MOLECULAR-DETAILED COMPUTATIONAL TOOLS TO STUDY HIV PATHOGENESIS AND DESIGN STEM-CELL BASED ANTI-HIV THERAPIES

by

Iraj Hosseini

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

July, 2015

© 2015 Iraj Hosseini
All Rights Reserved
Abstract

Combination antiretroviral therapy (cART) ensures that millions of people with HIV lead normal lives. However, cART is not a cure and if stopped, even after decades, HIV hidden in the latent reservoir can activate and lead to viral rebound. Given the drawbacks of cART particularly cost and difficulties of adherence to chronic treatment, HIV cures could significantly reduce the burden on patients while reducing the healthcare cost.

In 2008, the “Berlin Patient” was treated with myeloablative irradiation and hematopoietic stem cell transplant (HSCT) from a donor with a CCR5Δ32 mutation conferring resistance to HIV. Since then, the recipient has been “functionally cured”, i.e. has shown no signs of active HIV-1 replication in the absence of cART. This success renewed hope that replacing HIV-susceptible cells with more resistant cells by inserting genes or gene networks into patients’ or matched donors’ stem cells before transplantation could provide HIV-resistance to progeny target cells and lead to cure. This approach was recently shown in macaques to reduce viral load and return T cell counts to normal levels. Key questions remain: (a) given that donor chimerism occurs, what percentage of the cells must be HIV-resistant in order to block HIV; (b) what is the minimal level of anti-HIV activity needed in these cells; (c) which anti-HIV genes will work best, and for which patients; and (d) will combinations of anti-HIV genes synergize?

As few patients have undergone transplants, we built novel molecular-detailed mechanistic models of HIV infection to answer these questions. The models are validated against independent in vitro and in vivo experimental data. Using the models, we study the complex pathogenesis of HIV, design gene-augmented stem cell therapies, and
calculate the probability of cure for each therapy. We focus on HSCTs that include knocking out CCR5 and/or inserting anti-HIV genes or gene circuits such as the APOBEC3 family, SAMHD1, and on-demand apoptosis-inducing circuits. Instead of studying a single average course of HIV infection in a typical patient, we apply our models, parameterized using real patient data, to simulate a population of HIV-infected patients. Using this population of models, we run virtual clinical trials of different treatments. We validate the model by predicting recent clinical data from CCR5-modified T-cell therapy. The model has the ability to help design stem cell-based therapies and predict the results in clinical studies.

Advisor: Dr. Feilim Mac Gabhann, Assistant Professor, Institute for Computational Medicine and Department of Biomedical Engineering, Johns Hopkins University

Thesis Committee (alphabetically):

Dr. Joel Bader, Associate Professor, Institute for Computational Medicine and Department of Biomedical Engineering, Johns Hopkins University

Dr. Janice Clements, Professor, Dept. of Molecular & Comparative Pathobiology, Dept. of Neurology, and Dept. of Pathology, Johns Hopkins University
Note on Published Work

Content from the following chapters has been peer-reviewed and published, and is included with permission:

**Chapter 2.** Multiplexed Component Analysis to Identify Genes Contributing to the Immune Response During Acute SIV Infection (published in PLoS One [1])

**Chapter 3.** Multi-Scale Modeling of HIV Infection in vitro and APOBEC3G-Based Anti-Retroviral Therapy (published in PLoS Computational Biology [2])

**Chapter 4.** APOBEC3G-Augmented Stem Cell Therapy to Modulate HIV Replication: a Computational Study (published in PLoS One [3])
Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Feilim Mac Gabhann, for his kind support, constructive suggestions, and constant encouragement during the course of my PhD studies. He has been a fantastic mentor and great advisor, who patiently taught me to become a better scientist and generously allowed me to follow my passion in my time at Hopkins. Without Feilim, this dissertation simply would not have been possible. I consider myself incredibly fortunate for having had him as my mentor.

I also wish to express my appreciation to the members of my thesis committee, Dr. Janice Clements and Dr. Joel Bader for their suggestions and comments. Your feedback has made this work much stronger.

I would like to thank Dr. Lucio Gama at Hopkins for so many joyful scientific discussions on HIV and SIV and for always being enthusiastic about our work. I could not have asked for a better collaborator. Your friendship, good humor, and knowledge provided a real boost to this project. I am also grateful to Dr. Kapil Gadkar and Dr. Saroja Ramanujan at Genentech, for teaching me many computational and modeling skills, which were extremely useful in finishing this dissertation.

To Yasmin, Liz, Joe, Lindsay and other past and present members of the lab, thank you for your support and friendship. You are an intelligent group of colleagues, with whom I had many fruitful discussions about science and other topics. Special thanks to Dr. Yasmin Hashambhoy-Ramsay for being extremely welcoming when I joined the lab and for going out of her way to help me scientifically and professionally. I would also like to thank the dedicated faculty, staff and student members of BME EDGE, whom I
was very fortunate to have had the opportunity to work with for a common goal of making our great department even better for the students.

My most heartfelt appreciation goes to my dear parents, and wonderful brother and sister, who always supported me, loved me, and encouraged me in my academic pursuits. Undoubtedly, without their unconditional love, I would not be where I am today. Finally, my special thanks to my lovely wife, Shadi, for her love, patience, support, and understanding. Thank you for taking this journey with me.

This study was supported by the Ruth H. Aranow fellowship, Siebel Scholarship, and Willowcroft Foundation grant. I greatly appreciate these financial supports for making this research possible.
Table of Contents

Abstract ....................................................................................................................................... ii

Note on Published Work ........................................................................................................... iv

Acknowledgments .................................................................................................................. v

Table of Contents .................................................................................................................... vii

List of Tables .......................................................................................................................... xiii

List of Figures ......................................................................................................................... xiv

Chapter 1. Introduction ........................................................................................................... 1

1.1 What are HIV, AIDS, and cART? .................................................................................. 2

1.2 Anti-HIV stem cell therapies ......................................................................................... 2

1.3 Endogenous defense mechanisms against HIV ............................................................ 5

1.3.1 A3G-Vif interactions in CD4+ T cells ...................................................................... 5

1.3.2 SAMHD1 restricts the replication of HIV-1 in macrophages ................................. 9

1.3.3 Blocking CCR5 inhibits HIV entry into target cells .......................................... 9

Chapter 2. Multiplexed Component Analysis to Identify Genes Contributing to the Immune Response During Acute SIV Infection ................................................................. 12

2.1 Summary ......................................................................................................................... 13

2.2 Introduction ..................................................................................................................... 14

2.3 Methods .......................................................................................................................... 24

2.3.1 Animals and ethics statement ............................................................................... 24

2.3.2 RNA isolation from spleen, MLN, and PBMC ....................................................... 25

2.3.3 Quantification of SIV virions in plasma ................................................................. 25

2.3.4 NanoString nCounter gene expression system ..................................................... 25

2.3.5 Normalization methods ......................................................................................... 27
3.4 Results ................................................................................................................................. 111
  3.4.1 Single-cell model of APOBEC3G-Vif interactions .................................................... 111
  3.4.2 Multicellular model of HIV propagation in culture .................................................. 116
  3.4.3 Effects of A3G-based therapeutic strategies, drug penetrance and administration time on HIV replication ................................................................. 124
  3.4.4 Sensitivity analysis: determining critical model parameters that influence HIV replication ................................................................................................. 129

3.5 Discussion ......................................................................................................................... 132

Chapter 4. APOBEC3G-Augmented Stem Cell Therapy to Modulate HIV Replication:
  a Computational Study ......................................................................................................... 135
  4.1 Summary ....................................................................................................................... 136
  4.2 Introduction ................................................................................................................... 137
  4.3 Mathematical models ................................................................................................. 138
    4.3.1 Model I: the basic HIV model for A3G-augmented cells ........................................ 142
    4.3.2 Model IIa: the basic HIV model for WT and A3G-augmented cells ...................... 145
    4.3.3 Model IIb: the basic HIV model for WT and A3G-augmented cells with lower death rates for infected A3G-augmented cells ......................................................... 147
    4.3.4 Model IIc: the basic HIV model for WT and A3G-augmented cells with lower death rates for cells infected by A3G(+) viruses ....................................................... 149
    4.3.5 Model III: the basic HIV model for WT and A3G-augmented cells with auto-apoptosis capability ............................................................................................ 150
  4.4 Methods ....................................................................................................................... 151
    4.4.1 Estimation of the A3G-free virus release ratio, $p^{(w)}$, for WT CD4+ T cells ....... 151
  4.5 Results ......................................................................................................................... 153
    4.5.1 A3G gene therapy can effectively stop in vivo HIV replication .......................... 154
4.5.2 Blocking replication requires high percentage of A3G-augmented cells........ 156
4.5.3 Lower efficacy is achieved if infected A3G-augmented cells die more slowly than do infected WT cells.......................................................................................... 158
4.5.4 Lower efficacy is achieved if cells infected by A3G(+) viruses die more slowly than do cells infected by A3G(-) viruses................................................................................. 160
4.5.5 A3G-augmented cells with auto-apoptosis capability can stop replication more efficiently....................................................................................................................... 161
4.5.6 A3G\textsuperscript{ΔVif} outperforms A3G in blocking in vivo HIV replication ........164
4.5.7 Impact of mixed A3G overexpression levels on the A3G-SCT performance .... 165
4.6 Discussion................................................................................................................ 166
4.7 Appendix. calculating reproductive ratios in mathematical models............... 174
4.7.1 Model I: the basic HIV model for A3G-augmented cells (reduced burst size for cells infected by A3G(+) viruses) .................................................................................... 174
4.7.2 Model Ib: the basic HIV model for A3G-augmented cells (reduced infectivity rate for A3G(+) viruses)................................................................................................. 175
4.7.3 Model IIa: the basic HIV model for WT and A3G-augmented cells.............. 176
4.7.4 Model IIb: the basic HIV model for WT and A3G-augmented cells with lower death rates for infected A3G-augmented cells.......................................................... 179
4.7.5 Model IIc: the basic HIV model for WT and A3G-augmented cells with lower death rates for cells infected by A3G(+) viruses......................................................... 182
4.7.6 Model III: the basic HIV model for WT and A3G-augmented cells with auto-apoptosis capability............................................................................................................. 185
4.7.7 Model IV: the basic HIV model for A3G-augmented cells overexpressing A3G at low and high levels ............................................................................................ 188
Chapter 5. Mechanistic Models Predict Efficacy of CCR5-Deficient Stem Cell Transplants in HIV Patient Populations ................................................................. 191

5.1 Summary .............................................................................................................. 192

5.2 Methods ............................................................................................................... 193

5.2.1 Description of the model .............................................................................. 193

5.2.2 Model calibration ......................................................................................... 197

5.2.3 Virtual population ....................................................................................... 198

5.3 Results ................................................................................................................ 199

5.3.1 Clinical data (patient cohorts) ............................................................... 199

5.3.2 The “average” clinical profiles can be reproduced using parameter sets with biologically relevant, yet vastly different values ................................................. 202

5.3.3 Virtual patient populations capture the range of variability observed in the clinical data ......................................................................................... 209

5.3.4 Virtual clinical trials predict the efficacy of CCR5-based stem cell therapy ... 214

5.3.5 Validation of the virtual populations against clinical data from CCR5-based autologous T cell therapy in HIV-infected patients ...................................... 219

5.4 Conclusions and discussion .............................................................................. 220

Chapter 6. Computational Modeling of HIV Infection and Gene-Augmented Stem Cell-Based Cures .............................................................................................. 224

6.1 Summary .............................................................................................................. 225

6.2 Description of the extended model ................................................................... 225

6.3 Results ................................................................................................................ 228

6.3.1 Virtual clinical trials predict the efficacy of augmented stem cell therapies involving a single gene ................................................................. 228

6.3.2 Virtual clinical trials predict the efficacy of dual stem cell therapies .......... 232
6.4 Discussion..................................................................................................................233

Chapter 7. Future Directions..........................................................................................236

Bibliography .....................................................................................................................240

Curriculum Vitae ...............................................................................................................273
List of Tables

Table 2-1. List of genes measured by Nanostring technology ........................................ 15
Table 2-2. List of top two classifier PCs for each of the judges ...................................... 42
Table 2-3. Friedman test results for gene rankings .......................................................... 45
Table 2-4. Classification results using the top seven highly contributing genes ............... 73
Table 3-1. Values of $p$ and $c$ for different cases of viruses and cells. ......................... 109
Table 3-2. Estimations of burst size and virus infectivity rate for $t_{r,2} = 30$ hours .......... 121
Table 4-1. Parameter values used for simulations and calculations .............................. 140
Table 4-2. Estimations of burst size, infectivity rate and A3G-free virus release ratio .. 152
Table 5-1. Initial values for immune cells in each subpopulation ..................................... 198
Table 5-2. Subject characteristics and clinical measurements ......................................... 200
Table 5-3. Parameters of the model, their ranges for optimization, and definitions ....... 208
List of Figures

Figure 1-1. Anti-HIV stem cell therapy................................................................. 4

Figure 1-2. HIV life cycle....................................................................................... 7

Figure 2-1. Cytokine levels and SIV RNA in plasma versus time since infection.......... 20

Figure 2-2. Schematic of multiplexed component analysis (MCA) algorithm for evaluating gene expression datasets................................................................. 23

Figure 2-3. A hexagonal plot showing axes and vertices. ........................................ 33

Figure 2-4. Multivariate gene expression is predictive of output variables: score and loading plots for the spleen dataset analyzed by judge 2 - J2: (Orig, UV, PCA) ....... 36

Figure 2-5. The 12 judges identify both consensus and judge-specific genes: loading plots of 12 judges for the spleen dataset................................................................. 39

Figure 2-6. Percentage of variance captured by the top PCs. .................................. 41

Figure 2-7. Classification and cross validation in all datasets and for both classification schemes. ............................................................................................................... 43

Figure 2-8. Identification of tissue-specific and global genes: gene rankings across judges and datasets (tissues) for classification based on time since infection.................. 46

Figure 2-9. Identification of tissue-specific and global genes: gene rankings across judges and datasets (tissues) for classification based on SIV RNA in plasma................. 48

Figure 2-10. Gene rankings by the Pearson correlation method............................... 50

Figure 2-11. Gene rankings by the Spearman correlation method............................ 51

Figure 2-12. Gene rankings by the one-way ANOVA method................................... 52

Figure 2-13. Gene rankings by the significance analysis of microarrays (SAM) method.53

Figure 2-14. Classification results and evaluation of ranking similarity....................... 54
Figure 2-15. Classification results for the individual judges ........................................... 54

Figure 2-16. Classification results for the judges with log2-transformation .................. 54

Figure 2-17. Consensus among judges on gene contributions (spleen - time since infection). ..................................................................................................................... 55

Figure 2-18. Consensus among judges on gene contributions (MLN - time since infection). ..................................................................................................................... 56

Figure 2-19. Consensus among judges on gene contributions (PBMC - time since infection). ..................................................................................................................... 56

Figure 2-20. Consensus among judges on gene contributions (spleen – SIV RNA in plasma). ..................................................................................................................... 57

Figure 2-21. Consensus among judges on gene contributions (MLN - SIV RNA in plasma). ..................................................................................................................... 57

Figure 2-22. Consensus among judges on gene contributions (PBMC - SIV RNA in plasma). ..................................................................................................................... 57

Figure 2-23. Judge-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the spleen dataset ................. 59

Figure 2-24. Judge-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the MLN dataset ..................... 59

Figure 2-25. Judge-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the PBMC dataset ..................... 60

Figure 2-26. Judge-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the spleen dataset ..................... 60
Figure 2-27. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the MLN dataset.......................... 61

Figure 2-28. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the PBMC dataset. ....................... 61

Figure 2-29. Quantifying significance of gene ranking: *p*-value heatmap of the paired t-tests of gene rankings in all datasets and for both classification schemes.................. 63

Figure 2-30. Average correlation coefficient matrices in all datasets, for both classification schemes. .................................................................................................................. 66

Figure 2-31. The loading plots of average correlation coefficient matrices shown in Figure 2-30............................................................................................................................................. 67

Figure 2-32. The correlation coefficient matrices obtained from the loading plots in Figure 2-31............................................................................................................................................. 68

Figure 2-33. Correlations among genes simplified on a polar plot, illustrating MSD values and the ranking information................................................................. 70

Figure 2-34. Schematic of algorithm to plot Figure 2-33 (polar plots). ......................... 71

Figure 2-35. Tissue-specificity of genes: relative contribution genes to each tissue.... 74

Figure 3-1. HIV virion and its life cycle........................................................................ 99

Figure 3-2. Time-dependent profile of virus release from a single cell......................... 103

Figure 3-3. A schematic diagram showing state of the cells and a snapshot of the multicellular model. .................................................................................................. 108

Figure 3-4. Degradation profile of protein entities in the model. ............................... 114

Figure 3-5. Estimation of A3G and Vif production rates. ........................................... 114

Figure 3-6. HIV growth curves for WT and ΔVif viruses............................................. 119
Figure 3-7. Effects of HIV half-life and cell doubling time on infectivity rate and burst size. ........................................................................................................................................ 120

Figure 3-8. Distribution of cell states during the period of post infection.................. 123

Figure 3-9. Efficacy comparison of several drugs for different production rates........... 127

Figure 3-10. Effects of different production rates, penetrances and administration times of A3GΔVif on HIV growth curve.......................................................................................................................... 128

Figure 3-11. Comparison of model predictions using the two coupling methods........ 129

Figure 3-12. Sensitivity analysis in single-cell and multicellular models.................... 131

Figure 4-1. The basic HIV model: schematic diagram and simulations...................... 141

Figure 4-2. Extended HIV infection models for WT and A3G-augmented cells.......... 144

Figure 4-3. Estimation of $p^{(wt)}$, A3G-free virus release ratio, for WT CD4+ T cells... 153

Figure 4-4. Effects of A3G-free virus release ratio on HIV replication and reproductive ratio in Model I. ................................................................................................................................. 155

Figure 4-5. Effects of percentage of transfected cells, death rate ratios, and auto-apoptosis failure rate on HIV replication in Models IIa, IIb, IIc, and III.............................. 158

Figure 4-6. Effects of percentage of transfected cells, death rate ratios, and auto-apoptosis failure rate on reproductive ratio in Models IIa, IIb, IIc, and III. ..................... 163

Figure 4-7. Effects of A3G and A3GΔVif on reproductive ratio in Models IIa and III...] 164

Figure 4-8. Effects of $p^{(+lo)}$, A3G-free virus release ratio, on reproductive ratio. ....... 166

Figure 5-1. Clinical data from cART-naïve HIV-infected individuals in the multicenter AIDS cohort study (MACS). ........................................................................................................ 201

Figure 5-2. The values of objective function for the “optimized” set in each subpopulation. ................................................................................................................................. 203
Figure 5-3. Model calibrations................................................................. 204
Figure 5-4. Comparison of the 100 best parameter sets across subpopulations. ....... 205
Figure 5-5. Correlation analysis of the 100 best parameter sets in all populations. ...... 206
Figure 5-6. Dynamics of productive and latently infected T cells in the 100 “average” fits. ................................................................. 207
Figure 5-7. Dynamics of infected and uninfected macrophages and viremia produced by CD4+ T cells and macrophages in the 100 best “average” fits........................................ 207
Figure 5-8. Generation of virtual populations.................................................. 211
Figure 5-9. Temporal histograms of virtual patient profiles............................. 212
Figure 5-10. Comparison of the 1000 “virtual patient” parameter sets across subpopulations. .................................................................................................................. 212
Figure 5-11. Correlation analysis of the “virtual patient” parameter sets in all populations. .......................................................................................................................... 213
Figure 5-12. Dynamics of productive and latently infected cells in virtual patients. .... 213
Figure 5-13. Dynamics of infected and uninfected macrophages and viremia produced by CD4+ T cells and macrophages in virtual patients.................................................. 214
Figure 5-14. Performance prediction of CCR5-deficient stem cell therapy in rapid progressors. ........................................................................................................................................ 216
Figure 5-15. Calculating probability of cure for CCR5-deficient stem cell therapy in all populations. ....................................................................................................................... 218
Figure 5-16. Impact of variations in $\alpha_{\text{CCR5}}, \beta_{\text{CCR5}},$ and $\epsilon_{\text{CCR5}}$ on probability of cure for CCR5-deficient stem cell therapy in all populations. .................................................. 219
Figure 5-17. Validation of combined virtual populations against clinical data from CCR5-modified autologous T cell therapy in HIV-infected patients.......................... 220

Figure 5-18. Validation of each virtual population against clinical data from CCR5-modified autologous T cell therapy in HIV-infected patients.......................... 221

Figure 6-1. Viral load one year post-therapy................................................................. 230

Figure 6-2. Normalized CD4+ T cell counts one year post-therapy.............................. 231

Figure 6-3. Probability of cure for augmented stem cell therapies with a single gene... 231

Figure 6-4. Probability of cure for augmented stem cell therapies with two genes....... 233
Chapter 1. Introduction
1.1 What are HIV, AIDS, and cART?

Infection by the human immunodeficiency virus (HIV) is characterized by a dramatic and progressive infection and depletion of CD4+ T cells and a sustained state of chronic inflammation and immune activation. This leaves the body prone to many types of infections and cancers. If left untreated, HIV infection eventually will progress to AIDS (acquired immunodeficiency syndrome), clinically defined in terms of either a CD4+ T cell counts below 200 cells/µL or the occurrence of specific HIV-associated diseases.

Combination anti-retroviral therapy (cART) ensures that for millions of patients, HIV/AIDS is a chronic nonfatal disease. However, cART is not a cure and if stopped, even after decades, HIV hidden in the latent reservoir can activate and lead to viral rebound. It is also expensive and may lead to emergence of drug-resistant HIV strains. Cures are needed to eliminate maintenance costs and issues of adherence to long-term regimens.

1.2 Anti-HIV stem cell therapies

Thirty-two years after the discovery of HIV, there has been only one case of an HIV-infected individual reported to be functionally cured. This individual is known as the “Berlin Patient”, and in 2008 he was treated with myeloablative irradiation and hematopoietic stem cell transplant (HSCT) from a donor with a homozygous CCR5Δ32 mutation conferring resistance to HIV [4-6]. Since then, the recipient has not used combination anti-retroviral therapy (cART) and the virus appears to be eliminated. Recently, two Boston patients appeared HIV-free after reduced-intensity conditioning HSCT from donors without the rare CCR5Δ32 mutation; however their new immune
systems were vulnerable to re-infection and indeed the virus rebounded after 7 and 15 weeks in these patients [7]. Latent infection re-emerges because irradiation does not eliminate 100% of the recipient's immune system. Thus, irradiation and transplant are likely insufficient without the introduction of anti-HIV activity. Comparing the Berlin and Boston patients suggests that having a protective mechanism in the immune system is a necessary factor to achieve cure. This protection was provided by lack of a functional CCR5 gene in the Berlin patient; however, the new immune systems in Boston patients were as susceptible to HIV infection as their old immune systems. Notably, since 2008, at least 6 other patients received a graft from a donor with a homozygous CCR5Δ32 mutation [8-10]. However, none survived for longer than 1 year, suggesting that other key factors such as graft-vs-host effects are involved in the success of the therapy.

Finding a matched donor who has a homozygous mutation in CCR5 for each HIV patient is very challenging. However, the HIV-resistance conferred by the CCR5Δ32 mutation could be recapitulated in donor cells by augmenting the cells before transplant [11-13]. Rather than obtaining rare matched CCR5Δ32 donors, inserting anti-HIV genes ex vivo into patients’ or matched donors’ stem cells hematopoietic stem cells prior to transplant could provide HIV-resistance to the new immune system, and assist in viral elimination from the recipient’s system. This can be done by knocking out CCR5 using zinc fingers and/or by using ant-HIV genes or gene circuits such as APOBEC3 family, SAMHD1, and on-demand apoptosis-inducing circuits (Figure 1-1), which would provide a functional HIV cure as defined in [14]. This approach was recently shown to reduce plasma viremia and to return T cell counts to normal levels [15] in pigtail macaques that underwent bone marrow transplants augmented with mC46 [16,17], a virus fusion
inhibitor. This was in sharp contrast to control macaques, suggesting that protection of HSCs by anti-HIV genes plays a major role in HIV control. However, key unknowns remain to be cleared: (a) given that patients will have a chimeric immune system after the engraftment, what percentage of the cells must be HIV-resistant in order to clear the system; (b) what is the minimal level of anti-HIV activity needed in these cells to achieve cure? (c) which anti-HIV genes will work best, and for which patients; and (d) can we combine anti-HIV genes to get synergize effects? Given the small number of HIV patients who have undergone bone marrow transplants [18] and the limited number of animal experiments [15,18-21], we built a novel mechanistic model of HIV infection to answer these questions. Using the model, we study the complex pathogenesis of HIV, design and test anti-HIV stem cell therapies, and explore interpatient variability in response to these therapies.

**Figure 1-1. Anti-HIV stem cell therapy.**

Finding a matched donor who has a homozygous mutation in CCR5 for each HIV patient is very challenging. However, hematopoietic stem cells can be collected from the patient (autologous) or a matched donor (allogeneic) and treated to become HIV-resistant. This can be done by knocking out CCR5 using zinc fingers and/or by inserting ant-HIV genes or gene circuits such as APOBEC3 family, SAMHD1, and on-demand apoptosis-inducing circuits. Treated stem cells will be expanded *ex vivo* and re-infused back into the patient after bone marrow or total body irradiation, which kills the stem cells in the patient. Irradiation does not eliminate 100% of the recipient's stem cells or immune system (as evidenced by the observed post-transplant donor chimerism). Therefore, the post-engraftment immune system will be chimeric, i.e. it will be a mixture of new and old immune cells, which are progenies of the new and old stem cells.
1.3 Endogenous defense mechanisms against HIV

1.3.1 A3G-Vif interactions in CD4+ T cells

The innate immune system is a major line of defense against HIV-1, reducing viral replication and protecting neighboring cells from infection. Cytosolic host cell proteins with antiretroviral activities, termed restriction factors, are key in this battle between host and virus. The apolipoprotein B (apo B) messenger RNA (mRNA)-editing, catalytic polypeptide-like 3 (APOBEC3) family of proteins are known to be potent restriction factors and to counteract infection by HIV-1 (reviewed in [22-30]). While the seven APOBEC3 proteins have varying levels of potency, APOBEC3G (A3G) exhibits the highest activity against HIV-1 that lacks the viral infectivity factor (vif) gene in in vitro tissue culture [22,23].

The antiviral functions of A3G are still the subject of active research. Hypermutation of HIV cDNA via the deaminase functionality of A3G is thought to be the most important A3G mechanism against HIV-1. A3G can induce up to 10% guanosine to adenosine (G-to-A) mutations into viral reverse transcripts [23,27,31], by deaminating cytidine (C) to uridine (U) on the minus strand [32-36]. This high mutational frequency can destroy viral genome integrity, resulting in production of noninfectious virions. Several groups have suggested that deaminase-independent antiviral activities of A3G also play a role in blocking HIV-1 replication. These include, but are not limited to, inhibiting synthesis of viral cDNA by blocking translocation of reverse transcriptase along the template RNA [37-39], reduction in the ability of tRNA^{Lys3} primers to initiate reverse transcription [40,41], blocking integration of the double-stranded viral DNA by causing defects in cleavage of tRNA^{Lys3} primer [42], or inhibiting nuclear import of pre-
integration complex [43]. Note that it is the A3G from the cell in which the virus is made that binds to viral mRNA and gets encapsulated in progeny virions. It is only after the virus is released and infects another cell that the encapsulated A3G exerts both its deaminase-dependent and -independent activities. Evidence supporting this observation came from studies performed almost 10 years before discovery of A3G [44-46].

HIV-1 is a retrovirus, more specifically a lentivirus. It encodes nine genes, of which \( \text{gag}, \text{pol} \) and \( \text{env} \) are common to all retroviruses. The remaining 6 genes (\( \text{tat}, \text{rev}, \text{vif}, \text{vpr}, \text{vpu} \) and \( \text{nef} \)) encode proteins with accessory and/or regulatory roles crucial to HIV pathogenesis [47,48]. As mentioned earlier, HIV-1 has developed the ability to evade the antiviral activities of A3G through the expression of Vif, a viral-encoded protein [44-46]. Vif binds to A3G [49-52] and exerts multiple counter-mechanisms to block encapsulation of A3G into virions. One mechanism is Vif-induced degradation of A3G where Vif recruits an E3 ubiquitin ligase complex and facilitates degradation of A3G through the proteasomal pathway [36,53-59]. It has been suggested that Vif directly impedes encapsulation of A3G into virions [49,56,60]. Sequence analysis studies have shown that an Asp-Pro-Asp motif at positions 128–130 in A3G is crucially important for binding of Vif to A3G. The D128K mutation in A3G protects the protein from Vif-induced degradation [50,51,60]. Mutations at Tyr-124 or Trp-127 make the protein unable to bind viral RNA and therefore get packaged into viruses [52,61,62]. The 124-127 motif is located beside the 128-130 Vif-binding region in the 3D model structure of A3G [61], suggesting that Vif binding and RNA binding may be in competition. Figure 1-2 shows a diagram of HIV infection as well as interactions between HIV-1 and A3G in HIV producing and newly infected CD4+ T cells.
Figure 1-2. HIV life cycle.
Mechanism of HIV infection including viral entry, reverse transcription, integration of viral DNA, virion assembly and release of viral particles is schematically shown. A3G, a host protein and a restriction factor, binds to viral mRNA and gets encapsulated into the viral capsid. If viruses carrying A3G infect other cells, the packaged A3G will exert several antiviral activities, which include inducing G-to-A mutations into viral reverse transcripts by deaminating C to U on the minus strand, blocking multiple steps in reverse transcription and causing integration defects. Vif, a viral protein, binds to A3G and inhibits encapsulation of A3G into virions by facilitating degradation of this protein through the proteasomal pathway.

There is evidence from multiple sources that overexpression of A3G could be an effective HIV therapy. In vitro, elevated levels of A3G expression resulted in A3G overcoming the effects of Vif [36,60,63,64]. In addition, in vivo studies have also suggested that targeting the A3G-Vif pathway may provide a new class of anti-retroviral treatment, however, some clinical studies have provided controversial results [65-76], and to date, these studies on the effects of A3G on HIV disease progression have not covered large numbers of individuals. In a 2005 study, Jin et al. found that in a group of 25 untreated HIV+ patients, A3G mRNA levels were negatively correlated with HIV
viral loads and were significantly associated with CD4+ T cell counts [65]. Results reported by two other research groups found that subjects with high G-to-A hypermutation had lower plasma HIV RNA levels and higher CD4+ T cell counts; however A3G mRNA levels were not directly measured [66,67]. Interestingly, the association of reduced plasma HIV RNA levels with hypermutation was considerably greater than association of reduced plasma HIV RNA levels with the CCR5Δ32 allele [66]. Ulenga et al. found that the expression levels of A3G were correlated with the levels of hypermutation in the vif and env regions, but not in the gag region of the virus genome. On the other hand, their study suggested no correlation between plasma viral loads and the levels of hypermutation in the vif, env, and gag regions [68]. In contrast, another study published by the same group found that the expression levels of A3G mRNA in patients with low viral set point were significantly higher than those of patients with high viral set point [69]. While other clinical studies have also shown the ability of A3G to modulate in vivo HIV infection [70,71], some groups have not been able to reproduce the same results [72-76]. Further investigation on the role of A3G in HIV disease progression would greatly benefit the field.

In addition to the A3G-Vif axis being a potential therapeutic anti-HIV approach, recent studies have suggested that A3G may also be used as a preventive strategy against HIV-1. Biasin et al. demonstrated that HIV-exposed seronegative subjects had significantly increased A3G mRNA and protein expressions compared to HIV-seropositive patients and healthy control individuals [77]. This higher expression was associated with lower susceptibility of cells to in vitro HIV infection [77]. Similar results were reported in a 2009 study by Vázquez-Pérez et al., where the average A3G mRNA
expression was over 2-fold higher in exposed uninfected subjects that those of healthy control individuals [70].

1.3.2 SAMHD1 restricts the replication of HIV-1 in macrophages

Myeloid cells and monocyte-derived macrophages (MDM) are natural targets of HIV infection and serve as long-term HIV reservoirs that sustain virus replication. SIV studies have indicated that macrophages produce large amounts of HIV even after CD4+ T cells are depleted and are responsible for the sharp increase in the viral load during late stages of the disease [78,79]. Macrophages can disseminate virus to different tissues including the brain and are thought to be the key cells responsible for HIV-1 associated neurocognitive disorder (HAND), which is prevalent in up to 50% of HIV-infected patients, even for those who maintain suppression of virus replication [80,81].

Sterile alpha motif and histidine-aspartic domain containing protein 1, SAMHD1, is a deoxynucleoside triphosphate triphosphohydrolase that provides anti-HIV functionality by cleaving cellular dNTP and inhibiting early steps of reverse transcription [82,83]. SAMHD1 is highly expressed in non-cycling cells such as monocytes and macrophages, in which it exerts its anti-HIV functions [84]. Overexpression of SAMHD1, induced by IL-12 and IL-18, has shown to boost resistance of MDMs to HIV-1 infection [85].

1.3.3 Blocking CCR5 inhibits HIV entry into target cells

CCR5 is a major coreceptor for HIV entry into target cells [86,87] and its inhibition is an effective way to block HIV infection. CD4+ T cells with a CCR5 mutation are highly resistant to in vitro HIV infection [88]. HIV-infected individuals heterozygous for a 32-bp mutation in CCR5 generally have a slower CD4 T-cell decline and prolonged AIDS-
free survival [89,90]. Individuals homozygous for the mutation in CCR5 are resistant to HIV infection [91]. Finally, the Berlin patient, the only functionally cured HIV patient, underwent bone marrow transplant from a donor with a homozygous CCR5Δ32 mutation [4,6] before he stopped taking cART. Also, the CCR5 mutation does not appear to cause any significant pathologies [92,93]. Therefore, there is mounting evidence that knocking out CCR5 should provide protection against HIV in patients.

Several gene editing strategies have been developed to modify CCR5 in stem cells and make them resistant to HIV infection for autologous transplant in HIV-infected patients. These include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided endonucleases (CRISPR/Cas9 systems). Among these techniques, ZFNs present the most clinically advanced gene editing strategy. ZFNs are engineered to contain the endonuclease domain of a type 1 restriction enzyme and multiple DNA binding zinc fingers [94,95]. Each zinc-finger recognizes three or four base pairs of DNA, and hence an array of zinc fingers can be linked to extend the length of a DNA sequence that is recognized. For example, three to six zinc fingers can recognize 9 to 18-bp sequences, enough to uniquely target a sequence in the human genome. Since the nuclease domain of ZFNs cuts DNA when it dimerizes, two sets of zinc fingers are required. ZFNs create a double-strand break in the region containing the gene of interest, which is commonly repaired by an error-prone non-homologous end joining (NHEJ) cellular DNA repair mechanism. NHEJ introduces deletions and insertions with a high probability and hence disrupts the gene of interest. ZFNs have been extensively used to disrupt CCR5 in stem cells and CD4+ T cells intended for treating HIV-infected patients [20,96-98]. A recent clinical trial demonstrated that infusion of the
CD4+ T cells modified by ZFNs to become CCR5-deficient into HIV-infected patients appears to be safe [99].
Chapter 2. Multiplexed Component Analysis to Identify Genes Contributing to the Immune Response During Acute SIV Infection
2.1 Summary

Immune response genes play an important role during acute HIV and SIV infection. Using an SIV macaque model of AIDS and CNS disease, our overall goal was to assess how the expression of genes associated with immune and inflammatory responses are longitudinally changed in different organs or cells during SIV infection. To compare RNA expression of a panel of 88 immune-related genes across time points and among three tissues – spleen, mesenteric lymph nodes (MLN) and peripheral blood mononuclear cells (PBMC) – we designed a set of Nanostring probes. To identify significant genes during acute SIV infection and to investigate whether these genes are tissue-specific or have global roles, we introduce a novel multiplexed component analysis (MCA) method. This combines multivariate analysis methods with multiple preprocessing methods to create a set of 12 “judges”; each judge emphasizes particular types of change in gene expression to which cells could respond, for example, the absolute or relative size of expression change from baseline. Compared to bivariate analysis methods, our MCA method improved classification rates. This analysis allows us to identify three categories of genes: (a) consensus genes likely to contribute highly to the immune response; (b) genes that would contribute highly to the immune response only if certain assumptions are met – e.g. that the cell responds to relative expression change rather than absolute expression change; and (c) genes whose contribution to immune response appears to be modest. We then compared the results across the three tissues of interest; some genes are consistently highly-contributing in all tissues, while others are specific for certain tissues. Our analysis identified $CCL8$, $CXCL10$, $CXCL11$, $MxA$, $OAS2$, and $OAS1$ as top contributing genes, all of which are stimulated by type I interferon. This suggests that the
cytokine storm during acute SIV infection is a systemic innate immune response against viral replication. Furthermore, these genes have approximately equal contributions to all tissues, making them possible candidates to be used as non-invasive biomarkers in studying PBMCs instead of MLN and spleen during acute SIV infection experiments. We identified clusters of genes that co-vary together and studied their correlation with regard to other gene clusters. We also developed novel methods to faithfully visualize multi-gene correlations on two-dimensional polar plots, and to visualize tissue specificity of gene expression responses.

2.2 Introduction

Infection by the human immunodeficiency virus (HIV) is characterized by a dramatic and progressive depletion of CD4+ T cells and a sustained state of chronic inflammation and immune activation. Disease progression appears to be directly related to early events during acute infection, including an intense and coordinated production of plasma cytokines (“cytokine storm”) that is not observed in other chronic viral infections, such as Hepatitis type B and C [100]. Studies using macaques infected with simian immunodeficiency virus (SIV) corroborate these findings (Figure 2-1A-C), and provide insights on the complex network of immune regulatory genes that is triggered in response against the virus [101,102]. Because of the difficulties in establishing the precise time when an individual is infected by HIV, unraveling the effect of genes and their level of significance during acute SIV infection is key in understanding the mechanisms by which these viruses interact with the immune system. Using an SIV macaque model for AIDS and CNS disease, our group has been assessing how the expression of genes associated with immune and inflammatory responses are longitudinally changed in different organs.
or cells during SIV infection. Because of the large number of tissue samples and to be
cost effective, we designed a set of Nanostring probes to measure the expression of 88
immune-related genes that are routinely analyzed in several diseases (Table 2-1).

**Table 2-1.** List of genes measured by Nanostring technology.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Gene</th>
<th>Full name or Alias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Housekeeping Genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td>Actin</td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td></td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>HPRT</td>
<td></td>
<td>Hypoxanthine-guanine phosphoribosyltransferase</td>
</tr>
<tr>
<td>PBGD</td>
<td></td>
<td>Porphobilinogen deaminase</td>
</tr>
<tr>
<td><strong>Chemokines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL1</td>
<td>CCL1</td>
<td></td>
</tr>
<tr>
<td>CCL2</td>
<td></td>
<td>Monocyte chemotactic protein (MCP) 1</td>
</tr>
<tr>
<td>CCL3</td>
<td></td>
<td>Macrophage inflammatory protein (MIP) 1α</td>
</tr>
<tr>
<td>CCL4L1</td>
<td></td>
<td>MIP-1β (similar to human CCL4)</td>
</tr>
<tr>
<td>CCL5</td>
<td></td>
<td>RANTES</td>
</tr>
<tr>
<td>CCL7</td>
<td></td>
<td>MCP-3</td>
</tr>
<tr>
<td>CCL8</td>
<td></td>
<td>MCP-2</td>
</tr>
<tr>
<td>CCL11</td>
<td></td>
<td>Eotaxin-1</td>
</tr>
<tr>
<td>CCL13</td>
<td></td>
<td>MCP-4</td>
</tr>
<tr>
<td>CCL19</td>
<td></td>
<td>MIP-3β</td>
</tr>
<tr>
<td>Chemokine receptors</td>
<td>Cluster of differentiation</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>CCL20</td>
<td>MIP-3α</td>
<td></td>
</tr>
<tr>
<td>CCL24</td>
<td>Eotaxin-2</td>
<td></td>
</tr>
<tr>
<td>CX3CL1</td>
<td>Fractalkine</td>
<td></td>
</tr>
<tr>
<td>CXCL2</td>
<td>MIP-2α</td>
<td></td>
</tr>
<tr>
<td>CXCL3</td>
<td>MIP-2β</td>
<td></td>
</tr>
<tr>
<td>CXCL9</td>
<td>Monokine induced by gamma interferon (MIG)</td>
<td></td>
</tr>
<tr>
<td>CXCL10</td>
<td>Interferon gamma-induced protein 10 (IP10)</td>
<td></td>
</tr>
<tr>
<td>CXCL11</td>
<td>IP-9</td>
<td></td>
</tr>
<tr>
<td>XCL1</td>
<td>Lymphotactin</td>
<td></td>
</tr>
<tr>
<td>CCR1</td>
<td>CCR1</td>
<td></td>
</tr>
<tr>
<td>CCR2</td>
<td>CCR2</td>
<td></td>
</tr>
<tr>
<td>CCR3</td>
<td>CCR3</td>
<td></td>
</tr>
<tr>
<td>CCR4</td>
<td>CCR4</td>
<td></td>
</tr>
<tr>
<td>CCR5</td>
<td>CCR5</td>
<td></td>
</tr>
<tr>
<td>CCR7</td>
<td>CCR7</td>
<td></td>
</tr>
<tr>
<td>CXCR3</td>
<td>CXCR3</td>
<td></td>
</tr>
<tr>
<td>CX3CR1</td>
<td>CX3CR1</td>
<td></td>
</tr>
<tr>
<td>CD4</td>
<td>CD4</td>
<td></td>
</tr>
<tr>
<td>(CD)</td>
<td>CD14</td>
<td>CD14</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>CD16</td>
<td>CD16</td>
</tr>
<tr>
<td></td>
<td>CD68</td>
<td>CD68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interferons</th>
<th>IFNα1</th>
<th>IFNα1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IFNβ</td>
<td>IFNβ</td>
</tr>
<tr>
<td></td>
<td>IFNγ</td>
<td>IFNγ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type I Interferon receptors</th>
<th>IFNR1</th>
<th>IFNR1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IFNR2</td>
<td>IFNR2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interleukines</th>
<th>IL1B</th>
<th>IL1b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IL4</td>
<td>IL4</td>
</tr>
<tr>
<td></td>
<td>IL6</td>
<td>IL6</td>
</tr>
<tr>
<td></td>
<td>IL7</td>
<td>IL7</td>
</tr>
<tr>
<td></td>
<td>IL9</td>
<td>IL9</td>
</tr>
<tr>
<td>IL18BP</td>
<td>IL18bp</td>
<td></td>
</tr>
<tr>
<td>IL10</td>
<td>IL10</td>
<td></td>
</tr>
<tr>
<td>IL11</td>
<td>IL11</td>
<td></td>
</tr>
<tr>
<td>IL12A</td>
<td>IL12A</td>
<td></td>
</tr>
<tr>
<td>IL12B</td>
<td>IL12b</td>
<td></td>
</tr>
<tr>
<td>IL13</td>
<td>IL13</td>
<td></td>
</tr>
<tr>
<td>Cytokine receptors</td>
<td>IL17</td>
<td>IL17</td>
</tr>
<tr>
<td>-------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>IL25</td>
<td>IL25</td>
<td></td>
</tr>
<tr>
<td>IL27</td>
<td>IL27</td>
<td></td>
</tr>
<tr>
<td>IL28A/B</td>
<td>IL28A/B</td>
<td></td>
</tr>
<tr>
<td>IL29</td>
<td>IL29</td>
<td></td>
</tr>
<tr>
<td>IL8Rα</td>
<td>CXCR1 or CD128</td>
<td></td>
</tr>
<tr>
<td>IL12RB2</td>
<td>IL12Rb2</td>
<td></td>
</tr>
<tr>
<td>IL22Rα2</td>
<td>IL22Rα2</td>
<td></td>
</tr>
<tr>
<td>IL28R</td>
<td>IL28R</td>
<td></td>
</tr>
<tr>
<td>Interferon regulatory factors</td>
<td>IRF1</td>
<td>IRF1</td>
</tr>
<tr>
<td>IRF3</td>
<td>IRF3</td>
<td></td>
</tr>
<tr>
<td>IRF7</td>
<td>IRF7</td>
<td></td>
</tr>
<tr>
<td>Signal transducers and activators of transcription</td>
<td>STAT1</td>
<td>STAT1</td>
</tr>
<tr>
<td>STAT2</td>
<td>STAT2</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>STAT3</td>
<td></td>
</tr>
<tr>
<td>STAT5A</td>
<td>STAT5a</td>
<td></td>
</tr>
<tr>
<td>STAT5B</td>
<td>STAT5b</td>
<td></td>
</tr>
<tr>
<td>Kinases</td>
<td>JAK1</td>
<td>Janus kinase 1</td>
</tr>
<tr>
<td>TYK2</td>
<td>Tyrosine kinase 2</td>
<td></td>
</tr>
<tr>
<td>Other enzymes</td>
<td>IDO1</td>
<td>indoleamine 2,3-dioxygenase 1</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td>NOS2</td>
<td>Nitric oxide synthase 2</td>
</tr>
<tr>
<td>Interferon stimulated genes</td>
<td>MxA</td>
<td>Myxovirus resistance 1</td>
</tr>
<tr>
<td></td>
<td>OAS1</td>
<td>2'-5'-oligoadenylate synthetases</td>
</tr>
<tr>
<td></td>
<td>OAS2</td>
<td></td>
</tr>
<tr>
<td>Suppressor of cytokine signaling</td>
<td>SOCS1</td>
<td>SOCS1</td>
</tr>
<tr>
<td></td>
<td>SOCS3</td>
<td>SOCS3</td>
</tr>
<tr>
<td>Secreted phosphoprotein 1</td>
<td>SPP1</td>
<td>Osteopontin</td>
</tr>
<tr>
<td>Pathogen pattern recognition receptors</td>
<td>TLR2</td>
<td>Toll like receptors</td>
</tr>
<tr>
<td></td>
<td>TLR3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLR4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLR5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLR6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLR7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TLR9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MDA5</td>
<td>Melanoma Differentiation-Associated protein 5</td>
</tr>
<tr>
<td></td>
<td>ZBP1</td>
<td>Z-DNA-binding protein 1</td>
</tr>
<tr>
<td>Tumor necrosis factor-α</td>
<td>TNFα</td>
<td>Cachexin</td>
</tr>
<tr>
<td>Tumor necrosis factor ligand</td>
<td>TNFSF13</td>
<td>CD256</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Tumor necrosis factor receptor</td>
<td>TNFRSF13B</td>
<td>CD267</td>
</tr>
<tr>
<td></td>
<td>TNFRSF13C</td>
<td>CD268</td>
</tr>
<tr>
<td></td>
<td>TNFRSF8</td>
<td>CD30</td>
</tr>
</tbody>
</table>

**Figure 2-1. Cytokine levels and SIV RNA in plasma versus time since infection.**

The mean changes of cytokine mRNA levels in (A) Spleen, (B) MLN, (C) PBMC, relative to baseline in uninfected animals over time in groups of animals acutely infected with SIV. We used the cubic spline method to interpolate between the data points. Note that mRNA samples in our study were isolated from different animals euthanized at various time points and hence do not constitute a longitudinal study. (D) Viral load in plasma at 4, 7, 14, and 21 days post infection.
Univariate analysis of the gene expressions alone or studying the correlation between gene expressions and output variables such as time since infection and SIV RNA in plasma provides limited success in interpreting the data. This may be due to several reasons. First, the changes in gene expressions are essentially caused by SIV infection. This suggests that the mRNA measurements, regardless of the biological functions of genes, should be correlated with time since infection or SIV RNA in plasma, leading to many “hits” that are not biologically significant. In addition, the data could be noisy and focusing on the co-variance as the only metric can be misleading. Second, it is generally thought that multiple genes work together to orchestrate the immune response during acute SIV infection. Therefore, we use multivariate analysis techniques, which can compensate for the correlations between multiple genes, to study all the genes simultaneously. These techniques, including principal component analysis (PCA), independent component analysis (ICA), and partial least squares (PLS) regression, have been used in various biological applications such as tumor classification [103], biomarker identification in traumatic brain injury [104], predicting age of cytotoxic T cells [105], and classification of yeast gene expression data into biologically meaningful groups [106]. The main differences between univariate and multivariate analysis methods are addressed in a recent review by Saccenti et al. [107]. Note that prior quantitative knowledge of how the changes in expression of each gene impact the immune response during acute SIV infection is not available. For example, the system may be more sensitive to changes in the absolute values of mRNA measurement for some genes, but more sensitive to relative changes for other genes. Previous multivariate analysis studies emphasize only one of these possibilities, and select preferentially for genes that satisfy the assumption – for
example, selects for genes with high absolute changes, or only genes with high relative changes. Therefore, preprocessing the data to take into account various initial assumptions is a necessary step for performing a thorough study of the effect of genes on the immune response. Various normalization methods including mean-centering \cite{108,109}, autoscaling or unit-variance scaling \cite{109,110}, pareto scaling \cite{111,112}, maximum scaling \cite{113}, range scaling \cite{113,114}, vast scaling \cite{115}, and maximum likelihood scaling \cite{116,117} have been used prior to multivariate analysis methods. The advantages and disadvantages of these different normalization strategies were discussed in detail in \cite{112,118}.

We present a multiplexed component analysis (MCA) technique in which we combine a variety of preprocessing techniques with two popular multivariate analysis methods to develop a set of twelve “judges” (Figure 2-2A). Preprocessing emphasizes specific features of a dataset by using an array of methods such as mean-centering, unit-variance scaling, or coefficient of variation scaling (CV), applied on the original or log-transformed data. Using a multiplexed set of preprocessing techniques ensures that we incorporate multiple possibilities for how gene expression changes affect the immune response, and therefore do not artificially include or exclude potentially significant genes. We use PCA \cite{109,119-122} and PLS \cite{123,124} as multivariate analysis techniques, which are powerful tools in studying datasets where the variables (88 genes) outnumber the observations (24 animals). Each of the twelve judges observes the data distinctively from others, and provides a set of uncorrelated principal components (PCs). We identify top contributing genes in each tissue by ranking the overall weights (loadings) of genes on the top two classifier PCs. Combining the ranking information from all the judges, we
are able to identify genes that are consistently and statistically significantly ranked as top contributing genes. We also examine the relation between genes in the top two classifier PCs, to study the genes that co-vary together. Finally, we calculate the contribution of each gene to the classification in each tissue to evaluate whether mRNA measurements in PBMC can act as a possible surrogate of measurements in spleen and MLN.

Figure 2-2. Schematic of multiplexed component analysis (MCA) algorithm for evaluating gene expression datasets.

(A) Since there is no prior information on how the changes in gene expressions affect the immune response during acute SIV infection, we use an array of mathematical techniques to be able to observe the data from different viewpoints. A “judge” is defined as the combination of a transformation, a normalization technique and a multivariate analysis method. Each dataset is analyzed by 12 different judges, forming a Multiplexed Component Analysis (MCA). Each judge provides a model consisting of a set of principal components (PCs), which are used to classify datasets based on one of the two output variables: time since infection or SIV RNA in plasma (classification schemes). For each judge, the two PCs that provide the most accurate and robust classification are chosen for further analysis. (B) Normalization methods include mean-centering (MC), unit-variance scaling (UV), and coefficient of variation scaling (CV); each method results in a different representation of the data, emphasizing different characteristics of the original data set. The MC normalization method emphasizes the genes with the highest absolute variations; the UV normalization method gives equal weight to each gene in the dataset; the CV normalization method emphasizes the genes with the highest relative changes.
2.3 Methods

2.3.1 Animals and ethics statement

All animal studies were approved by the Johns Hopkins University Institutional Care and Use Committee (IACUC protocol #PR12M310), and all procedures followed the guidelines of the Weatherall Report, the USDA Animal Welfare Act, and the Guide for the Care and Use of Laboratory Animals. Twenty-four juvenile pigtailed macaques (Macaca nemestrina) were studied before and during acute SIV infection. Twenty of these animals were dual-inoculated with the molecular clone SIV/17E-Fr and the immunosuppressive swarm SIV/DeltaB670, as previously described [125]. Groups of six infected macaque were euthanized at 4, 7, and 14 days post infection (p.i.), and two animals were euthanized at 21 days p.i. Four animals were mock-inoculated and used as uninfected controls. The 4-day time point is included as the earliest time point that the virus could be found in the brain of infected macaques. All animals were considered SPF (specific pathogen free) before enrolling in our study, and were tested negative for SIV, Simian type D retrovirus, Herpesvirus simiae (B virus), Simian foamy virus and simian T-cell leukemia virus. Uninfected and infected macaques were housed in different rooms to prevent cross-contamination. Assigned veterinarians and trained technicians monitored the animals twice daily for signs of distress, including diarrhea, weight loss, and opportunistic infections, in order that early humane endpoints could be performed if necessary. Macaques were housed in facilities that are fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International (AAALAC), and fed a balanced commercial macaque chow (Purina Mills). Throughout the experiments, animals were housed in cages providing 6 square feet of space with
visual and auditory contact of conspecifics, and received environmental enrichment, such as manipulanda and novel foodstuffs. Before euthanasia, all macaques were perfused with sterile saline solution. Euthanasia was performed under veterinary supervision using an overdose of intravenous sodium pentobarbital while under deep ketamine sedation (10 mg/kg intramuscular), followed by perfusion with 1X PBS prior to tissue harvest. Organs, including spleen and mesenteric lymph node (MLN), were harvested, sectioned, and frozen at -80°C. Blood was collected for the isolation of peripheral blood mononuclear cells (PBMC) using a Percoll protocol (GE Healthcare Life Sciences, Pittsburgh, USA), and cell pellets were frozen and stored at -80°C.

2.3.2 RNA isolation from spleen, MLN, and PBMC

Total RNA was isolated from frozen spleen, MLN, and PBMCs using the RNeasy kit (Qiagen, Valencia, CA, USA), according to the manufacturer’s protocol. Samples were eluted in 60 µL of RNase-free water and frozen at -80°C until time to be analyzed.

2.3.3 Quantification of SIV virions in plasma

Plasma was collected during euthanasia, and SIV RNA was isolated using the QIAamp Viral RNA Mini Kit (Qiagen). Analysis was performed by qPCR as previously described [126].

2.3.4 NanoString nCounter gene expression system

Nanostring technology uses molecular "barcodes" (a string of fluorescent dyes that uniquely identify a specific transcript) and potent microscopic imaging to quantify genes of interest after a hybridization protocol. The technique does not require reverse
transcription or DNA amplification, and provides high reproducibility and sensitivity for
the detection of multiple transcripts [127]. CodeSets for 92 macaque genes (Table 2-1),
including four housekeeping genes, were designed according to the company's
specifications, based on rhesus macaques (Macaca mulata) annotated sequences. In
addition to the target-specific CodeSets, the kit also includes six positive probes for
quality control and seven negative controls whose sequences were obtained from the
External RNA Controls Consortium and are confirmed to not hybridize with mammalian
genomes. Isolated RNA was quantitated by spectrophotometry, and 250 ng of each sample
was sent for hybridization and consecutive quantitation to the Johns Hopkins Deep
Sequencing and Microarray Core. RNA counts were normalized by the geometric mean
of four housekeeping genes: actin (mean = 97053.6), GAPDH (24761.6), HPRT (1056.3),
and PBGD (259.2). Note that in this case, use of geometric mean is more appropriate than
the arithmetic mean or median, which are suited when variables take relatively similar
values. In the PBMC dataset, we replaced the measurements taking a value of zero with a
value of one after normalization. These include CCL11(1), CCL24(1), CX3CL1(1),
IL9(2), IL13(1), IL28A/B(1), and NOS2(1). The number in parentheses represents the
number of modified measurements out of 24 measurements for each gene. The
modifications avoid problems caused by log-transformation on mRNA counts equal to
zero. In the spleen and PBMC datasets, we observed few measurements taking values
lower than the average value of the negative controls. Although we did not change these
measurements, it is worth mentioning the genes with such features. These include
CCL7(1), IL13(1) and IL17(1) in the spleen dataset and CCL1(1), CCL24(2), IL11(1),
IL12A(1), IL13(1), IL25(2), IL9(1), NOS2(2), SPP1(1) in the PBMC dataset. The
number in parentheses represents the number of measurements taking values higher than zero and lower than the average value of the negative controls. We used normalized mRNA measurements from 88 genes as input variables in our analysis. The data sets supporting the results of this article are available in the NCBI Gene Expression Omnibus (GEO) database, [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51488].

2.3.5 Normalization methods

Mean-centering (MC) repositions the mean of the data to zero, by subtracting the average value of the variable from each measurement. If the data is not mean-centered before PCA/PLS, the first PC will be approximately in the direction of the mean of the data instead of being in the direction of maximum variance. Thus, genes that have similar distributions with different mean values will be equivalent after MC is performed on the dataset (Figure 2-2B). In the next sections, we will discuss that PCA and PLS attempt to find directions in the data with maximum variance. Therefore, in the MC-preprocessed datasets, genes with a higher variance tend to get higher loadings than other genes, i.e. these genes will be farther from the origin point (0, 0) in the loading plots, and therefore contribute more to the scores in the score plots. The MC normalization method emphasizes the genes with the highest absolute variations in mRNA measurements across animals.

The MC method may overestimate the influence of variables with a larger variance in the PCA/PLS models. To give equal weight to each variable in the model, the UV scaling method divides each mean-centered measurement by the standard deviation of the variable, resulting in unit variance for all the variables (Figure 2-2B). With UV scaling, the importance of higher-variance variables is reduced, while that of lower-variance
variables is increased. Thus while the impetus to use the UV method is to avoid skewing the analysis towards high-variance variables, this method introduces its own skewing. As PCA and PLS extract the direction of maximum variance, it would be difficult to predict what genes would get the highest loadings in the PCA/PLS models.

When the MC normalization method is used, the assumption is that changes in the absolute value of mRNA measurements affect the immune response. In contrast, when UV scaling is performed, the range of measurements for each variable is completely ignored and all the variables are given the same weight in modeling. This implies that no variable can dominate other variables in the model due to its large variance. To find a middle ground between these two cases, we apply another preprocessing method where each variable is divided by its mean. Then we subtract one from the results to make the data mean-centered (Figure 2-2B). In this case, each new variable will have a variance equal to the square of the coefficient of variation of the original variable. The coefficient of variation can be viewed as a normalized measure of variation because it measures the variability in the data with regard to the mean value. In our work, this method is called coefficient of variation (CV) scaling. The method emphasizes the genes with the highest amount of change relative to the mean value and will generally lead to a higher loading in the PCA/PLS models for genes with a high coefficient of variation.

We use an example to clarify the difference between the normalization methods. Assume that we have gene expressions for three genes. Depending on the normalization method, these genes can be of less, more or equal importance compared to others.

Gene A: [ 90, 100, 110]
Gene B: [990, 1000, 1010]
Gene C: [900, 1000, 1100]
If the data is normalized by the MC method, Genes A and B will have equal importance and their relative changes will be smaller than that of Gene C.

Gene A: \[ -10, 0, 10 \]
Gene B: \[ -10, 0, 10 \]
Gene C: \[ -100, 0, 100 \]

If the data is normalized by the CV method, Genes A and C will have equal importance and their relative changes will be larger than that of Gene B.

Gene A: \[ -0.1, 0, 0.1 \]
Gene B: \[ -0.01, 0, 0.01 \]
Gene C: \[ -0.1, 0, 0.1 \]

If the data is normalized by the UV method, Genes A, B, and C will all have equal importance.

Gene A: \[ -1, 0, 1 \]
Gene B: \[ -1, 0, 1 \]
Gene C: \[ -1, 0, 1 \]

As seen, the importance of a change in gene expression can be interpreted differently by different normalization methods.

2.3.6 Multivariate analysis methods and the judges

The gene expression datasets are first preprocessed using a transformation and a normalization method as described in the Results section. We analyze each preprocessed set of data, using both Principal Component Analysis (PCA) and Partial Least Squares regression (PLS). For PCA, we use the `princomp` function in Matlab. The two important outputs of this function are: 1) the loadings of genes onto each PC, which are the coefficients (weights) of the genes that comprise the PC; and 2) the scores of each PC for
each observation, which are the projected data points in the new space created by PCs. We impose orthonormality on the columns of the score matrix obtained by the `princomp` function and scale the columns of the loading matrix accordingly such that the score matrix multiplied by the transposed loading matrix still results in the original matrix of the data. This is necessary to study the correlation between genes in the dataset in a loading plot, provided that the two constructing PCs closely approximate the matrix of the data [128].

PLS regression is a method to find fundamental relations between input variables (mRNA measurements) and output variables (time since infection or SIV RNA in plasma) by means of latent variables called components [123,124]. In this chapter, we use the `plsregress` function in Matlab to perform PLS regression. This function returns PCs (loadings), the amount of variability captured by each PC, and scores for both the input and output variables. The columns of the score matrix returned by the `plsregress` function are orthonormal. Therefore one can study the correlation between genes in the dataset using the gene loadings in the loading plots.

We define a judge as the combination of a preprocessing method (transformation and normalization) and a multivariate analysis technique (Figure 2-2A). Each dataset, i.e. spleen, MLN, or PBMC, was analyzed by all 12 judges, forming a Multiplexed Component Analysis algorithm. The Matlab files for visualization and the MCA method are available at http://gforge.icm.jhu.edu/gf/project/mca_siv/docman/?subdir=140.

### 2.3.7 Classification and cross validation

In our analysis, we use a centroid-based clustering technique. We use two variables to cluster the animals into distinct groups: (1) time since infection; and (2) SIV RNA in
plasma (copies/ml) (Figure 2-1D). These variables thus define the 'classification schemes' discussed in the text. In score plots, we calculate the cluster centroids as averages of the scores within clusters. Observations are classified into clusters using shortest distance between observation and centroid. The true class information for each observation is known prior to the analysis, and if the cluster is assigned correctly, the classification is successful; otherwise it has failed. We perform the classification for all the 24 observations and calculate the classification rate.

For classification, all the data is used to train the model and the same data is classified into clusters. It is critical to measure the robustness of the classifiers to predict unknown observations. To do this, the dataset is divided into two sub-datasets: “training” and “test”. We use leave-one-out cross validation (LOOCV), in which a single observation from the dataset is selected as the test dataset. The remaining observations are used as the training dataset to build the model and to calculate the centroid for each cluster. Then, the test observation is projected onto the low-dimensional space created by the model and assigned to the nearest cluster. If the cluster is assigned correctly based on the prior class information, classification has succeeded; otherwise it has failed. We perform the cross validation for all 24 observations and calculate the estimated LOOCV rate.

2.3.8 Visualizing the relative contribution of genes in hexagonal plots

In a general scenario, each gene has three different overall ranks, ranging between 1 and \( L = 88 \). These three ranks could represent the “importance” to each class of the judges or represent the gene’s “contribution” to each tissue. The two cases follow the same mathematical formalism; here we will focus on the latter case, in which the three ranks are denoted by \((R_S, R_M, R_P)\), representing the gene’s contribution to spleen, MLN
and PBMC, respectively. Since genes contribute differently to each tissue, we measure the relative contribution of genes to identify which ones are tissue-specific and which ones contribute equally to all tissues. First we examine two extreme cases: 1) if $R_S = R_M = R_P$ for a particular gene, that gene is contributing equally to all tissues; 2) if $R_S = 1, R_M = 88,$ and $R_P = 88$ for a gene, this means that particular gene fully contributes to spleen, while it has very small contributions to MLN and PBMC. In addition, note that two hypothetical genes ranked $(1, 10, 10)$ and $(71, 80, 80)$ should not have the same relative contributions. While both genes have a higher contribution to spleen than MLN or PBMC, the relative contribution of the former is higher than the latter. We propose the following equations to convert gene overall ranks to relative gene contributions to each tissue.

$$R_M = \frac{R_S + R_M + R_P}{3}, \quad C_X = \frac{1}{3} + \frac{R_M - R_X}{L-1}$$

Where $R_M$ represents the baseline contribution of a gene to all tissues. The terms $R_M - R_X$ and $C_X$ represent the specific contribution and the relative contribution to tissue $X$, respectively. Note that $C_S + C_M + C_P = 1$ (the term $L-1$ in the denominator ensures that $-2/3 \leq r_X \leq 2/3$). If $C_S = C_M = C_P = 1/3$, this means that the gene is contributing equally to all tissues. The farther that a gene gets from $(1/3, 1/3, 1/3)$, there is more contribution to one tissue and less contribution to the other two. For example, $(1, 88, 88)$ and $(1, 1, 88)$ will be respectively converted to $(1, 0, 0)$ and $(2/3, 2/3, -1/3)$ using the mathematical formulation. We plot the relative gene contributions on a hexagonal plot, where $(1/3, 1/3, 1/3)$ is the center of the hexagonal. The hexagonal plot has three main vertices representing Spleen, MLN, and PBMC and three auxiliary vertices denoting PBMC&MLN, PBMC&Spleen, and MLN&Spleen. The arrows in Figure 2-3 show the...
directions in which the relative contribution of a gene decreases for each of the main vertices. For example, as we move from the main vertex Spleen to the auxiliary vertex PBMC&MLN, the relative contribution to spleen decreases linearly from 1 to -1/3. Grid lines are drawn perpendicular to the arrows and are used to help calculate the coordinates for a given point. To do so, one should draw parallel lines to the grid lines, find the intersection of those lines with the corresponding arrows and then calculate the distance from the corresponding main vertices.

Figure 2-3. A hexagonal plot showing axes and vertices.

2.4 Results

2.4.1 Data collection, preprocessing, and the twelve judges

In this study, we analyzed the RNA expression levels of 88 genes in spleen, mesenteric lymph node and PBMCs of macaques acutely infected with SIV. mRNA levels were quantified using Nanostring, a probe-based technique, and values were
normalized by the geometric mean of four housekeeping genes. The final counts were preprocessed, and then analyzed using PCA or PLS. Preprocessing the data had two steps: transformation and normalization. Transformation of raw data can be advantageous when some of the variables in the dataset have extreme measurements (outliers), resulting in a non-normal distribution for these variables. The outliers may exert a large impact on the model and overshadow other measurements. For datasets with non-zero values, one method to alleviate the non-normality of the data is to perform log-transformation [129]. We either use the original raw data (Orig) or perform log2-transformation on the data (Log2). Normalization of the data is common because the typical amount and the range of expression for each gene in the datasets can vary substantially. This can significantly affect analyses attempting to identify which genes are key during the acute SIV infection. The type of normalization used alters the type of gene expression changes that are assumed to be significant, which in turn is related to how these gene expression changes can affect the immune response. We use three preprocessing methods (Figure 2-2B): 1) Mean-centering (MC) subtracts the average value from each measurement to set the mean of the data to zero. The MC normalization method emphasizes the genes with the highest absolute variations in mRNA measurements across animals; 2) Unit-variance scaling (UV) divides the mean-centered variables by their standard deviation, resulting in unit variance variables. The UV normalization method is a popular method that gives equal weight to each variable in the dataset; 3) Coefficient of variation scaling (CV) divides each variable by its mean and subtracts one. This gives each variable the same mean, but a variance equal to the square of the coefficient of variation of the original variable. This method emphasizes the genes with the highest relative changes in mRNA measurements.
For a worked example illustrating the difference between the types of gene changes to which each normalization method is responsive, see the Methods section.

Each of our 12 judges is a combination of a preprocessing method (transformation and normalization) and a multivariate analysis technique, i.e. a judge can be represented by an ordered triple \((x, y, z)\) where \(x\) takes its value from \{Orig, Log\}, \(y\) takes its value from \{MC, UV, CV\} and \(z\) takes its value from \{PCA, PLS\} (Figure 2-2A). Therefore, there are 12 distinct judges in our analysis. We use * to denote all the possible options for a particular triple element; for example, \((Log, *, PCA)\) defines all the judges that use \(Log\)-transformation and the PCA analysis method. In this chapter, the dataset for each tissue (spleen, MLN, PBMC) was analyzed by all 12 judges, forming the Multiplexed Component Analysis algorithm.

2.4.2 Animals cluster into separate groups in the score plot

After PCA or PLS is performed on the preprocessed data, observations are projected onto a low dimensional space and are assigned new coordinates, called scores. Figure 2-4A shows a PC1-PC2 score plot of 24 observations (animals) in the spleen dataset analyzed by judge 2 - J2: \((Orig, UV, PCA)\). Together, these two principal components capture 57.5% of the variation in the dataset. Although PCA is an unsupervised method with no information on the time since infection, it is seen that dots with the same color (animals with the same time of infection) grouped together. The red dashed ellipse is drawn using Hotelling's \(T^2\) statistic [130] to determine the 95% confidence interval, which contains all the dots except animal #18. A circular pattern is seen in Figure 2-4A: uninfected animals (red dots) lie in the top left quadrant but they move to the top right quadrant 4 days after infection (green dots), and the bottom right quadrant at 7 days (blue
dots), and settle in the bottom left quadrant (brown and black dots), possibly demonstrating a new steady state at 14-21 days. Figure 2-4B is the corresponding loading plot, where the weight of each gene on PC1 and PC2 is shown. The loading and score plots are closely linked such that genes that are highly loaded in a specific direction in the loading plot contribute more to the observations that are located in that direction in the score plot. For example, type I interferons (IFNα and IFNβ) and interferon-stimulated genes (MxA, OAS1, OAS2) are rapidly and significantly upregulated during the first days of SIV infection and they share similar location with the 4-day group in the score plot. In general, genes that are far from the origin point (0, 0) in the loading plot, i.e. highly loaded on PC1 and PC2, contribute more to the scores in the score plots. The loading plots are used to find correlated genes through clustering genes that are located in a particular direction, provided that the two PCs constructing the loading plots satisfactorily approximate the matrix of the data [128]. For example, functionally related inflammatory genes such as type I interferons are located in the top right quadrant, which indicates the correlation between their expression profiles.

Figure 2-4. Multivariate gene expression is predictive of output variables: score and loading plots for the spleen dataset analyzed by judge 2 - J2: (Orig, UV, PCA).

(A) Score plot; each dot represents an observation (animal), projected onto PC1 and PC2. Although the PCA method is given no information on the time since infection, clearly animals cluster with their time points. The red dashed ellipse determines the 95% confidence interval, which is drawn using Hotelling's T² statistic. (B) Gene loadings (weights) for the top two PCs. Genes that are highly loaded on PC1 and PC2 (i.e. far from the origin) contribute more to the scores in the score plot than other genes. Genes located in the same direction are highly correlated. The results for other judges testing the spleen dataset are shown in the Appendix (score plots) and Figure 2-5 (loading plots)
2.4.3 The judges have distinct interpretations of the spleen

In this section, we focus in detail on the spleen dataset analyzed by all the 12 judges, as described in the methods; we applied the methods to the other tissues as well. Each judge emphasizes a unique type of change in gene expression, and hence the shapes of the gene clouds calculated by each judge are different, showing that different genes could be predicted to be significant depending on the underlying assumptions, which are different for each judge.

In Figure 2-5, the loading plots for the first two components are shown for the 12 judges. The loading plot constructed by $J1: (\text{Orig, MC, PCA})$ shows both $MxA$ and $CXCL10$ are loaded higher than other genes. For $J7: (\text{Orig, MC, PLS})$, where PCA is replaced by PLS, we see that, the gene cloud is approximately mirrored compared to the gene cloud of judge 1. Other high-loading genes for these two judges include $OAS1$, $OAS2$, $CXCL11$, and $IDO1$. Comparing no transformation ($\text{Orig}$) with log$_2$-transformation ($\text{Log2}$) in judges 1 and 4, we observe that judge 4 is less dominated by a small number of highly-loaded genes than judge 1 (the scales on the axes are different). A common feature of judges 1, 4, 7, and 10 is that the MC is the normalization method. Thus, if we assume that changes in the absolute value of gene expressions have significant impacts on the immune response, genes such as $MxA$, $CCL8$, and $CXCL10$ are highly contributing to the immunological events observed during acute SIV infection. Indeed, $MxA$ is one of the most reliable surrogates for the measurement of type I interferon response both in vitro and in vivo [131], and $CCL8$ and $CXCL10$ are important chemoattractants for monocytes and activated lymphocytes, respectively [132].
In our analysis, we either use the original data or log₂-transformed data. Before performing PCA or PLS, preprocessing methods are applied on the data to make it mean-centered (MC), UV-scaled, or CV-scaled. We used time since infection as the output variable for PLS. The loading plots on PC1 and PC2 are shown for each judge. In the 1st column (MC), there are a few highly loaded genes and the rest of genes are located in a dense cloud. In the 2nd column (UV), there is no single gene with a significantly high loading and the cloud of genes is more spread out. In the 3rd column (CV), we see a combination of the main features of the previous columns: a few highly loaded genes and spread gene clouds. The score and loading plots for the 12 judges in other datasets (and other classification schemes) are available in the Appendix.
Unlike (\(\ast, MC, \ast\)) in the 1st column, no single gene with a significantly high loading is seen in the 2nd column, constructed by (\(\ast, UV, \ast\)). Instead, we observe a group of genes that have higher loadings than others. This is somewhat expected since all the genes have equal variance when the UV scaling is performed. Comparing \(J2: (Orig, UV, PCA)\) and \(J5: (Log2, UV, PCA)\), we observe that there is a slight rotation between the two gene clouds, while they are similar in terms of the distance of genes from the origin and their relative location. If PC2 in judge 8 is multiplied by -1, the same scenario will be observed between judges 8 and 11, both of which use the UV normalization method. This indicates that UV scaling may alleviate the issue of non-normality and therefore log2-transformation has a lesser effect in this case.

The CV scaling method, used in the 3rd column, preprocesses genes to have their variance equal to the square of the coefficient of variation of the original genes. Therefore, it lies somewhere between the UV scaling method, which gives equal variance to each variable, and the MC normalization method, which does not modify the variance of variables at all. Here, we also observe that the 3rd column of judges, \(\ast, CV, \ast\), shares features with both the first and second columns, i.e., a few highly loaded genes as well as a spread cloud of genes. The preprocessing methods clearly impact the shape of the gene clouds constructed by PC1 and PC2, and hence changing the loading (importance) of genes under each assumption. In the next section, we define metrics to select the best pair of PCs for each judge to perform further analysis.
2.4.4 The choice of top classifier PCs varies between the judges

The score plots provided by PCA and PLS methods are used to cluster observations into separate groups based on the information on time since infection or SIV RNA in plasma. For each judge, dataset (tissue) and classification scheme (time since infection or SIV RNA in plasma), our goal is to find a score plot that provides the most accurate and robust classification of observations and to study the gene loadings in the corresponding loading plot. For each judge, we look at 28 score plots generated by all the combinations of two of the top eight PCs. This is because in all cases a high degree of variability, at least 76% and on average 87%, is captured by the top eight PCs (Figure 2-6).

![Figure 2-6. Percentage of variance captured by the top PCs.](image)

The total amount of variance captured by the top 4 PCs (dark colors) and the top 8 PCs (light colors) are shown for classification based on (A) time since infection and (B) SIV RNA in plasma in the spleen (green), MLN (blue), and PBMC (red) datasets. In all cases, at least 76% of the variance is captured by the top eight PCs.

Next, we perform centroid-based classification and cross validation to obtain classification and LOOCV rates, indicative of the accuracy and the robustness of the classification on a given score plot, respectively. The PCs representing the highest accuracy and robustness are chosen as the top two classifier PCs for that judge (Table...
PC1 and PC2 are the most commonly chosen classifier PCs, comprising 75% and 51% of all pairs, respectively. This is expected, as PC1 and PC2 capture the highest amount of variability among PCs. The PC1-PC2 pair is chosen in 25 out of 72 cases, followed by PC1-PC3 and PC1-PC4, each chosen in 9 cases.

Table 2-2. List of top two classifier PCs for each of the judges.

<table>
<thead>
<tr>
<th>Time since infection</th>
<th>SIV RNA in plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spleen</td>
</tr>
<tr>
<td><strong>J1: (Orig, MC, PCA)</strong></td>
<td>PC1-PC5</td>
</tr>
<tr>
<td><strong>J2: (Orig, UV, PCA)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J3: (Orig, CV, PCA)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J4: (Log2, MC, PCA)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J5: (Log2, UV, PCA)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J6: (Log2, CV, PCA)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J7: (Orig, MC, PLS)</strong></td>
<td>PC1-PC4</td>
</tr>
<tr>
<td><strong>J8: (Orig, UV, PLS)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J9: (Orig, CV, PLS)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J10: (Log2, MC, PLS)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J11: (Log2, UV, PLS)</strong></td>
<td>PC1-PC2</td>
</tr>
<tr>
<td><strong>J12: (Log2, CV, PLS)</strong></td>
<td>PC1-PC2</td>
</tr>
</tbody>
</table>
The results of clustering for both classification schemes are shown in the score plots in the Appendix and summarized in Figure 2-7. In most cases for time since infection (Figure 2-7A), the classification rates are higher than 75% (mean = 83.9%) and the LOOCV rates are higher than 60% (mean = 70.9%). For SIV RNA in plasma in most cases (Figure 2-7B), classification rates are higher than 60% (mean = 69.2%) and the LOOCV rates are higher than 54% (mean = 61.9%).

Figure 2-7. Classification and cross validation in all datasets and for both classification schemes.
The classification and LOOCV rates for the top classifier PCs are shown for each judge for classifications based on (A) time since infection and (B) SIV RNA in plasma. Light and dark colors represent the classification and the LOOCV rates, respectively. (C-H) The average classification and LOOCV rates are also shown for judges using a common feature, i.e. Orig vs. Log2, MC vs. UV vs. CV, and PCA vs. PLS. In general, we observe that clustering based on SIV RNA in plasma is less accurate and less robust than the classification based on time since infection.
We observe that clustering based on SIV RNA in plasma is generally less accurate and less robust than the classification based on time since infection. This may suggest that measuring SIV RNA in plasma alone does not provide a good indicator for the changes in immunological events during SIV infection due to the complex interactions between the virus and the immune system. Indeed, during HIV infection, markers for cellular activation are better predictors of disease outcome than plasma viral load [133].

In order to find whether there is a particular transformation, or preprocessing, or multivariate analysis that systematically provides more accurate and robust results than others, we calculated the average classification and LOOCV rates for judges that have a common feature, i.e. Orig vs. Log2, MC vs. UV vs. CV, and PCA vs. PLS (Figure 2-7C-H). In our datasets, the overall conclusion is that each of the judges has merit and can outperform others in some cases. It would be difficult to argue that one judge is clearly better than others when we consider both classification and LOOCV rates. Since each judge observes the data from a distinct viewpoint and we want to consider various assumptions on how the immune response is affected by the changes in gene expressions, we combine their opinions to identify significant genes during acute SIV infection.

In general, after the classification and cross validation are performed, the judges need to be evaluated based on their accuracy and robustness. If a judge has a low accuracy compared to others, that judge can be removed from further analysis. Alternatively, more accurate judges can be given higher weights when the results are combined. In this application, all the judges have high and approximately similar accuracy and robustness and hence we give them equal weights when we combine the results. Note that although
the judges have similar accuracy, each of them analyzes data differently and assigns distinguishably different loadings to the genes (loading plots in the Appendix).

2.4.5 CCL8 is identified as the top “contributing” gene by all the judges

Genes that are highly loaded (distant from the origin) contribute more to the scores that were used for classification, and hence are considered as top “contributing” genes. To find these genes, we calculate the distance of each gene from the origin in the loading plots (loading plots in the Appendix) and rank the values with the highest rank equivalent to the maximum distance, i.e. the highest contribution. Therefore for a given dataset and a classification scheme, each gene is assigned a rank (highest \(\equiv 1\); lowest \(\equiv 88\)) from each judge, resulting in a total of 12 ranks for each gene.

The first level of analysis is whether any of the genes are ranked consistently higher or lower than the other genes, across all judges. To answer this, we create a \(88\times12\) gene ranking table where rows and columns correspond to genes and judges, respectively. Using the Friedman test, we obtained extremely small \(p\)-values (Table 2-3), suggesting that in all three tissues and for both classification schemes there is at least one gene that is consistently ranked higher or lower than others.

Table 2-3. Friedman test results for gene rankings.

<table>
<thead>
<tr>
<th>Time of Infection</th>
<th>Plasma SIV RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spleen</td>
</tr>
<tr>
<td>Spleen</td>
<td>7.99e-83</td>
</tr>
<tr>
<td>MLN</td>
<td></td>
</tr>
<tr>
<td>PBMC</td>
<td></td>
</tr>
</tbody>
</table>
The genes are sorted based on the average of their 12 ranks in Figure 2-8A-C (time since infection) and Figure 2-9A-C (SIV RNA in plasma). To find the overall contribution of genes, the genes are also sorted based on the average of their three overall ranks (Figure 2-8D and Figure 2-9E). \textit{CCL8} is ranked as the highest contributing gene in both classification schemes. Albeit with a different order of contribution, \textit{CCL8} is followed by \textit{CXCL10}, \textit{CXCL11}, \textit{MxA}, \textit{OAS2}, and \textit{OASI} in the two classification schemes. These genes always appear among the top eleven contributing genes in all tissues and for both classification schemes. These genes are all stimulated by type I interferon, suggesting that the cytokine storm we here identify in lymphoid tissues - and that is also observed in the plasma of patients during acute HIV infection - reflects a systemic innate immune response against viral replication [100,134]. While there are genes that contribute highly to all three tissues, among the transcripts analyzed in this project we cannot identify a single gene that consistently appears in the lowest eleven contributing genes.

\textbf{Figure 2-8. Identification of tissue-specific and global genes: gene rankings across judges and datasets (tissues) for classification based on time since infection.}

The highly loaded genes contribute more to the scores that are used for classification, and hence are considered as the top “contributing” genes. To study genes based on their contribution, we calculate the distance of each gene from the origin in the loading plots and rank the distance values in a descending order with the highest rank equivalent to the maximum distance, i.e. the highest contribution. For a given dataset, each gene is assigned a rank (highest \(\equiv 1\); lowest \(\equiv 88\)) from each \textit{judge}, resulting in a total of 12 ranks for each gene. Then, we calculate the average of twelve ranks for each gene and sort the results from the high-ranking genes (dark blue) to the low-ranking genes (dark red) in the (A) spleen, (B) MLN and (C) PBMC datasets. This leads to an overall rank for each gene in each of the datasets. (D) We calculate the average value of the three overall ranks and sort the results in a descending order of contribution. We observe that
CCL8, followed by MxA, CXCL10, CXCL11, OAS2, and OAS1 are ranked as the top contributing genes in all datasets.
Figure 2-9. Identification of tissue-specific and global genes: gene rankings across judges and datasets (tissues) for classification based on SIV RNA in plasma.
To evaluate our MCA method, we compared its ranking results with those of other methods including the Pearson correlation (Figure 2-10), the Spearman correlation [135,136] (Figure 2-11), One-way analysis of variance (ANOVA) (Figure 2-12), and the significance analysis of microarrays (SAM) [137] (Figure 2-13) methods, all of which are used to rank the genes. Note that t-statistics and fold-change methods are also used in literature, but they are limited to classifications based on two groups. For each method, we selected the top five genes in each dataset and built decision trees to classify the observations using the selected genes. In most cases, the generated trees overfitted the dataset, and hence we pruned the trees and chose the sub-tree with the lowest cross validation error rate. The results indicate that, in 11 out of 12 cases, the top genes selected by MCA have substantially better classification power than those selected by the Pearson or Spearman correlation methods (Figure 2-14A and Figure 2-14C). The classification results of the SAM and ANOVA methods are similar to those of the MCA method. Furthermore, the Spearman's rank correlation coefficients, measuring the degree of similarity between the rankings of the MCA and other methods, indicate high correlations between the MCA and SAM methods (Figure 2-14B and Figure 2-14D). We also showed that in most cases the classification power top 5 average-ranked genes selected by all the judges is equally well or better than that of the top 5 genes selected by each individual judge (Figure 2-15) or that top 5 average-ranked genes selected by the judges with log2-transformation (Figure 2-16).
Figure 2-10. Gene rankings by the Pearson correlation method.
Figure 2-11. Gene rankings by the Spearman correlation method.
Figure 2-12. Gene rankings by the one-way ANOVA method.
Figure 2-13. Gene rankings by the significance analysis of microarrays (SAM) method.
Figure 2-14. Classification results and evaluation of ranking similarity.
(A and C) For each method, we selected the top five genes and built decision trees to classify the observations using those genes. (B and D) We used Spearman's rank correlation coefficients to measure the degree of similarity between the rankings of the MCA and other methods.

Figure 2-15. Classification results for the individual judges.
For each judge, we selected the top five genes in each dataset and built decision trees to classify the observations using the selected genes for (A) time since infection and (B) SIV RNA in plasma. The last column, labeled MCA, shows the results when we performed classification using the top five average-ranked genes by all the judges.

Figure 2-16. Classification results for the judges with log2-transformation.
For judges with log2-transformation, we selected the top five average-ranked genes in each dataset and built decision trees to classify the log-transformed observations for (A) time since infection and (B) SIV RNA in plasma. The second column, labeled MCA, shows the classification results using the top five average-ranked genes by all the judges.
The level of agreement between *judges* on the gene contributions varies substantially among genes. Similar colors across a row, such as *CXCL11* and *CCL2* in Figure 2-8B, show a high degree of consensus among *judges*, while there is a significant amount of disagreement between *judges* on rows with mixed colors, such as *CCL24* in Figure 2-8A. To measure the degree of consensus, we calculated the range and the standard deviation of the 12 ranks for each gene (Figure 2-17 - Figure 2-22). For a given gene, there is more agreement between *judges* when both the standard deviation and the range take low values. Typically, the high contributing genes tend to be located in the left bottom corner of Figure 2-17 - Figure 2-22, suggesting that there is a high degree of agreement between *judges* on the contribution of these genes.

![Figure 2-17. Consensus among *judges* on gene contributions (spleen - time since infection).](image)

For a given value of the standard deviation, there is more disagreement on the contribution of a gene with a higher value for the range, i.e. as we horizontally move from left to right, the degree of disagreement from *judges* increases. Also, for a given value of the range, there is more disagreement on the contribution of a gene with a higher value of standard deviation, i.e. as we vertically move from the bottom to the top, genes
experience a higher degree of disagreement from *judges*. The color of each dot represents its overall rank in the dataset. For example in Figure 2-18, *MxA* experiences a higher degree of disagreement compared to *CCL8* (the same horizontal line) and there is less consensus between *judges* on *STAT1* compared to *CCL3* (the same vertical line). For both classification schemes, there is a greater degree of agreement between *judges* in the MLN dataset than in spleen and PBMC. This can be visually seen in Figure 2-8 and Figure 2-9, where the gene rankings in the MLN dataset show the most consistency.

Figure 2-18. Consensus among *judges* on gene contributions (MLN - time since infection).

Figure 2-19. Consensus among *judges* on gene contributions (PBMC - time since infection).
Figure 2-20. Consensus among judges on gene contributions (spleen – SIV RNA in plasma).

Figure 2-21. Consensus among judges on gene contributions (MLN - SIV RNA in plasma).

Figure 2-22. Consensus among judges on gene contributions (PBMC - SIV RNA in plasma).
Furthermore, we evaluated how genes were assigned differential rankings by the judges with a common feature, specifically, MC- vs. UV- vs. CV-based judges. The average of 4 ranks given by each class of the judges was calculated. This results in three ranks for each gene, representing the importance of that gene to each class of the judges. To identify how different judges analyzed the datasets, we created a metric of the relative importance of each gene (see S6 Method). The results are shown in hexagonal plots (Figure 2-23 - Figure 2-28), where genes in the center have equal importance to all three classes of the judges. The proximity of a gene to a vertex indicates that the gene has more importance to the class or classes of the judges noted at that vertex. The inner color of each dot represents the average of the ranks, whereas the outer color represents the minimum (best) of the three ranks. The congested region in the center of the hexagon houses most of the genes and is amplified on the right-hand plot. For example, in Figure 2-24, genes in the center such as CXCL11, CCL8, CXCL10, and MxA have approximately the same blue color for the inner and outer circles, showing that these genes are important to all three classes and the level of importance to each class is the same. On the other hand, CCL24 has moderate importance when the decision of all the judges are combined, but it has a relatively high importance to CV-based judges. This suggests that CCL24 is one of the genes with the highest amount of change relative to the mean value. Note that if a gene is only important to CV-based judges, then it is likely to be biologically relevant only if high relative changes are the trigger for downstream effect. Such a gene would be ignored if only UV- or MC-based methods were used.
Figure 2-23. Judge-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the spleen dataset.
In each hexagonal plot, three main vertices represent MC-, UV-, and CV-based judges. Three auxiliary vertices denote CV&UV, CV&MC, and UV&MC. Genes close to one of these vertices are relatively more important to that class of judge. For example, genes that are close to CV&MC have equal importance to both CV- and MC-based judges. Genes at the center have approximately similar importance to each class of the judges. The coordinates are formatted as the relative gene importance, $C_{UV}$, $C_{MC}$, $C_{CV}$, taking values in the range $[-1/3, 1]$ and satisfy $C_{UV} + C_{MC} + C_{CV} = 1$ (see S6 Method for further explanation of coordinates).

Figure 2-24. Judge-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the MLN dataset.
Figure 2-25. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for time since infection in the PBMC dataset.

Figure 2-26. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the spleen dataset.
Figure 2-27. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the MLN dataset.

Figure 2-28. *Judge*-specificity of genes: relative importance of each gene using each normalization method, for SIV RNA in plasma in the PBMC dataset.

2.4.6 Gene rankings are more statistically significant in the MLN dataset

We study the statistical significance of the gene contributions by running a paired t-test for every two rows (genes) of the 88×12 table to evaluate the null hypothesis that the
two genes have equal contribution against the alternative hypothesis that one gene contributes significantly higher than the other one. If the \( p \)-value of the test takes sufficiently small values, it shows that one of the genes has a significantly higher contribution (Figure 2-29). Using linkage analysis (dendrograms), we identified clusters of genes that are statistically ranked higher than other succeeding gene clusters (\( \alpha = 0.05 \)).

For example in Figure 2-29A, the highest contributing group of genes consists of \( MxA, OAS2, OAS1, \) and \( CCL8 \). In this group, the sharpest statistical difference is between \( MxA \) and \( OAS1 \) with a \( p \)-value of 0.55, suggesting that none of the genes in this group are significantly more contributing than others. Similarly, in the second top contributing gene cluster, the lowest \( p \)-value, 0.23, belongs to the paired t-test between \( CXCL11 \) and \( IRF7 \), meaning that the genes in this group are also not statistically significantly different.

Instead, when we compare these two top gene clusters, we obtain a \( p \)-value of 0.012, meaning that the first gene cluster is significantly more contributing than the second gene cluster. For both classification schemes, the diagonal dark region for the MLN dataset is narrower than the other panels and the transition from the dark color to the light copper color is the sharpest. In agreement with our previous observations (compare Figure 2-8A-C), this suggests that the gene rankings in the MLN dataset are more statistically significant than in the other two datasets. We note that \( p \)-values of paired t-tests between consecutive single genes did not take sufficiently small values to show statistically significant difference among them. Instead, we were able to identify gene clusters that were statistically different compared to each other. mRNA measurements from more animals could lead to lower \( p \)-values, smaller gene clusters and more statistically significant gene rankings.
2.4.7 Polar plots provide a complete picture of the genes in the datasets

In the loading plots, we assign a vector to each gene from the origin to its location and study the correlation between genes using the cosine of the angle between their vectors, resulting in a matrix of size $88 \times 88$ (loading plots in the Appendix). This is possible because the columns of the score matrix are orthonormal and the top two classifier PCs provide an accurate and robust classification of the observations, and hence sufficiently approximate the dataset [128]. The angular correlation coefficients obtained this way do not necessarily match the pairwise correlation coefficients calculated using mRNA measurements in the dataset. Instead, they are calculated in the context of all other genes on planes that closely approximate the dataset. The average of 12 correlation coefficient matrices (one for each judge) for a given dataset and a classification scheme is shown in Figure 2-30, where each row or column shows the correlation coefficients between a specific gene and other genes. For each pair of genes, we calculated the standard deviation of the 12 correlation coefficients, resulting in 88 values for each gene. The mean of these values, indicative of the level of agreement between judges, is calculated for each gene and shown in the bar chart on the right hand side of each correlation matrix. Smaller values suggest higher degrees of agreement between judges on the correlation of that gene with other genes.

Figure 2-29. Quantifying significance of gene ranking: $p$-value heatmap of the paired t-tests of gene rankings in all datasets and for both classification schemes.
In our analysis, we perform paired t-tests of gene rankings in the spleen, MLN, and PBMC datasets. The results of the tests, $p$-values, range from 1 (black) to 0 (the light copper color). Lower $p$-values suggest a more statistically significant difference between the contribution of genes. On the bottom and left axes, genes are listed from the highest average rank (the left bottom corner) to the lowest as seen in Figure 2-8 and Figure 2-9. The clusters, colored alternately dark and light blue along the vertical axis, determine the genes that are significantly different from
genes in other clusters; the labels display the $p$-value of the paired t-tests between the cluster below the label and the cluster right to it.
For example in Figure 2-30A, the *judges* have the lowest degree of consensus about the correlation of *IL11* with other genes. For both classification schemes, the *judges* have a high degree of agreement on the gene correlations in the spleen dataset (Figure 2-30A, and D). Using linkage analysis (dendrograms), we identified 20 clusters comprising genes with approximately similar correlation patterns in the dataset (shown in different colors along the vertical axis). Interestingly, interferon-stimulated genes (*MxA, OAS1, OAS2*) always appear in the same group and in close proximity to type I interferon genes (*IFNα1* and *IFNβ*), suggesting correlated behavior during acute SIV infection.

To visualize the relative position of each gene compared to the other genes, we perform PCA on the average correlation coefficient matrix and construct the loading plot using the first two PCs scaled by the square root of their eigenvalues (Figure 2-31). Since the first two PCs capture more than 70% of the variance, they can create a plane that closely approximates the matrix, and hence the cosine of the angle between any two genes is approximately equal to the corresponding correlation coefficient in the matrix [128]. To validate this assumption, we calculated the angular correlation coefficients matrices from these plots, which provide a good approximation of the average correlation coefficient matrices with differences between some genes (compare Figure 2-30 and Figure 2-32). We measured the confidence on the angular position of a gene relative to others by calculating the mean-square-difference (MSD) between rows of the average correlation coefficient matrices in Figure 2-30 and their corresponding matrices in Figure 2-32. If the MSD of a gene takes small values, it suggests there is high confidence on the angular position of that gene in the loading plot.
Figure 2-30. Average correlation coefficient matrices in all datasets, for both classification schemes.
Dark blue and red colors represent positive and negative correlations, respectively.
Figure 2-31. The loading plots of average correlation coefficient matrices shown in Figure 2-30.
Figure 2-32. The correlation coefficient matrices obtained from the loading plots in Figure 2-31.
Polar plots summarize correlation information, MSD values and gene rankings in one place (Figure 2-33). The distance from the origin indicates the overall contribution of the genes in the dataset, obtained from Figure 2-8 and Figure 2-9. Therefore the high-ranking genes are located close to the perimeter while low ranking genes are located at the center. The angular position of genes is extracted from the loading plots constructed by the first two eigenvectors of the average correlation coefficient matrices (Figure 2-31). To make the comparisons easier, the clouds of genes are rotated such that CCL8, the top contributing gene, is located at zero degrees. The radial grid lines define the clusters obtained in Figure 2-29, each of which contains genes that are significantly more contributing than the genes in the lower neighboring cluster. Also, genes with the same color have similar patterns of correlation with other genes (the colors match the gene clusters shown in Figure 2-30). In Figure 2-33A, We plotted the expression profiles of representative genes from these clusters, showing the dynamic mRNA expression profiles as we move around the plot. Finally, the radius of each dot is linearly inversely proportional to the square root of MSD (rMSD), i.e. there is more confidence on the angular position of larger dots. The relationship between the dot size and the value of rMSD is shown on a scale at the bottom, where the largest and smallest circles correspond to rMSD = 0 and 9, respectively. We generally observe that the dots in the spleen dataset for classification based on time since infection have a larger size compared to other cases. This is because the first and second PCs capture more than 96% of the variance in the average correlation coefficient matrix (Figure 2-31A).
Figure 2-33. Correlations among genes simplified on a polar plot, illustrating MSD values and the ranking information.
Polar plots overview the information that can be obtained for any given gene from previous figures. For example, we observe that in the Spleen, MLN, and PBMC datasets, CCL8, MxA, CXCL10, CXCL11, OAS2 and OASI are located close to the perimeter in the top two clusters, meaning that they are all top contributing genes and their contribution is statistically significantly higher than that of other genes; they are clearly grouped in the same angular direction, suggesting strong correlations exist among them; they are all represented by large dots, implying high confidence on their locations relative to the other genes; and they are upregulated during the first days of SIV infection and their expression goes down after 4 days p.i. A schematic of the algorithm for obtaining polar plots is given in Figure 2-34.

Figure 2-34. Schematic of algorithm to plot Figure 2-33 (polar plots).

Note that we could not directly combine the information on the angular position of genes in the loading plots provided by the judges. This is because if a PC is multiplied by
-1, the new vector is still a principal component; however, all the relative positions of genes change in the loading plot. To avoid this problem, we converted the information on the angular position of genes to the correlation coefficients for each judge, took the average of the correlation coefficient matrices and converted it back using PCA to visualize positions of genes relative to each other.

2.4.8 Top contributing genes have approximately equal contributions to all tissues

Since genes contribute differently to each tissue, we measure the relative contribution of each gene to identify tissue-specific genes (see S6 Method). The results are shown in hexagonal plots (Figure 2-35), where genes in the center contribute equally to all tissues. The proximity of a gene to a vertex indicates that the gene contributes more to the tissue(s) noted at that vertex than to other tissues. The inner color of each dot represents the average contribution of the gene, whereas the outer color represents the highest contribution (lowest rank) of that gene. The common genes are seen close to the center of the hexagon, while the tissue-specific genes are located close to the vertices and near the edges. The congested region in the center of the hexagon houses most of the genes. To see this region more clearly, it is amplified on the right-hand plot. For both classification schemes, we observe the top contributing genes such as CCL8, MxA, CXCL10, CXCL11, OAS2, and OAS1 lie in the center of the plot with approximately the same blue color for the inner and outer circles, indicating their equal contribution to all tissues (Figure 2-35). This suggests that type I interferon responses are quite similar in the three compartments and that these genes could be used as biomarkers to be measured in PBMCs instead of spleen and MLNs during acute SIV infection. This can be tested by classifying the observations using the mRNA measurements of these genes in PBMCs and by evaluating
whether that classification is as accurate as the classifications using measurements in spleen or MLN. To this end, we built decision trees using the top seven highly contributing genes and chose the sub-trees with the lowest cross validation error rates in all tissues and for both classification schemes (Table 2-4). For time since infection and SIV RNA in plasma, the classification rates in the PBMC dataset are 87.5% and 83.3%, greater than or equal to the classification rates in spleen and MLN. This suggests that an analysis of gene expression in the more accessible PBMC can be used as a surrogate to understand the immunological events happening in the less accessible spleen and lymph nodes during acute SIV infection. However, each tissue has unique expression profiles, e.g. *XCL1*, a relatively high-contributing gene, contributes highly to spleen and MLN compared to PBMC, and hence analysis of selected top contributing tissue-specific genes could greatly inform about the mechanisms related to SIV infection in those tissues.

**Table 2-4.** Classification results using the top seven highly contributing genes.

<table>
<thead>
<tr>
<th>Time of Infection</th>
<th>Plasma SIV RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spleen</td>
<td>MLN</td>
</tr>
<tr>
<td>87.5%</td>
<td>87.5%</td>
</tr>
</tbody>
</table>
Figure 2-35. Tissue-specificity of genes: relative contribution genes to each tissue.
In each hexagonal plot, three main vertices represent Spleen, MLN, and PBMC. Genes close to one of these vertices show a strong contribution to the corresponding tissue. Genes at the center contribute approximately equally to each tissue. The inner color of each gene shows its overall rank in all tissues (Figure 2-8D and Figure 2-9E), while the outer color represents the minimum of each gene’s three ranks in the tissues.
2.5 Discussion and conclusions

Acute HIV infection is characterized by an exponential increase in plasma viremia with subsequent viral dissemination to lymphoid and non-lymphoid organs. As the innate immune system responds to viral replication, the expression of inflammatory cytokines in the plasma also rapidly increases, leading to a positive feedback where newly-induced activated cells allow for more viral infection. This hypercytokinemia is known as "cytokine storm", and it is not unique to HIV [138]. Other pathogens may also cause strong immune responses that lead to tissue damage, organ dysfunction and death. For instance, severe acute lung injury with respiratory failure can be observed after SARS-CoV and influenza infections, and are caused by cytokine storms in the lung alveoli and peripheral blood [138]. In HIV infection, this inflammatory response is not fatal but may cause irreparable impairment to the immune system, leading to massive CD4+ T cell depletion and chronic immune activation [100]. A similar cytokine storm is observed during acute infection in the brain of SIV-infected macaques [101], indicating that even immune-privileged organs are not shielded from the damage that such responses may cause during HIV and SIV infection. Understanding the pathways and components of these immunological events is essential for the development of therapeutic strategies aimed at reducing their harmful effects. Similar acute phase studies cannot be performed in HIV-infected patients for several reasons, including lack of precision regarding the exact time of transmission, limited access to organ biopsies, and HIV genotypical diversity [139]. Therefore, SIV macaque models represent a viable and efficient alternative to human studies, despite the biological differences between HIV and SIV [102,140].
In this study we used an accelerated and consistent macaque model of AIDS and HIV-associated neurocognitive disorders to analyze the expression of immune-related genes in three different lymphoid compartments during acute SIV infection. mRNA levels were quantitated by Nanostring, a novel technology that allows for the measurement of a large number of transcripts without reverse transcription or DNA amplification. Fluorescent bar-coded probes specifically hybridize with mRNAs that are then counted by a powerful scanner. The technique involves little sample manipulation and generates results faster, presented in a simple spreadsheet format. The Nanostring panel in this study was designed to understand how immune responses are longitudinally developed in different organs or cells during SIV infection. The panel includes genes that are commonly analyzed during inflammation and viral infection, and has been used to evaluate the longitudinal level variances in individual cytokines during SIV infection. Thus, the panel gives us insight into the host response to acute infection.

Studies that attempt to analyze changes of gene expressions over time or only examine bivariate correlations between two genes or a gene and a clinical parameter such as SIV RNA in plasma can result in limited (and often flawed) conclusions. This can be due to several reasons including lack of prior information on how changes in gene expressions affect the immune response, noisy measurements, and contribution of many genes, each of which has a minor impact but when considered together can create a significant response. In addition, after animals are infected by SIV, the changes in gene expressions are presumably caused by SIV infection. One can expect the mRNA measurements, regardless of their biological functions, to be correlated with SIV clinical parameters. This suggests that drawing conclusions based on only bivariate correlations
can be misleading. Therefore, multivariate analysis techniques are more appropriate tools to study a set of genes simultaneously.

Here, we introduced a novel multiplexed component analysis (MCA) technique to simultaneously analyze mRNA measurements under different assumptions for how the gene expression changes affect the immune response during acute SIV infection. In this method, mRNA measurements were studied by 12 judges, each of which consists of three successive modules: 1) transformation (Log2 or Orig), 2) preprocessing (MC, UV, or CV) and 3) multivariate analysis (PCA or PLS). The preprocessing module aims to emphasize specific features of the dataset, e.g. the MC normalization method emphasizes biological responses in which the immune response is affected by the genes with the highest absolute variations in expression across animals, whereas the CV normalization method puts emphasis on responses in which the relative changes in gene expressions are more important. Note that other normalizations, transformations and multivariate techniques could be combined to create more judges; however, the goal in this method is to have unique judges that observe the data from distinct viewpoints and hence the techniques that have similar effects on the data should not be included in the same analysis.

Each of the twelve judges provides a distinct set of uncorrelated principal components (PCs), capturing the directions in the data with maximum variance. From each set, we select two PCs that provide the most accurate and robust classification of the data in each of the classification schemes: time since infection and SIV RNA in plasma. These selection criteria result in PCs with gene loadings that robustly classify the animals at different stages of the disease. Our hypothesis is that highly loaded genes, which contribute the most to the classification, are those whose levels of expression are most
profoundly affected during acute SIV infection and therefore warrant further study. While the MCA technique does not by itself provide mechanistic insight into how these genes function in the immune system, it provides an impartial platform to compare genes and highlight those with the highest level of contribution during acute SIV infection, globally in the immune system or locally in specific tissues; and it can further be combined with mechanistic information about the immune response dependence on specific gene expression changes. Also, the MCA method can be used in genome-wide studies, where the number of genes is significantly higher. The transformation and normalization modules do not change in such applications. Also, the PCA and PLS methods are essentially dimension reduction algorithms and hence can be readily applied to large datasets to identify genes with significant contributions. One should note that the sets of significant genes selected by individual judges might be different when the number of genes is high, and hence extra attention should be paid when the gene ranking results from the judges are combined. In our datasets, the top eight PCs were enough to capture more than 76% of the variation within the dataset. When the number of genes increases, more than eight PCs may be needed to capture sufficient variance within the dataset.

We can combine the opinions of all the judges to sort genes based on their overall rank. As discussed above, the judges’ agreement on the gene rankings differs for each gene. When there is a high level of agreement among the judges for a gene, it suggests that the gene is accurately ranked, regardless of how the changes in gene expressions affect the immune response. On the other hand, there are genes that receive high ranks from some judges and low ranks from the others. This suggests that the specific way that gene expression changes are translated to the immune response matters, and that these
genes can hold less or more significance, which in turn generates new hypotheses for future experiments. The results also demonstrate differential ranking of some genes according to specific lymphoid compartments. IFNα1, for instance, is highly ranked in MLN but not in PBMCs or spleen. We hypothesize that this is due to the highly abundant population of IFNα-producing dendritic cells, which are responsible for antigen presentation and T cell activation in lymph nodes [141]. Similarly, CD68, a bona fide marker for macrophage activation ranks higher in spleen, an organ rich in macrophages [142]. An important point to make is that all three tissues here analyzed comprise mobile cell types, and therefore are subject to numerical changes in cell subpopulations during infection. Thus, changes in gene expressions do not reflect only transcription modulation, but also cell trafficking. Interestingly, three of the highest-ranking genes, CCL8, CXCL10 and CXCL11, are chemoattractants of cells susceptible to SIV infection (CCL8 for monocytes and CXCL10 and CXCL11 for activated lymphocytes) [143,144], and may be directly responsible for the trafficking of SIV-infected cells to organs and subsequent establishment of viral reservoirs during acute infection. Similar multi-gene analyses of cell type-specific transcripts may lead to methods for the precise quantitation of leukocytes in lymphoid compartments, and their contribution to inflammatory responses during pathological conditions.

One of the main advantages of our methodology is to provide a diverse set of perspectives on the evaluation of cellular and molecular events during infection in different tissues. For instance, gene-ranking analysis informs about the overall aspects of the immune response, but also identifies signature genes that are singularly relevant to cellular mechanisms in specific lymphoid compartments. In this report, similar high
ranking genes in spleen, MLN and PBMC reveal a systemic and concomitant type I interferon response during acute SIV infection, despite the diversity in cell populations in each tissue and the particular pathways by which cell phenotypes respond to viral infection. Therefore, the synchronous changes in gene expressions appear to be driven mostly by the crosstalk between cells and cytokines that are constantly trafficking through tissues than by viral replication per se [134].

Nonetheless, ranking gives somewhat limited information on how genes relate to each other and how transcription is longitudinally modulated in each tissue. Therefore, by combining the information on the angular position of genes provided by all the judges and depicting the results in polar plots (Figure 2-33), it is possible to identify genes with similar regulation patterns and evaluate whether these same genes are equally regulated in other lymphoid compartments. As an example, all putative interferon-stimulated genes (ISG) are grouped together in all three compartments, indicating a common regulatory process. On the other hand, based on the spleen results alone, it could be suggested that the transcription activator STAT5A is directly involved in the regulation of IL4, but this is not observed in the other two tissues, suggesting either tissue-specific regulation or an aliasing effect, and these are computationally-derived hypotheses for further study.

Finally, this methodology allows for the combination of results from three related but independent analyses into one cogent hexagonal plot (Figure 2-35), displaying the relative contribution of each gene to the overall changes in each compartment. This powerful visualization tool can be used to identify genes that uniquely and significantly contribute to immune responses in specific tissues, and also genes that could be selected as general inflammatory markers to be investigated during acute infection. The model
suggests that evaluation of a small selected panel of ISGs and chemokines in PBMCs may be sufficient to assess systemic inflammatory responses triggered by viral infection in secondary lymphoid tissues. On the other hand, \textit{IL10} and \textit{XCL1} appear to be highly significant in spleen and MLN but not in PBMCs, and therefore examining the levels of these cytokines in the blood may not provide accurate information regarding immunological events in lymphoid organs. The expression profiles of these genes in spleen and MLNs are strongly correlated, while they have little to no correlation with the expression profiles in PBMCs.

Note that mRNA samples in our study were isolated from different animals euthanized at 4, 7, 14, and 21 days post infection. Therefore, the obtained measurements at various time points do not constitute a longitudinal study. For example, the gene expression data points at days 4 and 7 are not inherently connected, but instead represent samples from populations of animals infected with SIV for 4 and 7 days, respectively. Therefore, the data at day 4 cannot be readily used to predict the gene expressions at day 7. In addition, there is a fundamentally different relationship between the input variables (mRNA measurements) and each of the two classification schemes. While time since infection as an output variable is intrinsically independent of the mRNA measurements, SIV RNA in plasma is completely dependent on the changes in gene expressions, as both inflammatory response genes and SIV are constantly in direct or indirect interactions in the immune system and hence cause changes in mRNA counts and SIV RNA in plasma. This, in addition to other factors, may partially explain why classification based on time since infection is more accurate than classification based on SIV RNA in plasma.
For most viral infections, the acute phase is a time of drastic physiological and immunological changes, especially at the beginning of adaptive immune responses. Further similar studies performed in samples collected at later time points, when infection is already established, would help to evaluate the relationship between cytokine expression and viral replication.
2.6 Appendix. score plots, loading plots and results of classification

Score plots provided by the judges are used to cluster observations into separate groups using a specific classification scheme (Figure A1 - Figure A6). For each judge, we study 28 score plots generated by all the combinations of two of the top eight PCs. In each score plot, we perform centroid-based classification and leave-one-out cross validation (LOOCV) to obtain classification and LOOCV rates. The classification and LOOCV rates are indicative of the accuracy and the robustness of the classification on a given score plot, respectively. The sum of these two rates is calculated and the PCs representing the plot with the highest sum are chosen as the top two classifier PCs for that judge. The classification and LOOCV rates (in parentheses) are written in the right bottom corner of each subplot. The top two classifier PCs are shown above the rates (Figure A1 - Figure A6). After choosing the top two classifier PCs for each judge, loading plots are constructed by the chosen PCs (Figure A7 - Figure A12).
Figure A1. Score plots and results of classification based on time since infection in the spleen dataset.
Figure A2. Score plots and results of classification based on time since infection in the MLN dataset.
Figure A3. Score plots and results of classification based on time since infection in the PBMC dataset
Figure A4. Score plots and results of classification based on SIV RNA in plasma in the spleen dataset
Figure A5. Score plots and results of classification based on SIV RNA in plasma in the MLN dataset.
Figure A6. Score plots and results of classification based on SIV RNA in plasma in the PBMC dataset
Figure A7. Loading plots constructed by the top two classifier PCs chosen by classification based on time since infection in the spleen dataset
Figure A8. Loading plots constructed by the top two classifier PCs chosen by classification based on time since infection in the MLN dataset
Figure A9. Loading plots constructed by the top two classifier PCs chosen by classification based on time since infection in the PBMC dataset.
Figure A10. Loading plots constructed by the top two classifier PCs chosen by classification based on SIV RNA in plasma in the spleen dataset.
Figure A11. Loading plots constructed by the top two classifier PCs chosen by classification based on SIV RNA in plasma in the MLN dataset
Figure A12. Loading plots constructed by the top two classifier PCs chosen by classification based on SIV RNA in plasma in the PBMC dataset.
Chapter 3. Multi-Scale Modeling of HIV Infection in vitro and APOBEC3G-Based Anti-Retroviral Therapy
3.1 Summary

The human APOBEC3G is an innate restriction factor that, in the absence of Vif, restricts HIV-1 replication by inducing excessive deamination of cytidine residues in nascent reverse transcripts and inhibiting reverse transcription and integration. To shed light on impact of A3G-Vif interactions on HIV replication, we developed a multi-scale computational system consisting of intracellular (single-cell), cellular and extracellular (multicellular) events by using ordinary differential equations. The single-cell model describes molecular-level events within individual cells (such as production and degradation of host and viral proteins, and assembly and release of new virions), whereas the multicellular model describes the viral dynamics and multiple cycles of infection within a population of cells. We estimated the model parameters either directly from previously published experimental data or by running simulations to find the optimum values. We validated our integrated model by reproducing the results of in vitro T cell culture experiments. Crucially, both downstream effects of A3G (hypermutation and reduction of viral burst size) were necessary to replicate the experimental results in silico. We also used the model to study anti-HIV capability of several possible therapeutic strategies including: an antibody to Vif; upregulation of A3G; and mutated forms of A3G. According to our simulations, A3G with a mutated Vif binding site is predicted to be significantly more effective than other molecules at the same dose. Ultimately, we performed sensitivity analysis to identify important model parameters. The results showed that the timing of particle formation and virus release had the highest impacts on HIV replication. The model also predicted that the degradation of A3G by Vif is not a crucial step in HIV pathogenesis.
3.2 Introduction

Mathematical models have proven valuable in understanding the dynamics of HIV-1 infection in vivo [145]. In most existing HIV infection models, subcellular events such as viral genome replication and integration, production of viral proteins, and release of new virions are often not explicitly reflected [146-152]. Instead, these extracellular models consider several cycles of infection where a population of cells can be infected by viruses and the effects of drug therapies on the number of viruses are studied. By contrast, intracellular models assuming only a single cycle of infection have been limited to the study of the kinetics of virus and host proteins and their interactions to understand the dynamics of viral replication inside the cells [153,154]. Both types of modeling give insights into how HIV disease progresses in the body, however, combining intracellular and extracellular models would greatly enhance our understanding in this area [155].

In our previous work [156], we used a model of a single cell that could undergo multiple re-infection as a surrogate for multicellular infection, to capture both intracellular and extracellular properties of HIV infection. In the present study, we have developed a multi-scale system integrating intracellular, cellular and extracellular processes. This integrated model explicitly includes concepts such as burst size (the number of viruses released by a cell), proliferation rate of cells, cell life cycle, virus clearance, and intracellular delays in viral formation and release from cells, which were not explicitly described in the previous model. The integrated model is used to simulate in vitro T cell HIV infection experiments.
Figure 3-1. HIV virion and its life cycle. (A) HIV particles are surrounded by fatty materials known as the viral envelope. The matrix formed from p17 protein is another layer underneath the viral envelope. The particles also contain two exact copies of RNA strands as well as three essential enzymes required for replication: reverse transcriptase, integrase and protease. (B) Mechanism of HIV infection including viral entry, genome integration, production and release of new viral particles is shown. If the released viruses carry A3G, they are denoted A3G(+) viruses, otherwise they are denoted A3G(-). When A3G(+) viruses infect the next cell, the packaged A3G has several activities such as hypermutating the minus strand of viral DNA, and inhibiting various steps of reverse transcription and integration. “Null” symbols inside the cell represent degradation of Vif, A3G, and A3G-Vif complex.

The intracellular (single-cell) model includes interactions between Vif, virus RNA and human A3G. Experimental data are used to establish system parameters such as degradation rate constants of proteins, life-span of infected cells, and viral generation time. The intracellular model sheds light on how changes in the intracellular parameters affect the production and release of new HIV viruses. The single-cell model results are integrated into a multicellular model to simulate T cell culture experiments. We estimate certain parameters such as viral burst size, HIV infectivity rate, and virus clearance rate using experimental data, and model predictions are verified using previously published experimental results. Biologically relevant levels of host and virus proteins in experiments are estimated using our multi-scale system. We monitor how the population of cells acts in response to virus infection. Several drugs targeting A3G and Vif pathways...
are studied to compare their efficacy at different doses. We also estimate drug efficacy under non-ideal conditions, such as when it is available to only a specific fraction of cells in the whole population, or delivered at later times following infection.

3.3 Methods

3.3.1 HIV biology and system model

HIV particles are surrounded by a fatty membrane known as the viral envelope. There is another layer underneath the viral envelope called matrix, which is formed from p17 protein. HIV has three essential enzymes required for replication: reverse transcriptase, integrase and protease. These enzymes along with two exact copies of RNA strands are packaged in the viral core or capsid, which also encapsulates A3G. The viral core is made from the protein p24. A generic structure of HIV virus is shown in Figure 3-1A.

Figure 3-1B shows the schematic model of HIV infection that is used to develop the computational model, capturing both intracellular and extracellular information. Each cycle of infection begins with a HIV virus attacking a healthy normal cell. After the virus entry into cells, the HIV genome is reverse transcribed into cDNA [157]. Some evidence suggests that HIV capsid remains intact during reverse transcription and that uncoating occurs at the nuclear pore upon completion of reverse transcription, reviewed in [158]. The resulting double-stranded DNA enters the nucleus along with the viral integrase, which splices the HIV DNA into the human genome. The integrated viral DNA, called provirus, is then transcribed into messenger RNA used as a blueprint for making new HIV proteins and enzymes. Some of the viral RNA remains as full-length RNA copies, to be incorporated as viral genetic material for new virions. We model the mechanisms from
virus entry to viral protein production using relevant kinetics and intracellular delay parameters. The focus of this study is the interaction between Vif and A3G, and their productions are explicitly included in our model.

HIV enzymes, structural proteins, and full-length RNA molecules are assembled into virions at the cell membrane. Human T cells can produce A3G as an intrinsic defense mechanism. This protein binds the viral RNA and gets encapsulated into the viral capsids while they are still inside the cell. Shortly after viral assembly, viruses get released from the cells and they are ready to infect new cells. In our model, if the released viruses carry A3G, they are denoted A3G(+) viruses, otherwise they are denoted A3G(-). The encapsulated A3G is assumed to not have effects on viral entry. This is because the entry process involves the binding of CD4 and chemokine coreceptors on the T cell surface to gp41 and gp120 on the viral envelope whereas A3G is encapsulated inside the capsid and doesn’t interact with the proteins on the viral envelope. When A3G(+) viruses infect the next cell, the packaged A3G can have various anti-retroviral activities. In this chapter, we focus on two downstream effects of A3G; 1) hypermutation in the minus strand of viral DNA; and 2) inhibition of viral cDNA production. This means that even though A3G(+) viruses can infect cells with similar rate of infectivity to A3G(-) viruses, infected cells produce fewer virions. HIV has evolved to combat A3G with Vif. The Vif protein binds A3G and facilitates its polyubiquitylation, and therefore increases its degradation rate. This Vif-induced degradation, and basal rates of degradation of both Vif and A3G, are included in the model. We assume A3G doesn’t affect replication of viruses in the producer cell. This has been observed in [44-46] and can be explained assuming that reverse transcription occurs inside the capsid and A3G doesn’t have access to transcripts.
3.3.2 Mathematical model development for intracellular interactions ("single-cell model")

Our model includes both proteins and virions in a generic human T cell. There is a differential equation for each entity, which describes its production, degradation, and interactions with other entities.

The A3G protein can be produced, degraded, and incorporated into progeny viruses. It also binds to and dissociates from Vif.

\[
\frac{d[A3G]}{dt} = P_{A3G} - k_{d,A3G}[A3G] - k_{on}[Vif][A3G] + k_{off}[A3G,Vif] - k_{A3G,HIV}s_{A3G} \{HIV(-)\}[A3G],
\]

where \( P_{A3G} \) is the production rate of A3G, \( k_{d,A3G} \) is the degradation coefficient of A3G, \( k_{on} \) and \( k_{off} \) are the binding and dissociation constants of the A3G-Vif complex, and \( k_{A3G,HIV} \) is the rate constant for A3G incorporation into A3G(-) HIV viruses, denoted by \( HIV(-) \). The stoichiometry of A3G proteins incorporated into virions is \( s_{A3G} \). Similarly, Vif concentration is governed by

\[
\frac{d[Vif]}{dt} = P_{Vif} - k_{d,Vif}[Vif] - k_{on}[Vif][A3G] + k_{off}[A3G,Vif],
\]

where \( P_{Vif} \) and \( k_{d,Vif} \) are the production rate and the degradation coefficient of Vif, respectively. The A3G-Vif complex can be formed from Vif binding to A3G or it can degrade.

\[
\frac{d[A3G,Vif]}{dt} = -k_{d,A3G,Vif}[A3G,Vif] + k_{on}[Vif][A3G] - k_{off}[A3G,Vif].
\]

In (3), the degradation coefficient of A3G-Vif complex is shown by \( k_{d,A3G,Vif} \). The number of HIV virions inside the cell is also modeled by
Viruses are produced at a rate of $P_{HIV}$ and budded off from the cell by a rate of $k_{rel}$. The A3G protein can get encapsulated into A3G(-) viruses and convert them to A3G(+) viruses. The number of intracellular A3G(+) viruses is governed by

$$\frac{d\{HIV^{(+)}\}}{dt} = k_{A3G,HIV}A3G - k_{rel}HIV^{(-)};$$

where $HIV^{(+)}$ refers to A3G(+) viruses. Finally, the release of newly-made HIV viruses is described by the following equations.

$$M^{(-)}(t) = \frac{d\{HIV_{rel}^{(-)}\}}{dt} = k_{rel}HIV_{rel}^{(-)};$$

$$M^{(+)}(t) = \frac{d\{HIV_{rel}^{(+)}\}}{dt} = k_{rel}HIV_{rel}^{(+)}.$$
In equation (6), $HIV_{rel,(\cdot)}$ and $HIV_{rel,(\pm)}$ represent released A3G(-) and A3G(+) viruses, respectively. $M_{\cdot}(t)$ and $M_{\pm}(t)$, the number of released A3G(-) and A3G(+) viruses at time \( t \), will be used later in the extracellular model. In the model, proteins are quantified in units of molar concentration, whereas viruses are quantified as discrete numbers of viral particles. Figure 3-2 shows the time evolution of total number of A3G(-) and A3G(+) viruses produced in a single cell after infection and released from it to the extracellular environment.

### 3.3.3 Mathematical model development for extracellular events (“multicellular model”)

Our multicellular model describes an extracellular pool of HIV viruses infecting a population of T cells, specifically, in cell culture. This model includes cellular and extracellular properties including the production rate of T cells, rate of infection by HIV viruses, variations in levels of A3G(-) and A3G(+) viruses, and burst size (which is defined as the average number of HIV viruses made by an infected cell). There is a strong link between the intracellular and multicellular models through the burst size and the release distribution of A3G(-) and A3G(+) viruses over time. The multicellular model can be described by a set of equations and constraints. In our model, we define $T_0$ as the initial number of “Normal” T cells. Each cell lives in the normal state until a HIV virus infects it. “Infected(+)” and “Infected(-)” states correspond to cells that have been infected by A3G(+) and A3G(-) viruses, respectively; however, infected(+) cells produce fewer viruses than infected(-) cells. It is assumed that there is no hyper-infection, that is, after a virus attacks and enters a healthy cell, the cell becomes infected, CD4 is down-regulated [159], and no more viruses attack it. Cells in infected(+) and infected(-) states
become “Productive(+)” and “Productive(-)” after $\Delta t = \tau_{\text{prod}}$ post infection, respectively and begin releasing viruses into the extracellular environment. The release continues until $\Delta t = \tau_{\text{dead}}$ after infection, when the cell dies and it is marked “Dead”. A schematic diagram of cell states is shown in Figure 3-3A.

Figure 3-3B is a snapshot of the multicellular model at a specific time, showing cells of different post-infection ages at different states. In our simulations, we keep time of infection for each cell in the multicellular model. This is represented by early and late infected cells in the set of infected cells and by early and late productive cells in the set of productive cells.

The number of healthy and infected cells in our model is governed by the following equations.

\[
\frac{dT(t)}{dt} = k_p T(t) - k_{\text{inf}}(V_{(-)}(t) + V_{(+)})(t)T(t),
\]

and

\[
\frac{dT_{\text{inf}(-)}(t)}{dt} = k_{\text{inf}}V_{(-)}(t)T(t) - k_{\text{inf}}V_{(-)}(t - \tau_{\text{prod}})T(t - \tau_{\text{prod}}),
\]

\[
\frac{dT_{\text{inf}(+))(t)}{dt} = k_{\text{inf}}V_{(+)}(t)T(t) - k_{\text{inf}}V_{(+))(t - \tau_{\text{prod}})T(t - \tau_{\text{prod}}),
\]

where $T$ is the number of healthy cells and the rate of infection is defined by $k_{\text{inf}}$. The proliferation rate of healthy cells is represented by $k_p = \ln(2)/\tau_{T,2}$ where $\tau_{T,2}$ is the cell doubling time. A3G(-) and A3G(+) viruses are denoted by $V_{(-)}$ and $V_{(+)},$ respectively, and $t$ represents time post inoculation of the T cell culture. Note that $V_{(-)}(t)$, $V_{(+))(t)}$, and $T(t)$ are all zero for $t < 0$. In (7), the number of normal T cells increases by cell proliferation and decreases as cells get infected.
In (8a) and (8b), \(T_{\text{inf},(-)}\) and \(T_{\text{inf},(+)}\) represent the number of cells infected by A3G(-) and A3G(+) viruses, respectively. As mentioned earlier, each infected cell begins releasing new viruses after a time \(t_{\text{prod}}\) post infection. At this point, the infected cells become actively productive, represented by \(T_{\text{prod},(-)}\) and \(T_{\text{prod},(+)}\). Note that in (8a), there are two mathematical terms determining the rate of change for \(T_{\text{inf},(-)}\). The first term, \(K_{\text{inf}}V_{(-)}(t)T(t)\), represents the number of cells that become infected by A3G(-) viruses at time \(t\) and enter the set of infected cells, whereas the second term, \(K_{\text{inf}}V_{(-)}(t - t_{\text{prod}})T(t - t_{\text{prod}})\), represents the number of cells that were infected at time \(t - t_{\text{prod}}\), i.e., they are productive at time \(t\) and leave the set of infected cells to join the set of productive cells (Figure 3-3B). The mathematical terms in (8b) are the same as those in (8a), except that they deal with infected(+) cells. The number of productive cells is described by

\[
\frac{dT_{\text{prod},(-)}(t)}{dt} = k_{\text{inf}}V_{(-)}(t - t_{\text{prod}})T(t - t_{\text{prod}}) - k_{\text{inf}}V_{(-)}(t - t_{\text{dead}})T(t - t_{\text{dead}}), \quad (9a)
\]

\[
\frac{dT_{\text{prod},(+)}(t)}{dt} = k_{\text{inf}}V_{(+)}(t - t_{\text{prod}})T(t - t_{\text{prod}}) - k_{\text{inf}}V_{(+)}(t - t_{\text{dead}})T(t - t_{\text{dead}}). \quad (9b)
\]

In (9a), the first term, \(K_{\text{inf}}V_{(-)}(t - t_{\text{prod}})T(t - t_{\text{prod}})\), represents the number of cells that become productive at time \(t\) and enter the set of productive cells. Once the cells are infected with HIV, they have an average life span of \(t_{\text{dead}}\). This means that productive cells release HIV viruses from \(t_{\text{prod}}\) until their death at \(t_{\text{dead}}\). The second term in (9a), \(K_{\text{inf}}V_{(-)}(t - t_{\text{dead}})T(t - t_{\text{dead}})\), represents the number of cells that were infected at time \(t - t_{\text{dead}}\), i.e., they are dead at time \(t\) and leave the set of productive cells (Figure 3-3B). The mathematical terms in (9b) are the same as those in (9a), except that they describe productive(+) cells. The number of dead cells is represented by \(T_{\text{dead}}\) and governed by
Along with the equations for different cell states, the model tracks extracellular viruses.

\[
\frac{dV_{(-)}(t)}{dt} = pT_{\text{prod},(-)}(t)M_{\text{avg}} + pcT_{\text{prod},(+)}(t)M_{\text{avg}} - k_{\text{inf}}V_{(-)}(t)T(t) - k_{v}V_{(-)}(t), \tag{11a}
\]

\[
\frac{dV_{(+)}(t)}{dt} = (1-p)T_{\text{prod},(-)}(t)M_{\text{avg}} + (1-p)cT_{\text{prod},(+)}(t)M_{\text{avg}} - k_{\text{inf}}V_{(+)}(t)T(t) - k_{v}V_{(+)}(t). \tag{11b}
\]

The average number of viruses released from a productive(-) cell is \( M_{\text{avg}} = B/(t_{\text{dead}} - t_{\text{prod}}) \) where \( B \) is the viral burst size. The percentage of released viruses that do not contain A3G is denoted by \( p \). The encapsulated A3G in the HIV virus has anti-viral activities in the target cell. This results in that a productive(+) cell produces fewer viruses than a productive(-) cell. The reduction in burst size of productive(+) cells is denoted by \( c \). Both \( p \) and \( c \) take values between 0 and 1. In (11a), the first and second terms refer to the number of A3G(-) viruses being produced from productive(-) and productive(+) cells, whereas the third term represents the number of viruses that are infecting cells. The last term shows the number of viruses that are being cleared from the culture. The clearance rate is represented by \( k_{v} = \ln(2)/t_{v,1/2} \) where \( t_{v,1/2} \) is HIV half-life in vitro. Mathematically similar terms describe the number of A3G(+) viruses in (11b).

Parameters \( p \) and \( c \) play important roles in our simulation. Since \( p \) has a direct effect on the shape of HIV replication curves (described later), we call this parameter HIV replicative potential. The value of \( p \) inversely correlate with intracellular A3G getting encapsulated in newly made viral particles. It is desirable for both \( p \) and \( c \) to have values as close as possible to zero to efficiently stop HIV replication. Nominal values of \( p \) and \( c \) are shown in Table 3-1 for different types of viruses and cells. Note that \( p \) is a property of

\[
\frac{dT_{\text{dead}}(t)}{dt} = \frac{dT_{\text{dead},(-)}(t)}{dt} + \frac{dT_{\text{dead},(+)}(t)}{dt} = k_{\text{inf}}V_{(-)}(t - t_{\text{dead}})T(t - t_{\text{dead}}) + k_{\text{inf}}V_{(+)}(t - t_{\text{dead}})T(t - t_{\text{dead}}). \tag{10}
\]
cells whereas $c$ is a property of viruses. To compute the number of A3G(-) and A3G(+) viruses in (11a) and (11b), we assumed that the release rate of viruses from a productive cell over period of $[t_{\text{prod}}, t_{\text{dead}}]$ is constant and also $p$, the HIV replicative potential remains constant during this period.

Figure 3-3. A schematic diagram showing state of the cells and a snapshot of the multicellular model.

(A) Each cell lives in the “Normal” state until a HIV virus infects it. An “Infected” cell doesn’t release new virions until a certain time point post infection, denoted $t_{\text{prod}}$. At this time point, the cell becomes “Productive” and begins releasing viruses into the extracellular environment until it dies at $t_{\text{dead}}$, when it is marked as “Dead”. The “Infected (+)” and “Infected(-)” states correspond to cells that have been infected by A3G(+) and A3G(-) viruses, respectively. The same concept applies to “Productive(+)” and “Productive(-)” cells. (B) The time of infection is known for each cell in our multicellular model. A snapshot of the multicellular model shows cells with different post-infection ages in the sets of infected and productive cells. Normal cells become infected and enter the set of infected cells as early-infected cells. The late-infected cells become productive and leave the set of infected cells to join the set of productive cells where they are shown as early-productive cells. Finally, late-productive cells die, exit the set of productive cells, and get marked as dead.
Table 3-1. Values of $p$ and $c$ for different cases of viruses and cells.

<table>
<thead>
<tr>
<th>Target cells</th>
<th>A3G(-) viruses</th>
<th>A3G(+) viruses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3G-expressing cells</td>
<td>$p = \text{Low, } c \approx 1$</td>
<td>$p = \text{Low, } c \approx \text{Low}$</td>
</tr>
<tr>
<td></td>
<td>Output: A3G(-) and A3G(+) viruses</td>
<td>Output: A3G(-) and A3G(+) viruses</td>
</tr>
<tr>
<td>Non A3G-expressing cells</td>
<td>$p \approx 1, c \approx 1$</td>
<td>$p \approx 1, c \approx \text{Low}$</td>
</tr>
<tr>
<td></td>
<td>Output: only A3G(-) viruses</td>
<td>Output: only A3G(-) viruses</td>
</tr>
</tbody>
</table>

As we will see in the results section, the release rate of viruses from a productive cell is not constant at all times. In fact, virus release begins at $t_{\text{prod}}$ post infection, increases for 8 hours and remains constant until the cell dies. Also the ratio of released A3G(-) viruses to total released viruses is not constant during the virus release period and changes over time as it can be seen in Figure 3-2. Therefore we use the following equations instead of (11a) and (11b) to accurately compute the number of viruses without having any assumptions on the release of viruses.

\[
\frac{dV_{(-)}(t)}{dt} = \sum_{i=1}^{T_{\text{prod},(-)}(t)} M_{(-)}(t - t_i) + \sum_{i=1}^{T_{\text{prod},(-)}(t)} cM_{(-)}(t - t_i) - k_{\text{inf}}V_{(-)}(t)T(t) - k_pV_{(-)}(t), \quad (12a)
\]

\[
\frac{dV_{(+)}(t)}{dt} = \sum_{i=1}^{T_{\text{prod},(+)}(t)} M_{(+)}(t - t_i) + \sum_{i=1}^{T_{\text{prod},(+)}(t)} cM_{(+)}(t - t_i) - k_{\text{inf}}V_{(+)}(t)T(t) - k_pV_{(+)}(t). \quad (12b)
\]

$M_{(-)}(t)$ and $M_{(+)}(t)$ are the number of released A3G(-) and A3G(+) viruses from a single cell at time $t$ after infection, defined in (6a) and (6b). The first term in (12a) is a summation over all the A3G(-) viruses that are being released at time $t$ from all the cells in the set of productive(-) cells. The time of infection for the $i$th cell in the set is denoted by $t_i$. The second term in (12a) is similar to the first term, but deals with the A3G(-)
viruses released from the productive(+) cells. Similar mathematical expressions are used for computing the number of A3G(+) viruses in (12b).

Viruses in cell culture can become non-infectious or dead after some time. The number of cleared viruses from the culture is given by

\[
\frac{dV_{\text{dead}}(t)}{dt} = k_r (V_{(-)}(t) + V_{(+)}(t)).
\]

(13)

### 3.3.4 Integration of single-cell model results into the multicellular model

As noted above, the single cell and multicellular models are linked through the burst size and the release distribution of A3G(-) and A3G(+) viruses over time. In this chapter, we use the following two methods to establish this link.

In the first method, the multicellular model assumes that release rate of viruses from a productive cell over period of \([t_{\text{prod}}, t_{\text{dead}}]\) is constant and also that the ratio of released A3G(-) viruses to total released viruses at each time point during virus release is constant and equal to \(p\). With these assumptions, \(V_{(-)}(t)\) and \(V_{(+)}(t)\) can be easily computed using (11a) and (11b). In the second method, the multicellular model makes no assumption on the rate of virus release and uses the actual time-dependent profile of virus release from a single cell to compute the total number of A3G(-) and A3G(+) viruses in culture supernatant by using (12a) and (12b). Although this method provides a comprehensive link between the two models, it would be difficult to optimize the system parameters. Therefore, we use the first method for optimizing the single-cell and multicellular model parameters. Having done that, we use the second method for our simulations regarding effects of A3G-based therapeutic strategies, drug penetrance and administration time on HIV replication.
3.4 Results

3.4.1 Single-cell model of APOBEC3G-Vif interactions

For our simulations, we obtained parameters from published biological experiments. Several groups have measured the degradation profiles of Vif and of A3G in the presence and absence of Vif [53,56,160,161]. First-order kinetic decay curves were used to approximate the degradation rate of Vif, A3G, and A3G-Vif complex from this experimental data as 0.25 hr$^{-1}$, 0.1 hr$^{-1}$, and 0.3 hr$^{-1}$, respectively (Figure 3-4A-C). The binding affinity of A3G to Vif has been estimated to be in the low micromolar range by surface plasmon resonance [162]. We assume a value of 1 µM for the binding affinity, and calculate $k_{on}$ by assuming $k_{off} = 3600$ hr$^{-1}$. The stoichiometry of A3G and HIV, i.e., the number of A3G proteins incorporated into a virion, $s_{A3G}$, has been estimated at seven molecules [163]. In an infected cell, viral protein production does not start as soon as HIV enters in the cell. Based on published experimental studies, Vif production begins approximately 12 hours after infection ($t_{prod,Vif} = 12$ hr), increases through 24 hours ($\Delta t_{rise,Vif} = 12$ hr) and remains roughly constant after that. This was experimentally observed by measuring expression of HIV-1 RNA transcripts during HIV infection in [164,165] and computationally verified in [153]. HIV particle formation is assumed to begin 16 hours post infection ($t_{form,HIV} = 16$ hr), continue to increase for 8 hours ($\Delta t_{rise,HIVform} = 8$ hr) and plateau after that. In addition, a 6-hour delay is assumed for the budding process, meaning that virus release begins at 22 ($t_{rel,HIV} = 22$ hr) and increases through 30 hours post infection ($\Delta t_{rise,HIVrel} = 8$ hr). These assumptions are consistent with
measurements of reverse transcriptase activity in cell culture supernatants indicating active release of viruses from cells [165] and with the model predictions in [153].

Using the above parameters as a basis, we estimated the production rates of A3G and Vif. We have not found experimental data quantifying these production rates, therefore, we estimate these parameters using data from [36]. In those experiments, 293T cells were used as a 'permissive' cell line, meaning that they did not express A3G intrinsically. These cells were co-transfected with two vectors: one either a wild-type (WT) or a Vif-deficient (ΔVif) X4 provirus; the other (at varying doses) encoding A3G. After a day, the levels of supernatant p24 protein in the culture were monitored. The number of viruses can be calculated from this, as the 24-kDa p24 protein is estimated to be present at 2000-4000 molecule per virion [166], therefore 1 pg p24 ≈ 12,500 HIV particles. The supernatant including viruses from the infected 293T cells was extracted, normalized by p24 content, and used to challenge indicator cell lines in a single-cycle infection assay in which expression of chloramphenicol acetyltransferase (CAT) indicated HIV infection. After 28 hours, the number of infected cells were measured and normalized [36]. We use the single-cell model to find the percentage of released A3G(-) viruses in culture in the transfection part of this experiment and employ the multicellular model with equations (11a) and (11b) for the inoculation of CAT-indicator cells. Although the first part of this algorithm is a transfection of a T cell culture, information on the number of cells and transfection efficiency was not available. Moreover, we know that cells begin releasing viruses 22 hours after infection and in this experiment the culture supernatant was extracted 24 hours after transfection. This means that viruses in the supernatant were released from cells that were infected in the first two hours after transfection. Therefore,
asynchronous infection would not be an issue in this case. This justifies using the single-cell model for the transfection part of this experiment.

Estimation of Vif and A3G production rates required an exhaustive search in $P_{Vif^{-}}P_{A3G}$ domain. The procedure is as follows. For a given pair of ($P_{Vif}$, $P_{A3G}$), the single-cell model computes the HIV replicative potential, the percentage of released viruses that are A3G(-). Next, in the multicellular model, CAT-indicator cells ($T_0 = 500,000$ cells) are inoculated by viruses corresponding to 5 ng p24 with the ratio of A3G(-) to total viruses equal to $p$. The simulations are run for three different doses of A3G as in the experiment: ($P_{Vif}$, $P_{A3G}$), ($P_{Vif}$, $P_{A3G}$/3), and ($P_{Vif}$, $P_{A3G}$/6), and the number of infected cells at 28 hours is computed, normalized and compared to the experimental results to calculate the fitness error which is defined as the square root of sum of squares of differences between experimental data and computed results. For the HIV-ΔVif case ($P_{Vif} = 0$ µM/hr), the error for a range of $P_{A3G}$ is depicted in Figure 3-5A and the minimum error is achieved at $P_{A3G} = 0.085$ µM/hr.

In contrast, Figure 3-5B shows the error for the HIV-WT case for a wide range of values of $P_{Vif}$ and $P_{A3G}$. The fitness increases from low to high as the color changes from red to dark blue. This figure gives us possible pairs of ($P_{Vif}$, $P_{A3G}$) producing best fits to the experimental data. Based on the value of $P_{A3G}$ obtained from Figure 3-5A, the best fit is achieved at $P_{Vif} = 33$ µM/hr. The experimental data from Sheehy et al. [36] used for calculating $P_{Vif}$ and $P_{A3G}$ is re-plotted as blue bars in Figure 3-5C and Figure 3-5D for WT and ΔVif viruses, respectively. The red bars show model predictions of percentage of CAT cells infected by using optimized $P_{A3G}$ and $P_{Vif}$ in our simulations. The estimates of
Figure 3-4. Degradation profile of protein entities in the model.
Using experimental data and first-order decay curves, degradation coefficients of (A) Vif, (B) A3G, and (C) A3G-Vif complex were estimated to be 0.25 hr$^{-1}$, 0.1 hr$^{-1}$, and 0.3 hr$^{-1}$, respectively (re-plotted from Benedict et al. [156]).

Figure 3-5. Estimation of A3G and Vif production rates.
(A) shows the optimum $P_{A3G}$ for the ΔVif case ($P_{Vif} = 0$), whereas (B) shows the fitness error heatmap for a wide range of values of ($P_{Vif}$, $P_{A3G}$) for the WT case. The error decreases as color changes from dark red to dark blue. The optimum $P_{A3G}$ can be read from (A) and projected to the dark blue region of (B) to find the optimum $P_{Vif}$. The experimental data from Sheehy et al. [36]
were used for estimating $P_{Vi}$ and $P_{A3G}$ is re-plotted as blue bars for (C) WT and (D) ΔVif viruses. The red bars show our predictions of percentage of CAT cells infected by using estimated $P_{A3G}$ and $P_{Vi}$ in our simulations. All the results in (A) and (B) are obtained for $k_{A3G,HIV} = 50 \, \mu M^{-1}/hr$. The optimum values of (E) PVif and (F) PA3G versus $k_{A3G,HIV}$ are also shown.
$P_{A3G}$ and $P_{Vf}$ were computed for an assumed value of $k_{A3G,HIV} = 50 \, \mu M^{-1}/hr$, because A3G incorporation rate is not known. We repeated the simulations for a range of $k_{A3G,HIV}$ values and estimated values of $P_{A3G}$ and $P_{Vf}$ with the minimum error are shown in Figure 3-5E and Figure 3-5F. We found that $P_{A3G}$ is inversely proportional with $k_{A3G,HIV}$, whereas $P_{Vf}$ remains approximately constant. For the rest of this paper, we use the average value of $P_{Vf}$, which is equal to 35.6 µM/hr. For the estimation of $P_{Vf}$ and $P_{A3G}$, optimal values for burst size, infectivity rate and HIV half-life in vitro were used in the multicellular model. These parameters will be discussed in the next section.

### 3.4.2 Multicellular model of HIV propagation in culture

For the HIV replication experiments that we are simulating in this section, unlike the single-round infectivity experiments above, there is sufficient time for new viruses released by the cells to infect other cells. In these simulations, we assume that permissive or non-A3G expressing CEM-SS cells are used, with $t_{T,2} = 30$ hours ($k_p = 0.5545$ day$^{-1}$) [167,168]. In the multicellular model, we focus on the release of new HIV. This begins approximately 0.9 days post infection [149,153,165], so $t_{prod} = 22$ hours. Also, on average a life-span of 2.3 days was estimated for infected cells [149], therefore $t_{dead} = 55$ hours. In exploring the multicellular model, we assume that the release rate of viruses is constant during the productive phase. Later, when we integrate the single-cell model, the actual time-varying distribution of virus release is employed. Several estimates using different techniques are available, ranging from a hundred to a few thousand viruses per cell [169-175]. HIV clearance rate in vivo has been estimated in the range of a few hours to a couple of days [149,176-178], however, measurements of HIV clearance rate in vitro are not available. Using experimental data in [36], we estimate $k_{inf}$, $B$, and $t_{v,1/2}$. In [36],
CEM-SS cells \((T_0 = 500,000)\) were stably co-transfected with either an A3G- or \textit{neo}-
encoding vector. Then, both cell lines were inoculated by either \(\Delta Vif\) or WT viruses with
an initial dose of 1 or 10 ng p24. Accumulation of p24 in the culture supernatants was
monitored over time. As expected, efficient replication of both WT and \(\Delta Vif\) viruses was
observed in the \textit{neo}-expressing cells. The A3G-expressing cells also supported WT virus
growth, however, very low replication of \(\Delta Vif\) viruses was observed.

Since WT HIV replication in both \textit{neo}- and A3G-expressing cells were almost the
same, we conclude that the amount of A3G was insufficient to have a large effect on WT
HIV replication, suggesting that Vif had completely inhibited A3G and most of newly
produced viruses did not contain A3G. Therefore, even though we have A3G-expressing
cells, we assume \(p = 1\). Hereafter, we dismiss \textit{neo}-expressing cells and only focus on
A3G-expressing cells. Data points taken from the published experiments describing the
increase in WT HIV numbers \textit{in vitro} are shown in Figure 3-6A with red and blue squares
representing 1 and 10 ng p24 input HIV, respectively. As seen in Figure 3-6A, data points
corresponding to 1 ng p24 start at 0.1 ng p24/ml on day 3 after inoculation of cell culture.
This means that the volume of cell culture was equal to 10 ml. In our simulations, the
blue data points (10 ng p24) are used as training data for parameter estimation. Then, we
change the initial dose to 1 ng p24 while keeping the estimated parameters fixed and run
the model to examine how well it can reproduce the red data points.

Given a specific value for viral half-life \(t_v,1/2\), we used the default nonlinear curve-
fitting function in Matlab to find the optimum pair of burst size and infectivity \((B, k_{inf})\)
such that the simulated HIV growth curve fits the experimental data (WT - 10 ng p24)
with the minimum fitness error. We studied the effects of HIV half-life and cell doubling
time on these estimates of optimal $B$ and $k_{inf}$ in Figure 3-7A and Figure 3-7B. As $t_{v,1/2}$ increases, the estimated value of $k_{inf}$ decreases while estimated burst size stays roughly the same. However, as $t_{T,2}$ changes from 18 to 48 hours, the estimates of $B$ increase approximately from 400 to 5900 while the values of $k_{inf}$ decrease less than an order of magnitude. This is reasonable in a sense that in order to get the same amounts of HIV output as $t_{T,2}$ increases, viruses must infect with lower rates ($k_{inf} \downarrow$), however, they must produce more progeny in the cells ($B \uparrow$). Since we know $t_{T,2} = 30$ hours, only the optimum values of $B$ and $k_{inf}$ for each value of $t_{v,1/2}$ are shown in Table 3-1. For each set of parameters in Table 3-1, simulated HIV growth curves are shown in Figure 3-6A with dark blue representing $t_{v,1/2} = 24$ hours and light blue representing other values for $t_{v,1/2}$ in the range of 4 hours to infinity. The curves overlap and are a good fit for the blue squares. HIV growth curves with the same parameters and 1 ng p24 HIV input are also depicted in Figure 3-6A with dark ($t_{v,1/2} = 24$) and light ($t_{v,1/2} = 4 \ldots \infty$) red colors. Again, it is observed that all the curves are similar and provide a good fit for the red data points, with the exception of the last square, which we will discuss later. Therefore, none of the tested values for viral half-life, $t_{v,1/2}$, can be dismissed at this point, since all of them have generated good fits for the experimental data.
Figure 3-6. HIV growth curves for WT and ΔVif viruses.

(A) Inoculation of cultures of 500,000 cells with WT HIV. Blue and red squares represent 1 and 10 ng p24 HIV input, respectively. For a given $t_{v,1/2}$, values of burst size and virus infectivity rate were estimated such that the resulting simulated HIV growth curve fitted the blue data point with the minimum fitness error (shown as blue lines for $t_{v,1/2} = 4, 8, 16, 24, 26, \text{Inf}$ hours). Then, the estimated numbers were used to predict experimental data points corresponding to 1 ng p24 (shown as red lines). (B) Inoculation by 10 ng p24 ΔVif HIV. For each triplet of ($t_{v,1/2}, B, k_{\text{inf}}$) from (A), the values of $p$ and $c$ between 0 and 1 were chosen such that the generated curve provided the smallest error. None of the values of $t_{v,1/2}$ produced good fits to the blue circles except $t_{v,1/2} = 24$ hours where $p$ and $c$ were estimated to be 0.008 and 0.008, respectively. (C) The estimated parameters for $t_{v,1/2} = 24$ hours from (A) and (B) were used to examine how well they could generate a curve to fit experimental data points corresponding to 1 ng p24 ΔVif input (red circles). The red dashed line provided an acceptable fit to the data points except for the last circle where the line diverged. (D) All the experimental data points as well as their HIV growth curves are shown in red and blue colors corresponding to 1 and 10 ng p24 HIV input. Also, we included crowding effects in our simulation by using a logistic function. The two new curves drawn in light red (1 ng p24 ΔVif) and light blue (10 ng p24 ΔVif) show the HIV growth curves for this case. It can be seen that these curves provide a better fit to the experimental data than the curves in (C).
Figure 3-7. Effects of HIV half-life and cell doubling time on infectivity rate and burst size. (A) $k_{\text{inf}}$ decreases as $t_{T,2}$ changes from 18 to 48 hours. The same trend is also observed as $t_{v,1/2}$ increases from 4 hours to infinity. (B) Estimated burst size remains almost the same for different values of $t_{v,1/2}$, however, it increases as $t_{T,2}$ goes up.

Now, we focus on the experimental data obtained from 10 ng p24 ΔVif viruses (blue circles in Figure 3-6B). HIV growth curves corresponding to different values of $t_{v,1/2}$ increasing from the lowest curve to the highest one are also shown in Figure 3-6B. These curves were generated as follows. For each triplet of values for $(t_{v,1/2}, B, k_{\text{inf}})$ from Table 3-1, the optimal values of $p$ and $c$ between 0 and 1 were chosen using Matlab’s built-in nonlinear curve-fitting function such that the generated curve provided the lowest error (deviation from experimental data). Although all the values of $t_{v,1/2}$ could provide a good fit to the experimental data points in Figure 3-6A, they all failed to produce good fits to the blue circles in Figure 3-6B except $t_{v,1/2} = 24$ which provided a good match with $p = 0.008$ and $c = 0.008$. Therefore, we conclude that HIV half-life in vitro is approximately 24 hours and we use this number and corresponding numbers from Table 3-1 for burst size and infectivity rate in the rest of this study. Note that these are the optimal values of $t_{v,1/2}$, $B$, and $k_{\text{inf}}$ that were used in the multicellular model in the last section where we estimated $P_{Vif}$ and $P_{A3G}$. 

120
Table 3-2. Estimations of burst size and virus infectivity rate for $t_{T,2} = 30$ hours.

<table>
<thead>
<tr>
<th>$t_{v,1/2}$ (hr)</th>
<th>$B$</th>
<th>$k_{inf}$ (day$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1900</td>
<td>$34 \times 10^{-10}$</td>
</tr>
<tr>
<td>8</td>
<td>1900</td>
<td>$19 \times 10^{-10}$</td>
</tr>
<tr>
<td>16</td>
<td>1900</td>
<td>$11 \times 10^{-10}$</td>
</tr>
<tr>
<td>24</td>
<td>1900</td>
<td>$9 \times 10^{-10}$</td>
</tr>
<tr>
<td>36</td>
<td>1900</td>
<td>$7 \times 10^{-10}$</td>
</tr>
<tr>
<td>$\text{Inf}$</td>
<td>1900</td>
<td>$4 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

In Figure 3-6C, red circles correspond to 1 ng p24 $\Delta$Vif input and a simulation curve using the estimated parameters is superimposed. This curve is a reasonable fit to the data points except for the last red circle. Compared to the 10 ng p24 data, the experimental data points for 1 ng p24 $\Delta$Vif input are noisy, possibly because of the small amounts of supernatant p24 which are initially close to the detectable level of the p24 ELISA assay.

At this point, we must ask: why do the simulated HIV growth curves saturate around 9-10 days after infection in the case of WT viruses? In order to explain this, we study the distribution of cell states during the period of post infection. As seen in Figure 3-8A for the case of 10 ng p24 WT HIV input, normal cells dominate from the beginning of infection until the 8th day. But as infection progresses, infected cells begin to take over. These cells begin production of new viruses at around 22 hours after infection so the dominant cells on the 9th day are productive cells. Finally, viruses kill the productive...
cells at around 55 hours post infection and thus after the 11th day, almost all the cells in the culture are dead. Therefore, no more viruses can be produced and the HIV growth curve plateaus. The same scenario is observed in Figure 3-8B (WT – 1 ng p24), however, infection progresses more slowly due to presence of fewer viruses initially. As seen in Figure 3-8B, the number of normal cells is much higher than the previous case in the period of days 1-9. This explains why the total number of viruses will eventually be higher than the case of 10 ng p24 WT in Figure 3-6A. The reason is that many more normal cells are available for HIV infection in the case of 1 ng p24 WT, which results in production of more viruses in the model. This is a testable prediction of the model, which would require the experiment to be continued after 10 days.

Looking at Figure 3-8D for ΔVif viruses, we see more than 80 percent of the cells are healthy on the 15th day and the percentage of dead cells is negligible, in clear contrast to what was seen for WT viruses in Figure 3-8A-B. This is because of the less efficient propagation of infection of ΔVif viruses allowing normal proliferation of healthy cells. This in turn provides yet more cells for HIV to infect compared to the WT case. In fact, it may explain the rise of the ΔVif growth curve in Figure 3-6C and its divergence from the experimental data points. In our model, we have assumed constant cell proliferation rates, however, cells might slow or even stop proliferation when the suspension becomes crowded. In order to test whether this was a possible explanation for the observed data, we included the crowding effects in our model by using the logistic function. Therefore, (7) is replaced by

\[
\frac{dT(t)}{dt} = k_p T(t)(1 - \frac{T(t)}{T_{\text{max}}}) - k_{\text{inf}}(V_-(t) + V_+(t))T(t),
\]

(14)
Figure 3-8. Distribution of cell states during the period of post infection.

We simulated cultures of 500,000 healthy normal cells inoculated by either (A) 10 ng p24 or (B) 1 ng p24 WT HIV input. The infected cells start producing new virions after 22 hours and eventually die around 55 hours after infection. For WT HIV input, most of the cells are dead by the 12th day. In contrast, if we inoculate the cultures with either (C) 10 ng p24 or (D) 1 ng p24 ΔVif viruses, the healthy cells will still be the majority ones and the number of dead cells is negligible on the 15th day. In a different scenario, we included effects of cell culture crowding in our multicellular model by using a logistic function. Such cultures inoculated with either (E) 10 ng p24 or (F) 1 ng p24 ΔVif viruses provide better fits to biological experiments.
where $T_{\text{max}}$ is the maximum possible number of normal cells in culture and we set it to 50,000,000 (5,000,000 cells/ml). Modified ΔVif growth curves are shown in Figure 3-6D with light blue and red colors for 10 and 1 ng p24 ΔVif input, respectively. These curves show better fits to the experimental data, suggesting that crowding effects and slow proliferation could explain the experimental results. The distributions of cells states corresponding to these two cases are shown in Figure 3-8E-F.

3.4.3 Effects of A3G-based therapeutic strategies, drug penetrance and administration time on HIV replication

Using the model parameterized as above, we can compare the predicted efficacy of several therapeutic approaches targeting Vif-A3G interactions. Here, we add four specific molecules to the model and simulate the effect of their intracellular expression. All four are large proteins, as opposed to small molecules, and expression would in most cases require gene therapy. However, small molecules that had similar properties or effects on the functional A3G-Vif network could be delivered orally or intravenously [179,180]. The molecules are: Ab-Vif, a high-affinity antibody to Vif [181]; A3G, APOBEC3G itself, which could be upregulated by cytokines such as IL-2 [182] or NFAT and IRF proteins [183]; A3G$^{\text{AUB}}$, a mutated A3G with lower Vif-induced degradation rate (e.g., C97A-A3G [184]); and A3G$^{\text{AVIF}}$, a mutated A3G that does not bind Vif (e.g. A3G/F126-129 [185] and D128K-A3G [186]).

In the single-cell model, Ab-Vif is modeled as a new protein with an affinity for Vif 100 times that of A3G. The degradation rate of the complex formed by antibody bound to Vif is assumed to be $k_{d,Vif}$. Upregulation of A3G is modeled by increasing $P_{A3G}$. A3G$^{\text{AUB}}$ is a mutated A3G that binds to Vif, but its complex with Vif is not degraded faster than
unbound A3G. Therefore, the degradation rate of A3G_{ΔUb}-Vif complex is assumed to be $k_{d_{A3G}}$. Finally, A3G_{ΔVif} has the binding site for Vif mutated, and so does not bind to Vif. Note that we also assumed that each of these therapeutic proteins has the same degradation rate as of A3G. In our simulations, each of the drugs is expressed in addition to the cellular A3G produced (at a rate of $P_{A3G}$). The efficacy of each drug in terms of reduction in HIV replicative potential versus various production rates is shown in Figure 3-9 for $k_{A3G,HIV} = 5, 50$ and $500 \text{ µM}^{-1}/\text{hr}$. Among the therapeutic approaches, Ab-Vif shows a very poor performance even at very high production rates. It should be noted that Ab-Vif on its own can only block Vif from binding to A3G; in other words, it can make A3G more available but cannot add to its function. At least some A3G must be present in the cells to get incorporated into HIV particles. This explains the characteristic plateau as Ab-Vif expression increases (Figure 3-9); beyond this point all A3G is available to be encapsulated. A3G and A3G_{ΔUb} have efficacy profiles that are similar to each other, both better than that of Ab-Vif. That A3G_{ΔUb} is predicted to work only slightly better than A3G suggests that Vif-induced degradation of A3G through proteasomal pathway is not central to Vif effectiveness. A3G_{ΔVif} is predicted to be the best therapeutic approach among these drugs, two orders of magnitude better than A3G and A3G_{ΔUb}. This further suggests that the binding of A3G to Vif is an important interaction that should be inhibited to block HIV replication.

We next study A3G_{ΔVif} in the multicellular model, assuming 1 ng p24 WT input and $k_{A3G,HIV} = 50 \text{ µM}^{-1}/\text{hr}$, although qualitatively similar results will be obtained for other values of HIV inputs and $k_{A3G,HIV}$ (data not shown). The intracellular and multicellular models were coupled using (12a) and (12b) to compute the total number of A3G(-) and
A3G(+) viruses in culture supernatant. Figure 3-10A-D show HIV growth curves corresponding to various production rates of $A3G^{ΔVif}$. The blue and red lines represent A3G(-) and A3G(+) viruses, respectively, whereas, the greens lines represent total viruses in culture supernatant including A3G(-), A3G(+), and dead ones. As seen in Figure 3-10D, for $P_{A3G^{ΔVif}} = 10^{2}$, HIV replication slows more than two orders of magnitude and A3G(-) viruses reach a level of $10^{-4}$ by the 10th day. However, A3G(+) viruses start boosting by the 12th day. As mentioned earlier, this is caused due to an unconstrained proliferation of normal cells that provides a huge number of susceptible cells for infection. Similar to before, we constrained proliferation by including crowding effects using a logistic function (dashed lines in Figure 3-10). In this case, A3G(+) viruses reach a stable level below $10^{-1}$ ng p24/ml and decrease very slowly up to the 15th day. Therefore, $A3G^{ΔVif}$ has actually been able to stop HIV replication. A comparison of model predictions using equations (11a/b) or (12a/b) as the coupling method is shown in Figure 3-11.

At this point, we are interested to study the effects of drug penetrance on HIV replication when only a specific fraction of cells have been transfected ($P_{A3G^{ΔVif}} = 10^{2}$). Figure 3-10E-H show HIV growth curves with penetrance = 10, 90, 95 and 100 % of cells. It is observed that even if the drug is available to 95% of the cells, viruses can still actively replicate until they kill all the cells in the solution. This suggests that the drug must be available to almost 100% of the cells in order to be effective. HIV growth curves with similar penetrances for the constrained proliferation case are also depicted in Figure 3-10E-H. Even in this case, there is a large gap between curves corresponding to 95 and 100% drug availability.
Figure 3-9. Efficacy comparison of several drugs for different production rates.

Efficacy of several proposed therapeutic proteins in reducing the parameter $p$ estimated using the single-cell model for (A) $k_{A3G,HIV} = 5$, (B) 50 and (C) 500 $\mu$M$^{-1}$/hr. For all cases, A3GΔVif shows a better performance than other drugs.

In the next set of simulations, we studied the effects of drug administration time on the virus replication ($P_{A3G\Delta Vif} = 10^2$, penetrance = 100%). As seen in Figure 3-10K, if the drug is administered on the 9th day, the HIV growth curve is almost similar to the case that no drug was available to the cells at all as shown in Figure 3-10L. Note that these results were obtained for in vitro cases, where a constant source of cell production is not available as opposed to in vivo cases where old cells proliferate and new cells are born. Also, it should be mentioned that more than 50% of the cells are either infected or productive on the 9th day (Figure 3-8B) and a lot of viruses are available in culture supernatant. Therefore, administration of drug to remaining cells cannot help the culture survive. However, cells in the culture can still survive if we administer the drug before the 7th day. This suggests that the drug must be available to the cells shortly after
inoculation in order for the drugs to be effective. *In vivo*, the situation would be different; the constant birth of new cells may give this therapy greater hope of success.

**Figure 3-10. Effects of different production rates, penetrances and administration times of A3GΔVif on HIV growth curve.**

In all the simulations, 500,000 cells were inoculated by 1 ng 24 WT HIV input. (A-D) A3GΔVif with different production rates were administered right after inoculation. The green and blue lines represent A3G(-) and A3G(+) viruses in the culture, respectively. The red lines characterize all the viruses including A3G(-), A3G(+), and dead ones. For $P_{A3GΔVif} = 10^2$, the amount of A3G(-) viruses decay to $10^{-4}$ ng p24/ml by the 10th day, however, the number of A3G(+) viruses rises on the 12th day. Dashed lines represent cultures with constrained proliferation (crowding effects modeled by using a logistic function). In this case, it is seen that A3G(+) viruses reach a stable level below $10^{-1}$ ng p24/ml and decrease very slowly up to the 15th day for $P_{A3GΔVif} = 10^2$. This suggests that A3GΔVif has been able to stop HIV replication. (E-H) Effects of drug penetrance on HIV growth curves. We simulated cases where the drug was only available to a fraction of cells ($P_{A3GΔVif} = 10^2$). Comparing cases corresponding to 95% and 100%, we can see that there is a gap larger than two orders of magnitude between the total levels of p24 on the 15th day. This implies that drugs should be available to all the cells to get the desired efficacy. The same qualitative effect is observed in the cultures with constrained proliferation for different drug penetrances. (I-L) Effects of administration time on HIV growth curves ($P_{A3GΔVif} = 10^2$ and penetrance = 100%). It is seen that administering drug on the 9th day is not effective and the results are similar to the case of no drug. However, if the drug is administered before the 7th day, cell could still survive. The same trend in effects of drug administration time is also observed in cultures with constrained proliferation.

128
Figure 3-11. Comparison of model predictions using the two coupling methods.
In all the simulations, 500,000 cells were inoculated by 1 ng 24 WT HIV input. A3G$_{ΔVif}$ with different production rates were administered right after inoculation. The red and blue lines represent A3G(-) and A3G(+) viruses in the culture, respectively. The green lines characterize all the viruses including A3G(-), A3G(+), and dead ones. Dashed lines represent cultures with constrained proliferation (crowding effects modeled by using a logistic function). The intracellular and multicellular models were coupled using either (A-D) equations (12a/b) or (E-H) equations (11a/b). Although the coupling method using equations (11a/b) assumes that release rate of viruses from a productive cell (and the ratio of A3G(-) to total viruses) is constant over the period $[t_{prod}, t_{dead}]$, it provides a very good approximation of the model predictions obtained by the second coupling method using equations (12a/b). Note that there is no assumption on the rate of virus release in the second method and the actual time-dependent profile of virus release from a single cell is used to compute the total number of A3G(-) and A3G(+) viruses in culture supernatant.

3.4.4 Sensitivity analysis: determining critical model parameters that influence HIV replication

In this section, we analyze the effects of parameter variations in both single-cell and multicellular models. For the intracellular model, we investigate the deviations of HIV replicative potential resulted from $\pm 5\%$ change in each of the 17 model parameters (Figure 3-12A). As seen, $t_{form,HIV}$, the particle formation starting time, had the highest positive impact on $p$, whereas $t_{rel,HIV}$, the virus release starting time, had the highest
negative impact. This suggests that if the assembly and budding process of HIV particles from the cells could somehow be slowed, it would have a significant effect on virus replication. On the other hand, some of the parameters such as $s_{A3G}$, burst size, and $k_{d_{A3G}}$ had very negligible effects on the intracellular model output.

For the extracellular model, two outputs are considered for sensitivity analysis; 1) number of healthy cells and 2) number of A3G(-) viruses on the 6th day. In our simulations, 1 ng p24 WT HIV input was used for infection without administering any drug. In terms of the 1st output, almost none of the parameters had a major impact except the proliferation rate, $k_p$, which produced high variations in the number of normal cells (Figure 3-12B). In contrast, all the multicellular parameters had significant effects on the levels of A3G(-) viruses (Figure 3-12C). Considering the combined effects on both outputs, $k_p$ was detected as the most sensitive parameter in the multicellular model.
Figure 3-12. Sensitivity analysis in single-cell and multicellular models.

(A) The values of all 17 parameters in the intracellular model have changed by ±5% and the percentage change in $p$ for each of them is shown. Two parameters representing the time origins of virus release and particle formation had the highest impact on $p$. In contrast, parameters such as burst size, the degradation rate of A3G, and the stoichiometry of A3G proteins incorporated in HIV particles had zero or negligible effects on $p$. For the extracellular model, the effects of parameter variations were studied on two outputs; (B) number of normal cells and (C) number of A3G(-) viruses on the 6th day. In both cases, variations of cells proliferation rate had the highest impact on the extracellular model outputs.
3.5 Discussion

We have developed a mathematical model of the HIV lifecycle inside and outside of cells, using differential equations. Our model is the first one developed to specifically couple molecular-level events within individual cells to the viral dynamics and multiple cycles of infection within a population of cells. In this chapter, we used two different methods to couple the two models. Estimation of the system parameters was done using the model with equations (11a) and (11b) in which the release rate of viruses was assumed to be constant over period of $[t_{\text{prod}}, t_{\text{dead}}]$ and also $p$ remained constant during this period. For the rest of our simulations to study effects of A3G-based therapies, the model with equations (12a) and (12b) was used in which the time-dependent distribution of virus release from a single cell was employed to compute the total number of A3G(-) and A3G(+) viruses. The multi-scale system allowed us to achieve a quantitative understanding of the Vif-A3G pathway in HIV pathogenesis. Experimental data were used to establish system parameters such as stoichiometry of molecules, degradation rates of proteins, production profiles of viral proteins, viral burst size, cell proliferation rate, life-span of infected cells, viral generation time and virus clearance rate. We validated our system by reproducing the results of in vitro T cell culture experiments. We found that both downstream effects of A3G (hypermutation and reduction of viral burst size) were important to replicate the experimental results in silico. Based on the model simulations, in vitro virus clearance was estimated to be 24 hours. The model also predicted that the average number of HIV viruses produced by an infected cell is 1900. We simulated two types of T cell cultures with unconstrained or constrained proliferation rate (including crowding effects by using a logistic function). It was observed that
simulated HIV growth curves provided better fits to the experimental data in the latter case suggesting that proliferation may slow down in cell culture after it gets crowded.

Several therapeutic molecules targeting the Vif-A3G pathway were tested in our system. These included a high-affinity antibody to Vif [181], APOBEC3G itself, a mutated A3G with lower Vif-induced degradation rate (A3G$^{ΔUb}$) [184], and a mutated A3G that does not bind Vif (A3G$^{ΔVif}$) [185,186]. It was found that A3G$^{ΔVif}$ was the most effective drug that could stop HIV replication. This also implied that inhibition of A3G binding to Vif is a crucial step in blocking HIV replication. We further studied A3G$^{ΔVif}$ with respect to effects of penetrance and administration time on HIV replication. The model predicted that the drug must be available to almost 100% of the cells in order to get the desired efficacy. Also it must be available to the cells shortly after inoculation in order for the cells to survive.

Sensitivity analysis of the single-cell and multicellular models helped us characterize parameters with significant impacts on the system. We did a local sensitivity analysis by changing each parameter by 5% and study their effects on the output parameters. In the single-cell model, we chose HIV replicative potential, the ratio of released A3G(-) viruses to the total number of released viruses, as the output parameter. This is a critical parameter in our system linking the two models together. We found that $t_{rel,HIV}$ and $t_{form,HIV}$ are the most sensitive parameters. This implies that slowing the assembly and budding process of HIV particles from the cells reduces the number of output A3G(-) viruses. In the multicellular model, two outputs were chosen for sensitivity analysis: 1) number of healthy cells and 2) number of A3G(-) viruses on the 6th day. We found that the proliferation rate of cells had the highest combined impact on both output parameters.
In this study, we primarily focused on molecular and cellular processes of HIV infection \textit{in vitro}, however, this provides the necessary requirements to expand the model and move towards \textit{in vivo} computation modeling of HIV. In the extended model, virus clearance \textit{in vivo} and the mechanisms of cell birth, proliferation, and death would be different and new topics such as latency would come into play. Also, CD4+ T cells in the immune system can function as memory cells. Therefore, they can latently carry integrated HIV for the duration of their lifetime. These cells can survive for years and possibly decades and upon withdrawal of antiretroviral therapy, they become active and HIV viral loads rebound quickly. So, this concept of latency stage should also be accounted for in the model by having a small population of dormant infected cells that live for a long period of time and infrequently become activated to produce HIV. In addition, the immune system is also hugely diverse and has many more cells than \textit{in vitro} cell culture experiments. Also, some tissues such as GI tract are more susceptible to HIV infection than others. Therefore, compartmentalization is essential and specific models need to be developed for each tissue and they must be closely linked to represent the whole body. The \textit{in vivo} model will be more complicated but can answer some more fundamental questions about HIV pathogenesis than we cannot cover with \textit{in vitro} modeling.
Chapter 4. APOBEC3G-Augmented Stem Cell Therapy to Modulate HIV Replication: a Computational Study
4.1 Summary

The interplay between the innate immune system restriction factor APOBEC3G and the HIV protein Vif is a key host-retrovirus interaction. APOBEC3G can counteract HIV infection in at least two ways: by inducing lethal mutations on the viral cDNA; and by blocking steps in reverse transcription and viral integration into the host genome. HIV-Vif blocks these antiviral functions of APOBEC3G by impeding its encapsulation. Nonetheless, it has been shown that overexpression of APOBEC3G, or interfering with APOBEC3G-Vif binding, can efficiently block in vitro HIV replication. Some clinical studies have also suggested that high levels of APOBEC3G expression in HIV patients are correlated with increased CD4+ T cell count and low levels of viral load; however, other studies have reported contradictory results and challenged this observation. Stem cell therapy to replace a patient’s immune cells with cells that are more HIV-resistant is a promising approach. Pre-implantation gene transfection of these stem cells can augment the HIV-resistance of progeny CD4+ T cells. As a protein, APOBEC3G has the advantage that it can be genetically encoded, while small molecules cannot. We have developed a mathematical model to quantitatively study the effects on in vivo HIV replication of therapeutic delivery of CD34+ stem cells transfected to overexpress APOBEC3G. Our model suggests that stem cell therapy resulting in a high fraction of APOBEC3G-overexpressing CD4+ T cells can effectively inhibit in vivo HIV replication. We extended our model to simulate the combination of APOBEC3G therapy with other biological activities, to estimate the likelihood of improved outcomes.
4.2 Introduction

In the previous chapter, we developed a multi-scale computational model of HIV infection in \textit{in vitro} T cell culture, consisting of intracellular, cellular and extracellular events \cite{2}. Figure 1-2 shows a basic diagram of HIV infection as well as interactions between HIV-1 and A3G in HIV producing and newly infected CD4+ T cells. One of the predictions of that model was that overexpression of A3G or of a mutated form lacking the Vif-binding site (termed A3G$^{\Delta Vif}$) \cite{185,186} can effectively stop \textit{in vitro} HIV replication. This prediction was in agreement with a number of studies in which elevated levels of A3G expression resulted in A3G overcoming the effects of Vif \cite{36,60,63,64}. The model also predicted that the degradation of A3G by Vif is not a crucial step in HIV pathogenesis; instead it is the binding of A3G to Vif that is the key step and must be targeted to improve A3G efficacy \cite{2}. Our goal in this study is to transpose our validated model of A3G-Vif interactions from simulations of \textit{in vitro} cell culture to simulations of \textit{in vivo} HIV infection and treatment.

Our central question is, can the transplantation of stem cells transfected with A3G or its variants halt \textit{in vivo} HIV replication? If yes, what percentage of the cells must be transfected and how much inhibitory potency (A3G overexpression), is required? In this chapter, we build upon the basic model of \textit{in vivo} HIV infection \cite{146,148,177,178,187,188} and extend it using the results of our multi-scale model of A3G-Vif interactions (\textit{Error! Reference source not found.} and \cite{2}). Other relevant biological phenomena such as accelerated cell death are also included into extended models in order to study their impacts on the performance of A3G-SCT. Then we analyze the models at steady state to evaluate the likely long-term effects of this treatment.
In this section, we briefly introduce the basic model of *in vivo* HIV infection [146,148,177,178,187,188] and then extend it into models capable of studying the effects of A3G-augmented stem cell therapy (A3G-SCT) on virus replication in the body. All variables are capitalized and represent concentrations of different cell types or viruses with units of $\mu l^{-1}$. The following equations describe the *in vivo* viral dynamics of HIV.

\[
\frac{dT}{dt} = \lambda - d_T T - kVT, \quad (1a)
\]

\[
\frac{dI}{dt} = kVT - d_I I, \quad (1b)
\]

\[
\frac{dV}{dt} = NI - d_V V. \quad (1c)
\]

In this model, free viruses infect susceptible CD4+ T cells and give rise to infected cells (Figure 4-1A). Variables $T$, $I$ and $V$ represent the concentration of uninfected CD4+ T cells, infected cells, and free virus, respectively. The rate of infection is proportional to the concentration of free virus and T cells and is equal to $kVT$ where $k$ is the infectivity rate constant with units $\mu l \times day^{-1}$. The infected cells produce and release on average $N$ new free virions per day. Uninfected cells are produced in the thymus at a constant rate, $\lambda$ cells/($\mu l \times day$). Uninfected and infected cells die at the rates $d_T$ and $d_I$, respectively. Free viruses are cleared at a rate of $d_V$. The last three parameters have units of $day^{-1}$. The values for these parameters in our models are consistent with previously published results [2,147,148,176,177,189-194] and are given in Table 4-1. Prior to infection, there are zero viruses and zero infected cells; Uninfected cells are at the steady state value balancing production and death, i.e., $T^* = \lambda / d_T$. After infection occurs (in the model, this is...
represented by an initial input of viruses), the system moves towards one of two possible equilibrium points,

\[ \overline{I} = \frac{\lambda kN - d_I d_D d_V}{d_I kN} \quad \text{or} \quad \overline{I} = 0. \]  

(2)

Where \( \overline{I} = 0 \) suggests that virus has been eradicated from the body while \( \overline{I} = (\lambda kN - d_I d_D d_V)/d_I kN \) implies that virus has grown and established an infection in the body. The latter equilibrium point is only stable if the concentration of infected cells is greater than zero, meaning that the numerator in equation (2) takes a positive value or

\[ R_0 = \frac{kN \lambda}{d_I d_D d_V} > 1, \]  

(3)

where \( R_0 \) is known as the basic reproductive ratio, describing the average number of secondary infected cells arising from each primary infected cell when almost all cells are uninfected. Therefore, if the reproductive ratio takes values greater than one, HIV infection spreads through the body; while values less than one imply that the infection dies out and the virus becomes eradicated. In a study by Little et al., the basic reproductive ratio of 4 patients was estimated in the range of 5.2 to 9.1 with a mean value of 7.1 [195]. Stafford et al. analyzed 10 patients and found a median of 5.7, with a range between 2.8 and 11.0 for \( R_0 \) [194]. Using a slightly different model, one group found \( R_0 \) estimates ranging from 7.4 to 34, with a mean value of 19.3 [195] and recently another group obtained an estimated mean of 8.6 for \( R_0 \) with 75% of 47 patients having \( R_0 \) values less than 11 and two infected individuals having \( R_0 \) values greater than 20 (highest value = 26.4) [196]. The goal of any strategy for HIV treatment or cure must be to reduce \( R_0 \) to values less than one. Figure 4-1B shows the impact of decreasing \( k \) and hence reducing \( R_0 \) on the system.
Table 4-1. Parameter values used for simulations and calculations.

<table>
<thead>
<tr>
<th>Parameter (units)</th>
<th>Values</th>
<th>References</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$ (µl⁻¹/day)</td>
<td>20</td>
<td>[177,193]</td>
<td>Production rate of uninfected CD4+ T cells</td>
</tr>
<tr>
<td>$d_T$ (day⁻¹)</td>
<td>0.02</td>
<td>[189,193]</td>
<td>Death rate of uninfected cells</td>
</tr>
<tr>
<td>$d_I$ (day⁻¹)</td>
<td>0.39</td>
<td>[194]</td>
<td>Death rate of infected cells</td>
</tr>
<tr>
<td>$k$ (µl×day⁻¹)</td>
<td>2.11×10⁻⁴</td>
<td>Calculated</td>
<td>Rate constant for infection of cells by free virus</td>
</tr>
<tr>
<td>$N$ (day⁻¹)</td>
<td>850</td>
<td>[194]</td>
<td>Average number of virus produced by an infected cell</td>
</tr>
<tr>
<td>$d_V$ (day⁻¹)</td>
<td>23</td>
<td>[148,176]</td>
<td>Clearance rate of free virus</td>
</tr>
<tr>
<td>$p^{(wt)}$</td>
<td>0.83</td>
<td>Estimated</td>
<td>A3G-free virus release ratio: fraction of A3G(-) viruses released from infected WT CD4+ T cells</td>
</tr>
<tr>
<td>$c$</td>
<td>10⁻³-10⁻²</td>
<td>[2]</td>
<td>Reduction in the average number of released virions from cells infected by A3G(+) viruses</td>
</tr>
<tr>
<td>$f$</td>
<td>0 to 1</td>
<td></td>
<td>Fraction of uninfected T cells overexpressing A3G</td>
</tr>
<tr>
<td>$t$</td>
<td>&gt; 1</td>
<td></td>
<td>Ratio of death rate of infected A3G-augmented cells to death rate of infected WT cells</td>
</tr>
<tr>
<td>$w$</td>
<td>&gt; 1</td>
<td></td>
<td>Ratio of death rate of cells infected by A3G(+) viruses to death rate of cells infected by A3G(-) viruses</td>
</tr>
<tr>
<td>$r$</td>
<td>0-1</td>
<td></td>
<td>Apoptosis failure rate: fraction of infected A3G-augmented cells that escape apoptosis</td>
</tr>
<tr>
<td>$R_0$</td>
<td>20</td>
<td>[194-196]</td>
<td>Basic reproductive ratio (output): number of new infections arising from a single infected cell when almost all the other cells are uninfected.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Varies</td>
<td>Calculated</td>
<td>Reduction factor (output): reduction in the basic reproductive ratio resulting from A3G-SCT</td>
</tr>
</tbody>
</table>

**Figure 4-1. The basic HIV model: schematic diagram and simulations.**

(A) The model consists of three entities: Free viruses, “uninfected” and “infected” CD4+ T cells. Before infection, only uninfected cells are present with the production rate of $\lambda$ and the death rate of $d_T$. In the model, infection occurs by introducing an initial amount of viruses to the body. Free viruses infect uninfected cells and give rise to infected cells with $k$ representing infectivity rate constant. Infected cells die at a rate of $d_I$; before death, these cells produce and release $N$ new free virions per day. The in vivo clearance rate of viruses is denoted by $d_V$. (B) The basic reproductive ratio, $R_0$, is defined as the number of new infections that arise from a single infected cell when almost all the other cells are uninfected. This important metric determines whether the infection spreads ($R_0 > 1$) or dies out ($R_0 < 1$) in the body. In the numerical simulation, $R_0 = 20$ initially and a hypothetical treatment is administered on day 100 which reduces the value of reproductive ratio by $\sigma = 5, 15, 20, 30$, and $\infty$. Although $\sigma = \infty$ results in the fastest decline in the viral load, the gap between curves associated with $\sigma = 30$ and $\infty$ is small.
4.3.1 Model I: the basic HIV model for A3G-augmented cells

In the first extension of the basic model of HIV infection, we assume all the stem cells are transfected with A3G and hence all the progeny CD4+ T cells in the body overexpress A3G in addition to their biological expression of A3G; we term these A3G-augmented cells as opposed to WT cells. When these cells are infected by viruses, A3G overexpressed in cells overcomes Vif and gets encapsulated in some of the budded viruses, meaning that infected cells produce two types of virus: those that carry A3G, hence called A3G(+) viruses, and those viruses that do not, dubbed A3G(-) viruses; the ratio of A3G(-) to total released viruses is an important factor in the model. Cells infected by A3G(+) viruses produce fewer virions compared to cells infected by A3G(-) viruses, because the A3G carried in A3G(+) viruses affects viral production by inhibiting several steps of the HIV life cycle inside the infected cell. The model of HIV infection for A3G-augmented cells can be described by the following equations.

\[
\frac{dT^{(+)}}{dt} = \lambda - d_T T^{(+)} - k(V_{(-)} + V_{(+)})T^{(+)} , \quad (4a)
\]

\[
\frac{dI^{(-)}}{dt} = kV_{(-)} T^{(+)} - d_I I^{(-)} , \quad (4b)
\]

\[
\frac{dI^{(+)}}{dt} = kV_{(+) }T^{(+)} - d_I I^{(+) } , \quad (4c)
\]

\[
\frac{dV_{(-)}}{dt} = p^{(+)}(NI^{(+)}_{(-)} + cNI^{(+)_{(+)}} ) - d_v V_{(-)} , \quad (4d)
\]

\[
\frac{dV_{(+)}}{dt} = (1 - p^{(+)})(NI^{(+)}_{(-)} + cNI^{(+)_{(+)}} ) - d_v V_{(+)}, \quad (4e)
\]
Where $T^{(\text{+})}$ represents the concentration of A3G-augmented CD4+ T cells and $V_{(-)}$ and $V_{(\text{+})}$ denote the concentration of A3G(-) and A3G(+) viruses, respectively. $I_{(-)}^{(\text{+})}$ and $I_{(\text{+})}^{(\text{+})}$ represent the concentration of A3G-augmented cells infected by A3G(-) and A3G(+) viruses, respectively.

As mentioned earlier, infected A3G-augmented cells release both A3G(-) and A3G(+) viruses. In the model, parameter $p^{(\text{+})}$ taking values between 0 and 1, denotes the fraction of A3G(-) viruses released from an infected cell and therefore is termed A3G-free virus release ratio. The superscript on $p^{(\text{+})}$ denotes that it is a property of $T^{(\text{+})}$. As seen in equations (4d) and (4e), for every $N$ viruses that are produced $p^{(\text{+})}N$ are A3G(-) viruses whereas $(1-p^{(\text{+})})N$ are A3G(+) viruses. The value of $p^{(\text{+})}$ depends on the concentrations of Vif and of A3G expressed in the cell, the kinetics of viral production and release and whether WT or mutated variants of A3G is expressed in the cell. To compute $p^{(\text{+})}$, we use our previously built single-cell model [2], where the A3G-Vif interaction along with other intracellular events such as production and degradation of host and viral proteins, and assembly and release of new virions are described using differential equations. The value of $p^{(\text{+})}$ is inversely associated with the intracellular A3G getting encapsulated in the released viral particles, i.e., the higher the production rate of A3G inside the cell, the lower the value of $p^{(\text{+})}$. It has been observed that in single-round infectivity assays, cells infected by A3G(+) viruses produce fewer virions compared to cells infected by A3G(-) viruses [36]. Parameter $c$ denotes the reduction in the number of virions released from cells infected by A3G(+) viruses, i.e., $NI_{(-)}^{(\text{+})}$ vs. $cNI_{(\text{+})}^{(\text{+})}$ in equations (4d) and (4e). This parameter takes values between 0 and 1. Estimates of $c$ have been obtained using computational models of in vitro T cell culture infectivity assays [2]. In the model, it is
also assumed that cells infected by A3G(-) or A3G(+) viruses have the same death rate.

Figure 4-2A shows a schematic illustration of Model I. This model provides a benchmark for the optimum performance that A3G-SCT can achieve in modulating \textit{in vivo} HIV replication for given values of $p$ and $c$.

The observed reduction in the number of produced viruses from cells infected by A3G(+) viruses could in fact be explained by decreased infectivity rate of A3G(+) viruses or decreased level of virus production in cells infected by A3G(+) viruses (represented by $c$ in Model I) – or both. To compare the effect of decreased infectivity rate of A3G(+)
viruses with that of decreased level of virus production in cells infected by A3G(+) viruses, we present a modified version of Model I, where A3G(+) viruses have reduced infectivity (represented by $\eta$ in Model Ib in the supplementary information) compared to A3G(-) viruses. But when infection occurs, unlike Model I, the number of viruses released from cells infected by A3G(+) viruses is the same as that of cells infected by A3G(-) viruses, i.e., all infected cells have the same burst size. As we will see later in the results section, the level of reduction in the reproductive ratio for both Models I and Ib are the same given that the reduction in infectivity rate of A3G(+) viruses is equal to the reduction in level of virus production in cells infected by A3G(+) viruses, i.e., $c = \eta$. This suggests that from a modeling point of view, both of these hypotheses could explain the results of single-round infectivity assays, where cells infected by A3G(+) viruses produce fewer virions compared to cells infected by A3G(-) viruses. However, from a mechanistic point of view, the encapsulated A3G would be unlikely to have effects on viral entry. This is because the entry process involves the binding of gp41 and gp120 on the viral envelope to CD4 and chemokine coreceptors on the T cell surface, whereas during fusion and entry, A3G is inside the viral capsid and does not interact with the proteins on the viral envelope. Therefore, viral entry and infectivity rate constant are assumed to be the same for both types of virus. On the other hand, A3G inhibits several steps during integration and reverse transcription, resulting in production of many nonfunctional viral particles, i.e., reduction in the number of functional viruses released from the cells infected by A3G(+) viruses.

### 4.3.2 Model IIa: the basic HIV model for WT and A3G-augmented cells

We next take into account that it may not be possible to transfect all the CD34+ stem
cells and therefore not all the CD4+ T cells would overexpress A3G. Therefore, we extend Model I to include two subpopulations of uninfected cells: A3G-augmented cells that overexpress A3G and WT cells that express A3G at normal levels. Initial infection occurs with a certain amount of A3G(-) viruses. These viruses can infect both WT and A3G-augmented cells. WT CD4+ T cells express A3G at low levels such that Vif can inhibit most of the A3G encapsulation into virions. When WT cells become infected they produce mostly A3G(-) virions and less A3G(+) virions, i.e., \( p^{(wt)} \) takes high values in the range \([0, 1]\) (see the Methods section and [197]). In contrast, infected A3G-augmented cells produce a higher fraction of A3G(-) and a lower fraction of A3G(+) virions, i.e., \( p^{(+)} < p^{(wt)} \). The superscript on \( p \) denotes whether it is a property of WT or A3G-augmented cells. The newly released A3G(+) viruses will similarly infect both WT and A3G-augmented cells. But these infected cells produce fewer virions. Note that WT cells even after infection by A3G(+) viruses still produce more A3G(-) virions than A3G(+) virions.

Model IIa can be defined by a system of eight differential equations.

\[
\frac{dT^{(+)}}{dt} = f\lambda - d_T T^{(+)} - k(V^{(-)} + V^{(+)})T^{(+)} , \\
\frac{dT^{(wt)}}{dt} = (1-f)\lambda - d_T T^{(wt)} - k(V^{(-)} + V^{(+)})T^{(wt)} , \\
\frac{dI^{(-)}}{dt} = kV^{(-)}T^{(+)} - d_1 I^{(-)} , \\
\frac{dI^{(+)}}{dt} = kV^{(+)T^{(+)} - d_1 I^{(+)}} , \\
\frac{dI^{(wt)}}{dt} = kV^{(+)T^{(wt)} - d_1 I^{(wt)}} ,
\]

(5a) (5b) (5c) (5d) (5e)
\[
\frac{dI^{(\text{wt})}_{(+)}}{dt} = kV_{(+)}T^{(\text{wt})}_{(-)} - d_{I}I^{(\text{wt})}_{(+)},
\]

(5f)

\[
\frac{dV_{(-)}}{dt} = p^{(+)}(NI^{(+)}_{(-)} + cNI^{(+)\text{wt}}_{(+)} + cNI^{(+)\text{wt}}_{(+)}) - d_{V}V_{(-)},
\]

(5g)

\[
\frac{dV^{(+)}_{(+)}}{dt} = (1 - p^{(+)})(NI^{(+)}_{(-)} + cNI^{(+)\text{wt}}_{(+)} + cNI^{(+)\text{wt}}_{(+)}) - d_{V}V^{(+)}.
\]

(5h)

Where the concentration of A3G-augmented and WT cells are denoted by \(T^{(+)}\) and \(T^{(\text{wt})}\), respectively. Parameter \(f\) represents the fraction of uninfected A3G-augmented cells. In this model there are four subpopulations of infected cells. \(I^{(+)}_{(-)}\) and \(I^{(+)\text{wt}}_{(+)}\) represent the concentration of A3G-augmented cells that are infected by A3G(-) and A3G(+) viruses, respectively. Similarly, WT cells infected by A3G(-) and A3G(+) viruses are represented by \(I^{(-)}_{(-)\text{wt}}\) and \(I^{(+)\text{wt}}_{(+)\text{wt}}\), respectively. In general, for \(I\) variables, superscripts represent whether infected cells are WT or A3G-augmented, while subscripts denote what type of virus is the cause of infection. The infectivity rate constant is assumed to be the same for all virus-cell pairs. All the infected cells have the same death rate. As described, Model IIa has two submodels describing HIV infection in A3G-augmented and WT cells, drawn in Figure 4-2A and Figure 4-2B, respectively. The purpose of this model is to investigate what percentage of the cells must overexpress A3G to block \(\text{in vivo}\) viral replication.

4.3.3 Model IIb: the basic HIV model for WT and A3G-augmented cells with lower death rates for infected A3G-augmented cells

As mentioned above, A3G(+) viruses are mostly produced by infected A3G-augmented cells. These viruses are considered to be less harmful than A3G(-) viruses. This is because they cause the infected cells to produce fewer virions than do A3G(-)
viruses. Therefore, it can be hypothesized that if infected A3G-augmented cells, as the main source for producing A3G(+) viruses, live longer (die more slowly) compared to infected WT cells, we may achieve a better performance in blocking replication. We test the effect of this possible difference in cell lifespan by customizing A3G-augmented cells to have a lower death rate than WT cells after they get infected. To change the death rate of infected WT and A3G-augmented cells in Model IIa, we only need to replace equations (5c-5f) with equations (6c-6f).

\[ \frac{dI^{(+)}}{dt} = kV^{(+)T} - d^{(+)}_t I^{(+)}, \]

(6c)

\[ \frac{dI^{(+)}}{dt} = kV^{(+)T} - d^{(+)}_t I^{(+)}, \]

(6d)

\[ \frac{dI^{(wt)}}{dt} = kV^{(wt)T} - d^{(wt)}_t I^{(wt)}, \]

(6e)

\[ \frac{dI^{(wt)}}{dt} = kV^{(wt)T} - d^{(wt)}_t I^{(wt)}. \]

(6f)

Where \( d^{(+)}_t \) and \( d^{(wt)}_t \) represent the death rate of infected A3G-augmented and WT cells, respectively. Similar to our notations for \( I \) variables, superscripts on \( d \) parameters denote whether infected cells whose death rate is represented are WT or A3G-augmented. Note that \( d^{(wt)}_t \) is the same as \( d_t \) defined in the basic model of HIV infection; however, \( d^{(+)}_t \) takes smaller values than \( d_t \).
4.3.4 Model IIc: the basic HIV model for WT and A3G-augmented cells with lower death rates for cells infected by A3G(+) viruses

After infection, A3G(+) viruses cause cells to produce and release fewer virions than do A3G(-) viruses. Therefore, these viruses might be considered less toxic for cells and as a result cells infected by A3G(+) viruses might live longer. An important question to ask is therefore, how would the efficacy of A3G-SCT change in a model that has a lower death rate for cells infected by A3G(+) viruses? To find the answer, we replace equations (5c-5f) with equations (7c-7f).

\[
\frac{dI^{(+)}}{dt} = kV^{(+)} T^{(+)} - d_{I^{(+)}},
\]

(7c)

\[
\frac{dI^{(-)}}{dt} = kV^{(-)} T^{(-)} - d_{I^{(-)}},
\]

(7d)

\[
\frac{dI^{(wt)}}{dt} = kV^{(wt)} T^{(wt)} - d_{I^{(wt)}},
\]

(7e)

\[
\frac{dI^{(+)}_{(wt)}}{dt} = kV^{(+)} T^{(+)} - d_{I^{(+)}_{(wt)}},
\]

(7f)

Where the death rate of cells infected by A3G(+) and A3G(-) viruses are denoted by \(d_{I^{(+)}}\) and \(d_{I^{(-)}}\), respectively. Similar to our notations for \(I\) variables, subscripts on \(d\) parameters denote whether cells whose death rate is represented were infected by A3G(-) or A3G(+) viruses. Note that \(d_{I^{(+)}}\) is the same as \(d_I\) defined in the basic model of HIV infection; however, \(d_{I^{(+)}}\) takes smaller values than \(d_I\).
4.3.5 Model III: the basic HIV model for WT and A3G-augmented cells with auto-apoptosis capability

For stem cell transfection, the effector gene, in this study a WT A3G or a functional A3G variant, can be included in a conditional gene circuit with an appropriate biosensor (as opposed to under the control of a constitutive promoter). One such biosensor is the HIV LTR promoter: the HIV protein Tat binds to the promoter and initiates expression of the effector gene [198-200]. In this way, transfected A3G is unexpressed until virus entry is detected, and only then is anti-HIV protein expression boosted. This provides on-demand antiviral activity at the cellular level with low potential for side effects. Other functionalities can also be added to the circuit to make it more potent against HIV infection. For example, an additional effector gene can be added to induce activation of the apoptosis pathway upon infection, causing the infected cell to die, significantly reducing viral production. Model III as described by the following differential equations is an extension of Model IIa to study the effects of such apoptotic functionality on modulating viral production.

\[
\frac{dT^{(+)}}{dt} = f \lambda - d_T T^{(+)} - k(V_{(-)} + V_{(+)} )T^{(+)} , \tag{8a}
\]

\[
\frac{dT^{(wt)}}{dt} = (1 - f) \lambda - d_T T^{(wt)} - k(V_{(-)} + V_{(+)} )T^{(wt)} , \tag{8b}
\]

\[
\frac{dI_{(-)}^{(+)}}{dt} = rkV_{(-)} T^{(+)} - d_I I_{(-)}^{(+)}, \tag{8c}
\]

\[
\frac{dI_{(+)}^{(+)}}{dt} = rkV_{(+)} T^{(+)} - d_I I_{(+)}^{(+)}, \tag{8d}
\]
\[
\frac{dI^{(\text{wt})}_{(-)}}{dt} = kV_{(-)}T^{(\text{wt})}_{(-)} - d_{I^{(\text{wt})}_{(-)}}, \quad (8e)
\]
\[
\frac{dI^{(\text{wt})}_{(+)}}{dt} = kV_{(+)}T^{(\text{wt})}_{(+)} - d_{I^{(\text{wt})}_{(+)}}, \quad (8f)
\]
\[
\frac{dV_{(-)}}{dt} = p^{(+)}(NI^{(\text{wt})}_{(-)} + cNI^{(\text{wt})}_{(+)} + p^{(\text{wt})}(NI^{(\text{wt})}_{(-)} + cNI^{(\text{wt})}_{(+)} - d_{V_{(-)}}), \quad (8g)
\]
\[
\frac{dV_{(+)}}{dt} = (1 - p^{(+)})(NI^{(\text{wt})}_{(-)} + cNI^{(\text{wt})}_{(+)} + (1 - p^{(\text{wt})})(NI^{(\text{wt})}_{(-)} + cNI^{(\text{wt})}_{(+)} - d_{V_{(+)}}, \quad (8h)
\]

A3G-augmented cells with apoptosis capability are assumed to die after infection without producing infectious virions. However, it is also considered that some inefficiencies may be involved with this process, suggesting that some A3G-augmented cells can survive after infection. This is captured by the inefficacy rate \(r\) in equations (8c) and (8d), where \(r\) is the fraction of infected A3G-augmented cells that have escaped auto-apoptosis.

Model III submodels are visualized in Figure 4-2B and Figure 4-2C.

4.4 Methods

4.4.1 Estimation of the A3G-free virus release ratio, \(p^{(\text{wt})}\), for WT CD4+ T cells

Refsland et al. recently generated a new cell line called CEM2n [197]. The cell line is a diploid derivative of the tetraploid non-permissive human T cell line CEM. CEM2n expresses high levels of CD4 and CXCR4 and has non-permissive characteristics of CEM line. Using two rounds of rAAV-mediated gene targeting, they generated two A3G-null derivatives called A3G\(\Delta1\) and A3G\(\Delta2\) from CEM2n cell line. A3G-null derivatives and the parental cell line had similar growth rates. To evaluate the effect of A3G-knockout on HIV replication, cells were inoculated at a multiplicity of infection (MOI) of 1% with
Vif-proficient HIV. Accumulation of p24 in the culture supernatants was monitored over time (Figure 4-3 and [197]). We have previously developed a multicellular model describing an extracellular pool of HIV viruses infecting a population of T cells in cell culture [2]. Using our model and experimental data from Refsland et al. [197], we estimate the A3G-free virus release ratio for CEM2n cell line. In our simulations, the total transducing units (TU) can be calculated by multiplying MOI and total number of cells. It was estimated that 1 ng of p24 corresponds to 1000 transducing units [201]. Assuming $T_0 = 500,000$, we calculated the initial number of viruses at 5 ng p24. Since A3G-null cell lines do not express A3G, when they get infected they do not release A3G(+) viruses and hence $p^{(A3G\Delta 1)} = p^{(A3G\Delta 2)} = 1$. Using the data points corresponding to A3G-null cells (blue squares in Figure 4-3), we found the optimum pair of burst size and virus infectivity rate such that the simulated HIV growth curves (blue lines) fit the experimental data with the minimum fitness error (Table 4-2). While keeping the estimated parameters fixed, we then used data points corresponding to CEM2n cells (red squares in Figure 4-3) to estimate the A3G-free virus release ratio for these cells. Since Q-PCR results showed that CEM2n has roughly the same A3G mRNA expression as primary CD4+ T cells [197], we use the average value of our estimates for $p^{(CEM2n)}$ as the A3G-free virus release ratio for WT CD4+ T cells, i.e., $p^{(wt)} = 0.83$.

**Table 4-2.** Estimations of burst size, infectivity rate and A3G-free virus release ratio.

<table>
<thead>
<tr>
<th>A3G-null cell line</th>
<th>Burst size</th>
<th>Virus infectivity rate</th>
<th>$p^{(CEM2n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3GΔ1</td>
<td>4100</td>
<td>$26\times 10^{-10}$</td>
<td>0.81</td>
</tr>
<tr>
<td>A3GΔ2</td>
<td>4100</td>
<td>$28\times 10^{-10}$</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Figure 4-3. Estimation of $p^{(w)}$, A3G-free virus release ratio, for WT CD4+ T cells.
Blue squares represent p24 measurements in (A) A3GΔ1 and (B) A3GΔ1 cell lines after inoculation with Vif-proficient HIV at MOI = 0.01. Values of burst size and virus infectivity rate were estimated for each A3G-nul cell line by curve-fitting the blue data points to reach the minimum residual error. Keeping our estimates fixed for each cell line, we then estimated the value of $p_{(CEM2n)}^{(w)}$ such that the generated red curve fitted the red data points with the minimum fitness error.

4.5 Results

In all the simulations, the initial concentration of uninfected WT CD4+ T cells is 1000 cells/µl. Without loss of generality, the concentrations of A3G(+) and A3G(-) viruses are given initial values of zero and one virions/µl, respectively. A3G-augmented stem cell therapy (A3G-SCT) is introduced on day 100, long after the system has reached the steady state. $R_0 = 20$ in our simulations (Table 4-1). However, in figures, we also plot the level of reduction required to block in vivo HIV replication with $R_0 = 70$, a conservatively high value for the reproductive ratio that might occur temporarily during the course of infection. This shows the performance of A3G-SCT in the worst-case scenarios. In Figure 4-4A, Figure 4-4B, and Figure 4-5, A-L, $T^{(+)}_{tot}$ refers to the total concentration of uninfected and infected A3G-augmented cells, i.e., $T^{(+)}_{tot} = T^{(+)} + I^{(+)}_{(i)} + I^{(+)}_{(i)}$.
in Models I, II, and III. Analogously, in Models II and III, $T_{\text{tot}}^{(\text{wt})} = T_{\text{tot}}^{(\text{wt})} + I_{(\text{wt})}^{(+) + I_{(\text{wt})}^{(-)}}$, while in Model I, $T_{\text{tot}}^{(\text{wt})}$ refers to the total concentration of uninfected and infected WT CD4+ T cells before introducing the therapy. We use $p^{(\text{wt})} = 0.83$ in our simulations throughout this paper. The value of $p^{(\text{wt})} = 0.83$ was estimated using our previously published model [2] and WT HIV replication data in A3G-expressing and A3G-knockout cells [197] (see the Methods section).

4.5.1 A3G gene therapy can effectively stop in vivo HIV replication

Analyzing Model I at steady state and assuming that the virus grows and establishes infection, we can derive the reproductive ratio, given by the following equation.

$$R_1 = \frac{\lambda k N (p^{(+) + c(1-p^{(+)})})}{d_{d_T} d_V} \Rightarrow \sigma = \frac{R_1}{R_0} = p^{(+)} + c(1-p^{(+)}) \quad (9)$$

Where $\sigma$ is the reduction factor. Detailed derivation of equation (9) can be found in the Appendix. Since both $p^{(+)}$ and $c$ take values between zero and one, the reproductive ratio for Model I is reduced compared to the reproductive ratio for the basic model of HIV infection. With 100% of stem cells transfected with A3G and $p^{(+)} = 0.1$ (90% of the budded virions from each cell carry A3G), the therapy fails to reduce $R_1$ to values less than one ($R_1 = 2.02$), however, it causes a 4.8-fold increase in the concentration of T cells and a 15.3-fold decrease in the total concentration of viruses (Figure 4-4A). By decreasing $p^{(+)}$, the A3G-free virus release ratio, to 0.01 (increasing the percentage of budded virions carrying A3G to 99%), the therapy successfully eradicates the virus from the body and the concentration of CD4+ T cells goes back to the initial concentration of 1000 cells/µl (Figure 4-4B, $R_1 = 0.22$).
Figure 4-4. Effects of A3G-free virus release ratio on HIV replication and reproductive ratio in Model I.
Infection occurs on day 0 and A3G-SCT begins on day 100 (Model I assumes that all the cells overexpress A3G). The impact of the therapy on the total concentration of viruses and cells is shown for (A) $p^{(+) = 0.1}$ and (B) 0.01. Light and dark red lines represent $V(-)$ and $V(+) variables, respectively, while light and dark blue lines represent $T_{tot}^{(wt)}$ and $T_{tot}^{(+) variables, respectively. The left and right axes show the virus and cell concentrations. Parameter $c$ is given the value of 0.001. Although (A) $p^{(+)} = 0.1$ decreases the viral load and increases the T cell count, it cannot eradicate the virus ($R_1 = 2.02$). However, (B) $p^{(+)} = 0.01$ successfully reduces $R_1$ to 0.22 and the infection dies out. (C) shows the level of reduction in the reproductive ratio that can be achieved by A3G-SCT for different values of $p^{(+)}$ and $c$. The green and black dashed lines represent the minimum level of reduction needed to stop HIV replication for $R_0 = 20$ and 70, respectively. Note that from left to right on the bottom axis, the A3G-free virus release ratio decreases from $10^{-1}$ to $10^{-4}$. Simulation results suggest that for small values of $p^{(+)}$ and $c$, up to three orders of magnitude reduction can be achieved.

Figure 4-4C shows that A3G-SCT can reduce the basic reproductive ratio by three orders of magnitude for small values of $p^{(+)}$ and $c$. However, in order to reduce the reproductive ratio of 20 and 70 to less than one, $p^{(+)}$ only needs to take values less than
0.049 and 0.013, respectively when $c = 0.001$, suggesting the potential of A3G-SCT to successfully achieve a functional cure for HIV infection. Note that parameter $c$ plays a limiting role in reducing the reproductive ratio such that as $p^{(+)}$ goes to zero, the maximum amount of reduction that can be achieved is $1/c$ fold. Depicted curves in Figure 4-4C represent the best performance that can be achieved with this treatment for given values of $p^{(v)}$ and $c$. The modified version of Model I assumes that A3G(+) viruses have reduced infectivity (represented by $\eta$ in Model Ib in the Appendix) compared to A3G(-) viruses. But, the number of viruses released from cells infected by A3G(+) viruses is the same as that of cells infected by A3G(-) viruses. The reproductive ratio for Model Ib is exactly the same as equation (9) if $c$ is replaced with $\eta$ (compare equation (9) with equation (AIb-9) in the Appendix). This suggests that reduction in infectivity rate of A3G(+) viruses has the same effect on the reproductive ratio as does the reduction in level of virus production in cells infected by A3G(+) viruses. However, the later is more relevant from a mechanistic point of view.

4.5.2 Blocking replication requires high percentage of A3G-augmented cells

Since complete stem cell transfection may not be feasible, Model IIa splits CD4+ T cells into two subpopulations of WT and A3G-augmented cells to evaluate the limitations of imperfect transfection. Studying the model at steady state (see the Appendix), we find that the reproductive ratio can be written by

$$R_{2a} = \frac{f\lambda kN\left(\left(p^{(v)} + c(1 - p^{(+))}\right) + \frac{1 - f}{f}\left(p^{(wt)} + c(1 - p^{(wt))}\right)\right)}{d_v d_t d_r},$$

$$\Rightarrow \sigma = \frac{R_{2a}}{R_0} = f\left(p^{(v)} + c(1 - p^{(+))}\right) + (1 - f)\left(p^{(wt)} + c(1 - p^{(wt))}\right).$$
$R_{2a}$ is reduced compared to the basic reproductive ratio because parameters $p^{(+)}$, $p^{(wrt)}$, $c$, and $f$ take values less than one. However, it takes values greater than $R_1$ for $f < 1$, suggesting that imperfect transfection reduces the efficacy of A3G-SCT. The reduction factor, $\sigma$, shown in equation (10) can be intuitively calculated from the reduction factor obtained in Model I. This is because now there exists two cell subpopulations in the system: A3G-augmented cells with the reduction factor equal to $(p^{(+)} + c(1-p^{(+)}))$; and WT cells with $\sigma = (p^{(wrt)} + c(1-p^{(wrt)}))$. The total reduction factor is hence the sum of the reduction factor for each cell subpopulation multiplied by its fraction in the total population. Given that $p^{(wrt)}$ takes a relatively high value of 0.83, one should note that the reduction caused by WT cells in the reproductive ratio has a lesser impact than that of A3G-augmented cells.

Although A3G-SCT with $f = 90\%$ ($R_{2a} = 1.86$) and $94\%$ ($R_{2a} = 1.20$) causes 18.2- and 77-fold reduction in viral load and improves total T cell concentration by 5.2- and 7.8-fold, respectively, the virus is not eradicated and the infection is still spread in the body (Figure 4-5A and Figure 4-5B, $p^{(+)} = 0.01$). Increasing the percentage of A3G-augmented cells to 99% reduces $R_{2a}$ to 0.38 and causes the HIV infection to die out (Figure 4-5C, $p^{(+)} = 0.01$). Reduction in the reproductive ratio in Model IIa is shown for several values of $f$ in Figure 4-6A. For $f = 95-99\%$, A3G stem cell transfection can reduce the reproductive ratio by a factor of 23.5 to 107.6. Note that as $p^{(+)}$ and $c$ go to zero, the fold reduction in the reproductive ratio approaches $1/(p^{(wrt)}(1-f))$. Therefore regardless of how small $p^{(+)}$ and $c$ values are, the percentage of WT cells will determine the performance of the treatment. Figure 4-6A also shows that large reductions in the reproductive ratio can be achieved as $f$ approaches 1.
4.5.3 Lower efficacy is achieved if infected A3G-augmented cells die more slowly than do infected WT cells

Model IIb describes a special case when A3G-augmented CD4+ T cells can live longer after they get infected, resulting in lower death rates for these cells compared to infected WT cells. Calculating the reproductive ratio for Model IIb illustrates whether this strategy is beneficial. Modification of the death rate of infected A3G-augmented cells results in a minor change in equation (10). The new reproductive ratio can be written as (see the Appendix)

\[
R_{2b} = \frac{f \lambda k N \left(t \left(p^{+} + c(1 - p^{+}) \right) + \frac{1 - f}{f} \left(p^{wt} + c(1 - p^{wt}) \right) \right)}{d^{wt}_i d_r d_v}
\]

\[
\Rightarrow \sigma = \frac{R_{2b}}{R_0} = tf \left(p^{+} + c(1 - p^{+}) \right) + (1 - f)\left(p^{wt} + c(1 - p^{wt}) \right),
\]

where \(t = \frac{d^{wt}_i}{d^{+}_i} > 1\). Based on our assumption, \(t\) must take values greater than one, which suggests lower therapeutic performance if infected A3G-augmented cells live longer than infected WT cells, i.e., \(R_{2b} > R_{2a}\) for \(t > 1\).

Figure 4-5. Effects of percentage of transfected cells, death rate ratios, and auto-apoptosis failure rate on HIV replication in Models IIa, IIb, IIc, and III.

In all simulations, infection occurs on day 0 and A3G-SCT begins on day 100. In all the subfigures, light and dark red lines represent \(V_{(-)}\) and \(V_{(+)}\) variables, respectively, while light and dark blue lines represent \(T_{tot}^{(-)}\) and \(T_{tot}^{(+)}\) variables, respectively. The left and right axes show the virus and cell concentrations. Parameter \(c\) is given the value of 0.001. \(f = 99\%\) for all the subfigures except (A) and (B) where \(f = 90\%\) and 94\%, respectively. For the first three rows, \(p^{(+)} = 0.01\) (Models IIa, IIb, and IIc), while it is set to 0.1 for the last row (Model III). (A-C, \(R_{2a} = 1.86, 1.2, \text{and } 0.38\)) simulation results for Model IIa suggest that high percentage of A3G-augmented cells is required to stop in vivo HIV replication. (D-F, \(R_{2b} = 0.38, 0.6, \text{and } 1.25\)) Model IIb assumes lower death rates for infected A3G-augmented cells compared to infected WT cells, i.e., \(t > 1\). Simulation results suggest that the efficacy of the therapy is degraded as the value of \(t\) increases. (G-I, \(R_{2c} = 0.38, 0.46, \text{and } 0.66\)) A3G(+) viruses are assumed to be less toxic in
Model IIc. Therefore, cells infected by these viruses die more slowly compared to cells infected by A3G(-) viruses, i.e., $w > 1$. Model IIc predicts that lower death rates for cells infected by A3G(+) viruses have a diminishing effect on the performance of the therapy. ($J-L$, $R_3 = 2.16$, 0.87, and 0.37) In Model III, cells are equipped with an additional gene circuit that activates apoptosis pathway upon infection; however, the circuit has a failure rate of $r$. Simulation results indicate that providing cells with this additional gene circuit enhances the performance of the therapy and can reduce the reproductive ratio to values less than one even in cases that the A3G-free virus release ratio does not take very small values.
Figure 4-5, D-F show that as the value of $t$ increases, the performance of A3G-SCT diminishes ($p^{(t)} = 0.01, f = 99\%$). For $t = 2$, the A3G treatment can still eradicate the virus from the body (Figure 4-5E, $R_{2b} = 0.60$), but the rate of decline in viral load is slower than that of $t = 1$ (Figure 4-5D, $R_{2b} = 0.38$). Nonetheless, eradication is not possible for $t = 5$ (Figure 4-5F, $R_{2b} = 1.25$). As seen in Figure 4-6B, the highest reduction in the reproductive ratio is achieved when $t = 1$, i.e., when infected A3G-augmented cells have the same death rate as infected WT cells. Also as $p^{(t)}$ decreases, the gap between the performance of A3G-SCT with $t = 1$ and that of therapy with $t > 1$ decreases, meaning that for small values of $p^{(t)}$, the importance of difference between death rate of infected A3G-augmented and WT cells is less significant.

4.5.4 Lower efficacy is achieved if cells infected by A3G(+) viruses die more slowly than do cells infected by A3G(-) viruses

Model IIc focuses on a scenario where A3G(+) viruses are assumed to be less toxic. This is because they cause the infected cells to produce fewer virions than do A3G(-) viruses. Therefore, cells infected by A3G(+) viruses may live longer after getting infected, i.e., they have a lower death rate than that of cells infected by A3G(-) viruses. By calculating the equilibrium point for Model IIc in the Appendix, the reproductive ratio is obtained by

\[
R_{2c} = \frac{f \lambda k N \left( p^{(t)} + wc(1 - p^{(t)}) \right) + \frac{1 - f}{f} \left( p^{(wt)} + wc(1 - p^{(wt)}) \right)}{d_{I_c} d_T d_F} \]

\[
\Rightarrow \sigma = \frac{R_{2c}}{R_0} = f \left( p^{(t)} + wc(1 - p^{(t)}) \right) + (1 - f) \left( p^{(wt)} + wc(1 - p^{(wt)}) \right),
\]
where $w = \frac{d_{i(+)}}{d_{i(+)}} > 1$. Since $w$ is assumed to be greater than one, the performance of A3G-SCT is degraded compared to Model IIa ($R_{2c} > R_{2a}$ for $w > 1$). However, the effect of $w$ on the performance of A3G-SCT is less severe than that of $t$ in Model IIb presented in the previous section; for $f \geq 50\%$, $c = 0.001$ and $t = w > 1$, we always have $R_{2c} < R_{2a}$, unless $p^{(+) > 1.7 \times 10^{-4}}$. This can be seen by comparing Figure 4-6, B and C. In order to reduce the reproductive ratio from 70 to less than one with $f = 99.9\%$, the A3G-free virus release ratio, needs to be decreased by 1.1 or 1.5-fold when $w$ goes from 1 to 2 or 5, respectively (Figure 4-6C). This shows that $w$ has a mild effect on the performance. In contrast, these numbers are 2.2- and 7.4-fold when $t$ goes from 1 to 2 or 5, respectively (Figure 4-6B).

Comparing Figure 4-5, G-I, we observe that the efficacy of A3G-SCT decreases as the value of $w$ increases ($p^{(+) = 0.01, f = 99\%}$). For $w = 5$ (Figure 4-5H, $R_{2c} = 0.46$) and $w = 15$ (Figure 4-5I, $R_{2c} = 0.66$), the therapy can eradicate the virus from the body but the rate of decline in viral load is slower than when $w = 1$ (Figure 4-5G, $R_{2c} = 0.38$). By comparing Figure 4-5H and Figure 4-5F, it can be seen that for the same values of $w$ and $t$, the rate of decline in virus concentration is faster, the T cell count is higher, and the viral load is lower in Model IIc compared to Model IIb.

4.5.5 A3G-augmented cells with auto-apoptosis capability can stop replication more efficiently

Finally, Model III explores the possibility of enhancing efficacy of the treatment by co-transfecting stem cells with A3G and a gene circuit that induces activation of the apoptosis pathway in progeny CD4+ T cells upon infection by HIV. The gene circuit causes the cell to die after infection and hence it cannot produce any infectious virions.
However, if the infected cell somehow escapes the apoptosis pathway, then A3G overexpressed in that cell gets encapsulated into virions to induce its antiviral activities. The reproductive ratio for Model III is given by (see the Appendix)

\[
R_3 = \frac{f \lambda k N \left[ r \left( p^{(+)} + c(1 - p^{(+))}) \right] + \frac{1-f}{f} \left( p^{(wr)} + c(1 - p^{(wr))}) \right) \right]}{d_id_r d_v} \Rightarrow \sigma = \frac{R_3}{R_0} = rf \left( p^{(+)} + c(1 - p^{(+))}) \right) + (1-f) \left( p^{(wr)} + c(1 - p^{(wr))}) \right).
\]

Since \( r \) takes values less than one, the reproductive ratio of Model III is reduced compared to that of Model IIa, suggesting that a better efficacy can be achieved using co-transfection of stem cells with A3G and an auto-apoptosis gene circuit. Reduction in the reproductive ratio for several values of \( r \) is shown in Figure 4-6D. For \( f = 99\% \) and \( r = 0.1 \), A3G-SCT reduces the reproductive ratio of 70 to less than one for any value of \( p^{(+) less than 0.059. For \( r = 0.01 \), \( p^{(+) should only take values less than 0.60 to be able to eradicate the virus from the body. This greatly relaxes the pressure on parameter \( p^{(+) \), the A3G-free virus release ratio, to take values as small as 0.005 for \( r = 1 \) to eradicate the virus, suggesting that using this therapy A3G or its variants can be overexpressed at much lower concentrations, and still be effective in blocking HIV replication. Note that the improvement in the performance of therapy becomes less significant for small values of \( p^{(+) (Figure 4-6D). For \( f = 99\% \) and \( p^{(+) = 0.1 \), when \( r = 1 \), i.e., the apoptosis gene circuit fails all the time, the treatment reduces the viral load but cannot eradicate the virus from the body (Figure 4-5J, \( R_3 = 2.16 \)). As \( r \) decreases to 0.35 (the apoptosis gene circuit works 65\% of the time), the treatment can successfully eradicate the virus and the infection goes away (Figure 4-5K, \( R_3 = 0.87 \)). For a smaller value of \( r \) in Figure 4-5L
while $f = 99\%$ and $p^{(+)} = 0.1$ ($R_3 = 0.37$), the rate of decline in viral load is even faster than the case of $f = 99\%$ and $p^{(+)} = 0.01$ in Figure 4-5C ($R_1 = 0.38$).

**Figure 4-6. Effects of percentage of transfected cells, death rate ratios, and auto-apoptosis failure rate on reproductive ratio in Models IIa, IIb, IIc, and III.**

The level of reduction in the reproductive ratio that can be achieved by A3G-SCT for different values of $p^{(+)}$ is shown for each model. In all the subfigures, the green and black dashed lines represent the minimum level of reduction needed to stop HIV replication for $R_0 = 20$ and 70, respectively. Note that from left to right on the bottom axis, the A3G-free virus release ratio decreases from $10^{-1}$ to $10^{-4}$. Parameter $c$ is given the value of 0.001. (A) Model IIa suggests that $f = 95\%$ is required to block HIV replication for $R_0 = 20$. Higher values of $f$ are needed to block HIV replication for larger values of $R_0$. (B) Simulation results of Model IIb predict that the performance of the therapy will be degraded if infected A3G-augmented cells die more slowly compared to infected WT cells, i.e., when $t > 1$. (C) Model IIc also suggests that the therapy achieves lower efficacy if cells infected by A3G(+) viruses die more slowly than cells infected by A3G(-) viruses, i.e., when $w > 1$. However, the performance degradation is less severe than that of Model IIb. (D) Finally, Model III indicates that A3G-SCT can achieve better efficacy if infected A3G-augmented cells activate apoptosis pathway upon their infection.
4.5.6 A3G\textsuperscript{ΔVif} outperforms A3G in blocking \textit{in vivo} HIV replication

Using the results of our previously published single-cell model [2], we study the effects of A3G and A3G\textsuperscript{ΔVif} (an A3G variant that does not bind Vif such as A3G/F126-129 [185] and D128KA3G [186]) overexpression on the reproductive ratio. Figure 4-7A shows the reduction in the reproductive ratio achieved in Model IIa versus the production rate of A3G and A3G\textsuperscript{ΔVif}. The APOBEC production rate in Figure 4-7 is in addition to the normal level of A3G production in WT cells. For \( f = 99\% \), in order to reduce the reproductive ratio of 70 to values less than one, the production rate of A3G should be at least 26 \( \mu M/hr \) while this number is 0.252 \( \mu M/hr \) for A3G\textsuperscript{ΔVif}. Analogously, a large gap is seen between red (A3G) and blue (A3G\textsuperscript{ΔVif}) curves for other values of \( f \), suggesting that nearly two orders of magnitude lower production rate of A3G\textsuperscript{ΔVif} is required to achieve the same efficacy as that of A3G.

![Figure 4-7. Effects of A3G and A3G\textsuperscript{ΔVif} on reproductive ratio in Models IIa and III.](image)

The level of reduction in the reproductive ratio that can be achieved by overexpression of A3G (red) and A3G\textsuperscript{ΔVif} (blue) is shown. In the two subfigures, the green and black dashed lines represent the minimum level of reduction needed to stop HIV replication for \( R_0 = 20 \) and 70, respectively. Parameter \( c \) is given the value of 0.001. (A) Simulation results for Model IIa show that almost two orders of magnitude lower production of A3G\textsuperscript{ΔVif} compared to that of A3G is required to achieve the same performance. (B) Model III predicts that by decreasing the apoptosis failure rate, lower production rate of A3G and A3G\textsuperscript{ΔVif} is required to stop \textit{in vivo} HIV replication.
Simulation results for Model III show that as \( r \) decreases, the gap between red (A3G) and blue (A3G\(^{\Delta \text{Vif}}\)) curves widens (Figure 4-7B, \( f = 99\% \)). For \( r = 1 \), in order to reduce the reproductive ratio of 70 to one, A3G\(^{\Delta \text{Vif}}\) production rate is 103-fold lower than that of A3G; this number is 331-fold lower for \( r = 0.01 \). Note that for all the curves in Figure 4-7, A and B, the reduction factor does not decrease further after a certain production rate. This is because the percentage of WT cells, \( 1-f \), determines the maximum achievable performance of the treatment, regardless of how much A3G or A3G\(^{\Delta \text{Vif}}\) is overexpressed.

### 4.5.7 Impact of mixed A3G overexpression levels on the A3G-SCT performance

In Models I, II and III, it is assumed that cells are either WT (expressing A3G at normal levels) or A3G-augmented (overexpressing A3G at high levels). However, stem cell transfection is not an all-or-none phenomenon, i.e., after the transfection, some of the progeny CD4\(^+\) T cells overexpress A3G at high levels while others may express A3G at lower levels. Therefore, it is noteworthy to evaluate the performance of A3G-SCT when all CD4\(^+\) T cells express A3G higher than WT cells, but the overexpression can be either low or high. Model IV (defined in the Appendix) gives insights into how the performance of the therapy would change in this scenario. \( T^{(+\text{lo})} \) and \( T^{(+\text{hi})} \) represent two subpopulations of cells overexpressing A3G at low and high levels, respectively. We assume that \( p^{(+\text{hi})} < p^{(+\text{lo})} < p^{(\text{wt})} \). Similar to Model IIa, the reproductive ratio for Model IV can be mathematically derived as (see the Appendix)

\[
R_4 = \frac{f \lambda k N \left( (p^{(+\text{hi})} + c(1 - p^{(+\text{hi})})) + \frac{1-f}{f} (p^{(+\text{lo})} + c(1 - p^{(+\text{lo})})) \right)}{d_t d_t d_V}
\]

\[
\Rightarrow \sigma = \frac{R_4}{R_0} = f \left( p^{(+\text{hi})} + c(1 - p^{(+\text{hi})}) \right) + (1-f) \left( p^{(+\text{lo})} + c(1 - p^{(+\text{lo})}) \right). \tag{14}
\]
As seen in Figure 4-8, for $f = 90\%$ and $p^{(+h)} = 0.001$, the other 10\% of the cells that overexpress A3G at low levels need to have $p^{(+lo)}$ taking values less than 0.48 in order for A3G-SCT to reduce $R_0$ from 20 to one, i.e., only 52\% of the viruses released from these cells must carry A3G for a successful therapy. Note that $p^{(w)} = 0.83$, hence 17\% of viruses released from WT infected cells are already A3G(+). To achieve $p^{(+lo)} = 0.48$, production rate of transfected A3G can be 119-fold lower than that of cells expressing A3G at high levels. Increasing $p^{(+h)}$ to 0.01 while keeping $f$ unchanged results in $p^{(+lo)} = 0.40$ to have a reduction factor equal to 20. For $p^{(+h)} = 0.01$, when $f$ decreases to 50\%, i.e., the maximum achievable reduction in the reproductive ratio is only 2.4-fold in Model IIa, the therapy can cause 20-fold reduction only if $p^{(+lo)}$ takes values less than 0.09.

**Figure 4-8. Effects of $p^{(+lo)}$, A3G-free virus release ratio, on reproductive ratio.**

The level of reduction in the reproductive ratio that can be achieved by A3G-SCT is shown when $p^{(+h)} = 0.001$ (red) and 0.01 (blue). In our simulations, $p^{(+h)} < p^{(+lo)} < p^{(w)}$. The green and black horizontal dashed lines represent the minimum level of reduction needed to stop HIV replication for $R_0 = 20$ and 70, respectively. From left to right on the bottom axis, $p^{(+lo)}$ decreases from 0.83 to $10^{-3}$. Parameter $c$ is given the value of 0.001.

### 4.6 Discussion

Despite decades of research, 33 million people live with HIV; 2.6 million people are infected annually, with 1.8 million fatalities [202]. Current leading-edge treatment regimens such as highly active antiretroviral therapy (HAART) ensure that millions of
people with HIV lead normal lives, and prevent millions of additional infections by reducing infectivity. However, HAART is expensive, is not a cure, and only a fraction of HIV sufferers worldwide receive it. There remains a critical need for new therapies, especially cures. Cures are preferable because they can eliminate long-term maintenance costs and issues of adherence to long-term regimens.

The idea of gene therapy and stem cell transfection has recently renewed hope to achieve a functional cure for HIV. The reported cure of the “Berlin Patient” was achieved through transplantation of hematopoietic stem cells from a donor with two key characteristics: (a) donor was tissue matched for transplantation; (b) donor had a genetic mutation that conferred resistance to HIV [4-6]. Finding such a donor would be difficult in general, but matched donor stem cells could be augmented, to provide the HIV-resistance by inserting genes or gene networks into those cells before transplantation. Small molecules (such as HAART drugs) cannot be encoded, but overexpression of endogenous anti-HIV proteins such as APOBEC3G, which can efficiently inhibit viral reproduction, is possible; alternatively or in addition, encoding a pro-apoptotic stimulus would induce HIV-infected CD4+ T cells to die, shortening lifespan and limiting HIV production. Gene therapy has the potential to counter problems associated with current anti-HIV treatments such as drug side effects, patient adherence, and emergence of drug resistant viruses. For some patients, multiple rounds of low adherence and viral resistance can lead to exhaustion of all antiviral regimens. Gene therapy can be promising in these cases; however, in order for it to become a standard of care in treating HIV, issues such as safety and persistence of genetically modified cells in the body must be addressed.
Here, we developed mathematical models extending the basic model of \textit{in vivo} HIV infection to describe the impacts that delivery of stem cells transfected with wild type A3G or its variants can have on HIV replication. A key novel feature of this work is the incorporation of previously developed biologically validated model of A3G-Vif interactions in a single cell [2] into the established model of \textit{in vivo} HIV infection [146,148,177,178,187,188]. Our models can be generalized to describe and simulate other anti-HIV therapies. Two crucial parameters of these models are $p$ and $c$. In the general form, parameter $p$ is the fraction of released viruses from a single cell that are unaffected by a drug, an anti-HIV protein, or a restriction factor. On the other hand, parameter $c$ is the reduction in the number of viruses released from cells infected by affected viruses. Knowing $p$ and $c$ for a restriction factor, we can calculate the efficacy of the therapy.

In our simulations, A3G-augmented stem cell therapy (A3G-SCT) is introduced on day 100 after initial infection by changing the production of uninfected CD4$^+$ T cells in thymus to generate both WT and A3G-augmented CD4$^+$ T cells with the ratio of 1-$f_f$. In reality, for stem-cell based gene therapy, CD34$^+$ stem cells would be mobilized using granulocyte colony stimulating factor (GCSF) and harvested from the blood for \textit{ex vivo} gene modifications. Patients may or may not undergo myeloablative conditioning. This procedure is performed prior to bone marrow transplantation using chemotherapy or total body irradiation with the purpose of killing all the stem cells and suppressing the immune response. This in turn leaves the body prone to infections but reduces the risk of graft-versus-host disease. Finally, transfected stem cells are reinfused back into patients. After transplantation, these cells differentiate into cell types such as CD4$^+$ T cells and
macrophages that are now able to overexpress anti-HIV genes. All these would result in temporary changes in the values of the system parameters such as production and death rates of CD4+ T cells. However, several weeks after bone marrow transplantation, stem cells and their progenies are sufficiently expanded to restore the immune system and thus we assume that the system parameters take on values similar to those pre-therapy. Therefore, in this chapter, we focus on the steady state response rather than the transient response immediately after A3G-SCT to study the impact of therapy on modulating \textit{in vivo} HIV replication.

Model I assumes that all CD34+ stem cells in the body are transfected with A3G or one of its variants. The results for this model demonstrate that A3G-SCT can reduce the reproductive ratio to values less than one for sufficiently small values of $p^{(+)}$, i.e., the concentration of A3G or its variants in HIV-producing cells must be high enough such that it gets encapsulated in most viruses. In the next model, it is assumed that a fraction of stem cells remain untransfected. Model IIa describes the viral dynamics in a mixed population of A3G-augmented and WT CD4+ T cells and predicts that the percentage of stem cells transfected with A3G or its variants must be 95% or higher for the therapy to be able to successfully stop \textit{in vivo} HIV replication when the pre-therapy reproductive ratio is 20. This critical result suggests that A3G-SCT can be an effective functional cure for HIV infection and would provide successful results in most clinical settings where $R_0$ takes values in the range of 10 to 20 [194-196]. Note that the value of $f$ can be affected by two major factors: efficiency of stem cell transfection with A3G or its variants; and performance of bone marrow transplantation. Viral vectors can achieve high transfection efficiencies but they raise safety concerns. On the other hand, non-viral delivery systems
such as biomaterials are considered to be safe but further research and optimizations are needed to improve their efficiencies. Selection of cells transfected \textit{ex vivo} should lead to high incorporation. In terms of bone marrow transplantation, multiple rounds of reinfusion of CD34+ stem cells may be required in order to establish a high ratio of stem cells that overexpress A3G. This is crucial because the maximum achievable reduction by A3G-SCT is limited by the fraction of WT CD4+ T cells in the body.

Model IIb describes a hypothetical scenario, where infected A3G-augmented cells die more slowly (while the number of released viruses from infected cells per day remains unchanged) compared to infected WT cells. One might expect that the extra condition on infected A3G-augmented cells could change the balance of A3G(-) and A3G(+) viruses at steady state such that the reproductive ratio would be reduced more efficiently. However, the model predicts that having infected A3G-augmented cells live longer leads to lower performances. This unexpected result can be explained if we look at the two subpopulations of target cells. Since there is no interaction/feedback between the two subpopulations, WT and A3G-augmented cells appear to act as isolated subsystems. Therefore, the kinetics of WT subpopulation of cells is untouched, while infected A3G-augmented cells live longer and produce more virions compared to the previous case represented in Model IIa, resulting in a lower decrease in the reproductive ratio.

As mentioned, A3G(+) viruses are considered to be less harmful than A3G(-) viruses. Model IIc predicts that the performance of the therapy will be diminished if cells infected by A3G(+) viruses have a lower death rate than that of cells infected by A3G(-) viruses. Similar to Model IIb, in order to understand this result, we have to take into account that although cells infected by A3G(-) viruses are unchanged, other cells infected by A3G(+)
viruses live longer and release more virions. Therefore, the reduction in the reproductive ratio is lower compared to Model IIa, resulting in a lower performance. In the literature, we have not found *in vitro* measurements of death rates for cells infected by A3G(+) viruses. Therefore further studies can clarify whether cells infected by A3G(+) viruses indeed live longer, though the model suggests that the impact of lower death rates for cells infected by A3G(+) viruses is not severe (Figure 4-6C).

Effective gene therapy against HIV will likely require a combination of anti-HIV mechanisms. Therefore, in addition to transfecing stem cells with A3G, other gene circuits can also be employed that may improve the performance of the therapy. Model III focuses on a case where stem cells are co-transfected with A3G and an auto-apoptosis gene circuit. Simulation results of Model III demonstrate that the addition of the auto-apoptosis gene circuit eases the requirement to have very small values for $p^{(+)}$ to achieve efficacy. This is an important result because it shows that the auto-apoptosis gene circuit can be beneficial in cases where overexpression of A3G or its variants is toxic or not desirable for some reasons. Model IV relaxes the assumption of all-or none transfection and studies the performance of A3G-SCT when all stem cells are transfected but progeny CD4+ T cells overexpress A3G at varying levels. The model predicts that the performance of A3G-SCT can be significantly improved if cells that overexpress A3G at low levels have somewhat higher A3G production rates than WT cells.

In our model, parameters $p^{(+)}$ and $p^{(wt)}$, termed A3G-free virus release ratio, create an interface between the previously built single-cell model of A3G-Vif interactions [2] and the *in vivo* models of HIV infection described in this chapter. The A3G-free virus release ratio is dose-dependent and directly measures the capability of WT A3G or its variants to
get encapsulated into released virions. Therefore, it effectively avoids the necessity of taking into account the concentration of anti-HIV proteins in building the in vivo models. Having simplified the model, the A3G-free virus release ratio does not shed light on how much protein production is needed to achieve the required reduction in the reproductive ratio. Therefore, integrating the single-cell model results into the in vivo models of HIV infection helps us evaluate the required production rate of A3G and A3G$_{\Delta Vif}$ (a vif resistant A3G by a single amino acid mutation [185,186]) to stop HIV replication. Simulation results suggest that the production rate of A3G$_{\Delta Vif}$ can be nearly two orders of magnitude lower than that of A3G to achieve the same performance, suggesting that A3G$_{\Delta Vif}$ has great potential to be used in gene therapy [203]. This can be useful in cases where overexpression of A3G is toxic to cells. Other therapeutic approaches such as using high-affinity antibodies that bind to Vif [181] or blocking Vif dimerization by small peptides [204] have been suggested in the literature. One should note that both of these approaches attempt to inhibit Vif binding to A3G and hence maximize the efficacy of A3G encapsulation into virions. But the impact of these strategies is limited by the amount of A3G that is expressed in cells. In other words, these molecules cannot reduce the A3G-free virus release ratio sufficiently unless A3G is also overexpressed. In addition, they would not have any effect on the performance of A3G$_{\Delta Vif}$, because this protein is a variant of A3G that does not bind to Vif in the first place. One potential obstacle in using A3G$_{\Delta Vif}$ against HIV infection is the high mutation rate of the virus. HIV evolves rapidly and it could mutate Vif protein to restore its ability to bind A3G$_{\Delta Vif}$ and to suppress anti-HIV activities of this restriction factor. In order to address this issue, similar to the rationale of HAART, other anti-HIV genes should be combined with
A3GΔVif to provide several independent layers of protection against HIV infection, and hence reduce the chance of viral escape [203].

Although more studies need to be done on in vivo antiviral effects of A3G, our models suggest that A3G and its variants hold great promise to be used in stem cell-based anti-HIV gene therapy clinical trials. Our work presented here takes a computational approach to give insights into the logistics of a successful A3G-SCT. Using the viral dynamics of each infected individual, we could also personalize the therapy to be highly effective. A3G-SCT may be an option for HIV patients undergoing bone marrow transplantation due to other complications such as AIDS-related lymphoma.
4.7 Appendix. Calculating reproductive ratios in mathematical models

4.7.1 Model I: the basic HIV model for A3G-augmented cells (reduced burst size for cells infected by A3G(+) viruses)

\[
\frac{dT^{(+)}}{dt} = \lambda - d_t T^{(+) - k(V(-) + V(+)T^{(+)})} \Rightarrow \overline{V(-) + V(+) = \frac{\lambda}{k T^{(+)}} - \frac{d_t}{k}} \quad (AI-1)
\]

\[
\frac{dI^{(+)}}{dt} = kV(-)T^{(+) - d_t I^{(+) - \overline{I^{(+) = \frac{kV(-)T^{(+)}}{d_t}}}}} \quad (AI-2)
\]

\[
\frac{dI^{(+)}}{dt} = kV(+)T^{(+) - d_t I^{(+) - \overline{I^{(+) = \frac{kV(+)T^{(+)}}{d_t}}}}} \quad (AI-3)
\]

\[
\frac{dV(-)}{dt} = p^{(+)}(NI^{(+) - cNI^{(+)}) - d_v V(-) \Rightarrow \overline{V(-)} = (NI^{(+) + cNI^{(+)})/d_v} \quad (AI-4)
\]

\[
\frac{dV(+)^{(+)}}{dt} = (1 - p^{(+)})NI^{(+) - cNI^{(+)}) - d_v V(+ \Rightarrow \overline{V(+) = (NI^{(+) + cNI^{(+)}) (1 - p^{(+)})/d_v} \quad (AI-5)
\]

\[
(AI-2) \& (AI-4) \Rightarrow \overline{I^{(+) = \frac{p^{(+)kT^{(+)N(I^{(+) + cI^{(+)})}}}{d_t d_v}}} \quad (AI-6)
\]

\[
(AI-3) \& (AI-5) \Rightarrow \overline{I^{(+) = \frac{(1 - p^{(+)})kT^{(+)N(I^{(+) + cI^{(+)})}}}{d_t d_v}}} \quad (AI-7)
\]

\[
(AI-6) \& (AI-7) \Rightarrow \overline{T^{(+) = \frac{kT^{(+)N(p^{(+) + c(1 - p^{(+)}) + cI^{(+)})}}{d_t d_v}}} \quad (AI-8)
\]

\[
\overline{T^{(+) = \frac{d_t d_v}{kN(p^{(+) + c(1 - p^{(+)}})}} \quad (AI-9)
\]

\[
\overline{V(-) + V(+) = \frac{\lambda kN(p^{(+) + c(1 - p^{(+)}}) - d_t d_v d_v > 0} \quad (AI-9)
\]

\[
R_t = \frac{\lambda kN(p^{(+) + c(1 - p^{(+)})}}{d_t d_v d_v} = R_0(p^{(+) + c(1 - p^{(+)})} > 1
\]

174
4.7.2 Model Ib: the basic HIV model for A3G-augmented cells (reduced infectivity rate for A3G(+) viruses)

\[
\frac{dT^{(+)}(t)}{dt} = \lambda - d_T T^{(+)} - k(V^{-} + \eta V^{(+)} )T^{(+)} + \frac{\lambda}{kT^{(+)} + \frac{d_T}{k}}
\]

(Alb-1)

\[
\frac{dI^{(+)}(t)}{dt} = kV^{-}T^{(+)} - d_I I^{(+)} + \frac{kV^{-}T^{(+)}}{d_I}
\]

(Alb-2)

\[
\frac{dI^{(+)}(t)}{dt} = \eta kV^{(+)}T^{(+)} - d_I I^{(+)} + \frac{\eta kV^{(+)}T^{(+)}}{d_I}
\]

(Alb-3)

\[
\frac{dV^{-}(t)}{dt} = p^{(+)}N(I^{(+)}(t) + I^{(+)}(t)) + d_I V^{-} \rightarrow V^{-} = (I^{(+)}(t) + I^{(+)}(t))p^{(+)}N/d_v
\]

(Alb-4)

\[
\frac{dV^{(+)}(t)}{dt} = (1 - p^{(+)} )N(I^{(+)}(t) + I^{(+)}(t)) - d_v V^{(+)} \rightarrow V^{(+)} = (I^{(+)}(t) + I^{(+)}(t))(1 - p^{(+)} )N/d_v
\]

(Alb-5)

(Alb-2) & (Alb-4) \rightarrow \frac{I^{(+)}(t)}{d_I} = \frac{p^{(+)}kT^{(+)}N(I^{(+)}(t) + I^{(+)}(t))}{d_I d_v}

(Alb-6)

(Alb-3) & (Alb-5) \rightarrow \frac{I^{(+)}(t)}{d_I} = \frac{(1 - p^{(+)} )\eta kT^{(+)}N(I^{(+)}(t) + I^{(+)}(t))}{d_I d_v}

(Alb-7)

(Alb-6) & (Alb-7) \rightarrow \frac{I^{(+)}(t) + I^{(+)}(t)}{d_I} = \frac{kT^{(+)}N(p^{(+)} + \eta (1 - p^{(+)} ))(I^{(+)}(t) + I^{(+)}(t))}{d_I d_v}

(Alb-8)

\rightarrow T^{(+)} = \frac{d_I d_v}{kN((p^{(+)} + \eta (1 - p^{(+)} )))}

(Alb-9)

(Alb-1) & (Alb-8) \rightarrow \frac{V^{-} + \eta V^{(+)}(t)}{d_I d_v} = \frac{\lambda kN(p^{(+)} + \eta (1 - p^{(+)} )) - d_I d_v k}{d_I d_v k} > 0 \rightarrow R_{ib} = \frac{\lambda kN(p^{(+)} + \eta (1 - p^{(+)} ))}{d_I d_v k} = R_0 (p^{(+)} + \eta (1 - p^{(+)} )) > 1

175
4.7.3 Model IIa: the basic HIV model for WT and A3G-augmented cells

\[
\frac{dT^{(+)}(t)}{dt} = f \lambda - d_r T^{(+)}(t) - k(V_{(-)} + V_{(+)} ) T^{(+)}(t) \rightarrow f \lambda = T^{(+)}(t) (d_r + k(V_{(-)} + V_{(+)})) \tag{AIIa-1}
\]

\[
\frac{dT^{(wt)}(t)}{dt} = (1-f) \lambda - d_r T^{(wt)}(t) - k(V_{(-)} + V_{(+)} ) T^{(wt)}(t) \rightarrow
\]

\[
(1-f) \lambda = T^{(wt)}(t) (d_r + k(V_{(-)} + V_{(+)})) \tag{AIIa-2}
\]

\[
\frac{dI^{(+)}(t)}{dt} = kV_{(-)} T^{(+)}(t) - d_i I^{(+)}(t) \rightarrow \frac{I^{(+)}(t)}{d_i} = \frac{kV_{(-)} T^{(+)}(t)}{d_i} \tag{AIIa-3}
\]

\[
\frac{dI^{(wt)}(t)}{dt} = kV_{(-)} T^{(wt)}(t) - d_i I^{(wt)}(t) \rightarrow \frac{I^{(wt)}(t)}{d_i} = \frac{kV_{(-)} T^{(wt)}(t)}{d_i} \tag{AIIa-5}
\]

\[
\frac{dI^{(+)}(t)}{dt} = kV_{(+)} T^{(+)}(t) - d_i I^{(+)}(t) \rightarrow \frac{I^{(+)}(t)}{d_i} = \frac{kV_{(+)} T^{(+)}(t)}{d_i} \tag{AIIa-4}
\]

\[
\frac{dI^{(wt)}(t)}{dt} = kV_{(+)} T^{(wt)}(t) - d_i I^{(wt)}(t) \rightarrow \frac{I^{(wt)}(t)}{d_i} = \frac{kV_{(+)} T^{(wt)}(t)}{d_i} \tag{AIIa-6}
\]

\[
\frac{dV^{(+)}(t)}{dt} = p^{(+)} (NI^{(+)}_{(-)} + cNI^{(+)}_{(+)} ) + p^{(wt)} (NI^{(wt)}_{(-)} + cNI^{(wt)}_{(+)} ) - d_i V^{(-)} \rightarrow
\]

\[
\overline{V^{(-)}} = \frac{N}{d_v} \left( p^{(+)} \left( I^{(+)}_{(-)} + cI^{(+)}_{(+)} \right) + p^{(wt)} \left( I^{(wt)}_{(-)} + cI^{(wt)}_{(+)} \right) \right) \tag{AIIa-7}
\]

\[
\frac{dV^{(+)}(t)}{dt} = (1-p^{(+)})(NI^{(+)}_{(-)} + cNI^{(+)}_{(+)} ) + (1-p^{(wt)})(NI^{(wt)}_{(-)} + cNI^{(wt)}_{(+)} ) - d_i V^{(+)} \rightarrow
\]

\[
\overline{V^{(+)}} = \frac{N}{d_v} \left( (1-p^{(+)})(I^{(+)}_{(-)} + cI^{(+)}_{(+)} ) + (1-p^{(wt)})(I^{(wt)}_{(-)} + cI^{(wt)}_{(+)} ) \right) \tag{AIIa-8}
\]

\[(AIIa-3) \& (AIIa-7) \rightarrow \overline{I^{(+)}_{(-)}} = \alpha \left( p^{(+)}(I^{(+)}_{(-)} + cI^{(+)}_{(+)} ) + p^{(wt)}(I^{(wt)}_{(-)} + cI^{(wt)}_{(+)} ) \right) \tag{AIIa-9}\]
where $\alpha = \frac{kT^{(+)}}{d_1 d_1'}$

(AIIa-4) & (AIIa-8) \Rightarrow
\[ \overline{I}_{(-)}^{(+)} = \alpha \left( (1 - p^{(+)}) (\overline{I}_{(-)}^{(+)}) + c \overline{I}_{(-)}^{(+)}) + (1 - p^{(wt)}) (\overline{I}_{(-)}^{(wt)} + c \overline{I}_{(-)}^{(wt)}) \right) \] (AIIa-10)

(AIIa-5) & (AIIa-7) \Rightarrow
\[ \overline{I}_{(-)}^{(wt)} = \beta \left( p^{(+)}) (\overline{I}_{(-)}^{(+)}) + c \overline{I}_{(-)}^{(+)}) + p^{(wt)} (\overline{I}_{(-)}^{(wt)} + c \overline{I}_{(-)}^{(wt)}) \right) \] (AIIa-11)

where $\beta = \frac{kT^{(wt)}}{d_1 d_1'}$

(AIIa-6) & (AIIa-8) \Rightarrow
\[ \overline{I}_{(-)}^{(wt)} = \beta \left( (1 - p^{(+)}) (\overline{I}_{(-)}^{(+)}) + c \overline{I}_{(-)}^{(+)}) + (1 - p^{(wt)}) (\overline{I}_{(-)}^{(wt)} + c \overline{I}_{(-)}^{(wt)}) \right) \] (AIIa-12)

(AIIa-1) & (AIIa-2) \Rightarrow
\[ \frac{T^{(+)}}{T^{(wt)}} = \frac{f}{1 - f} \] (AIIa-13)

(AIIa-9) & (AIIa-11) \Rightarrow
\[ \frac{\overline{I}_{(-)}^{(+)}}{\overline{I}_{(-)}^{(wt)}} = \frac{\alpha}{\beta} = \frac{T^{(+)}}{T^{(wt)}} \] (AIIa-14)

(AIIa-10) & (AIIa-12) \Rightarrow
\[ \frac{\overline{I}_{(-)}^{(+)}}{\overline{I}_{(-)}^{(wt)}} = \frac{\alpha}{\beta} = \frac{T^{(+)}}{T^{(wt)}} \] (AIIa-15)

(AIIa-9) & (AIIa-10) & (AIIa-14) & (AIIa-15) \Rightarrow
\[ \overline{I}_{(-)}^{(+)}} + c \overline{I}_{(-)}^{(+)}) = \alpha \left( (p^{(+)}) + c(1 - p^{(+)}) (\overline{I}_{(-)}^{(+)}) + c \overline{I}_{(-)}^{(+)}) + (p^{(wt)} + c(1 - p^{(wt)}) (\overline{I}_{(-)}^{(wt)} + c \overline{I}_{(-)}^{(wt)}) \right) \] (AIIa-16)
\[ 1 = \alpha \left( p^{(+)} + c(1 - p^{(+)}) \right) + \frac{T^{(\text{tw})}}{T^{(\text{+})}} \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \]

\[ \Rightarrow \]

\[ = \frac{kN}{d_i d_y} \left( p^{(\text{+})} + c(1 - p^{(+)}) \right) + \frac{T^{(\text{tw})}}{T^{(\text{+})}} \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \]

(AIIa-13) & (AIIa-16) \Rightarrow

\[ \frac{T^{(\text{+})}}{kN} \left( p^{(\text{+})} + c(1 - p^{(+)}) \right) + \frac{1 - f}{f} \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \]

(AIIa-1 & (AIIa-17) \Rightarrow

\[ \frac{V^{(-)} + V^{(+)}}{d_i d_y} = \frac{f \lambda}{k T^{(\text{+})}} - \frac{d_i}{k} \]

\[ = \frac{f \lambda kN \left( p^{(+) + c(1 - p^{(+)})} + \frac{1 - f}{f} \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \right) - d_i d_i d_y}{d_i d_y k} > 1 \]

(AIIa-18)

\[ \Rightarrow R_{2a} = \frac{f \lambda kN \left( p^{(+) + c(1 - p^{(+)})} + \frac{1 - f}{f} \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \right)}{d_i d_y d_y} \]

\[ = R_0 \left( f \left( p^{(+) + c(1 - p^{(+)})} \right) + \left( 1 - f \right) \left( p^{(\text{tw})} + c(1 - p^{(\text{tw})}) \right) \right) > 1 \]
4.7.4 Model IIb: the basic HIV model for WT and A3G-augmented cells with lower death rates for infected A3G-augmented cells

\[
\frac{dT^{(+)}(t)}{dt} = f \lambda - d_V T^{(+)}(t) - k(V_{(-)} + V_{(+)}))T^{(+)}(t) \Rightarrow f \lambda = T^{(+)}(t)(d_V + k(V_{(-)} + V_{(+)})) \quad (\text{IIb-1})
\]

\[
\frac{dT^{(\text{wt})}(t)}{dt} = (1 - f) \lambda - d_V T^{(\text{wt})}(t) - k(V_{(-)} + V_{(+)}))T^{(\text{wt})}(t) \Rightarrow (1 - f) \lambda = T^{(\text{wt})}(t)(d_V + k(V_{(-)} + V_{(+)})) \quad (\text{IIb-2})
\]

\[
\frac{dI^{(+)}(t)}{dt} = kV_{(-)}T^{(+)}(t) - d_I^{(+)} I^{(+)}(t) \Rightarrow \overline{I^{(+)}(t)} = \frac{kV_{(-)}T^{(+)}(t)}{d_I^{(+)}(t)} \quad (\text{IIb-3})
\]

\[
\frac{dI^{(\text{wt})}(t)}{dt} = kV_{(+)T^{(\text{wt})}}(t) - d_I^{(\text{wt})} I^{(\text{wt})}(t) \Rightarrow \overline{I^{(\text{wt})}(t)} = \frac{kV_{(+)T^{(\text{wt})}}(t)}{d_I^{(\text{wt})}(t)} \quad (\text{IIb-4})
\]

\[
\frac{dI^{(-)}(t)}{dt} = kV_{(-)}T^{(-)}(t) - d_I^{(-)} I^{(-)}(t) \Rightarrow \overline{I^{(-)}(t)} = \frac{kV_{(-)}T^{(-)}(t)}{d_I^{(-)}(t)} \quad (\text{IIb-5})
\]

\[
\frac{dI^{(\text{wt})}(t)}{dt} = kV_{(+)T^{(\text{wt})}}(t) - d_I^{(\text{wt})} I^{(\text{wt})}(t) \Rightarrow \overline{I^{(\text{wt})}(t)} = \frac{kV_{(+)T^{(\text{wt})}}(t)}{d_I^{(\text{wt})}(t)} \quad (\text{IIb-6})
\]

\[
\frac{dV_{(-)}(t)}{dt} = p^{(+)}(NI^{(+)}(-) + cNI^{(+)}(0)) + p^{(\text{wt})}(NI^{(\text{wt})}(-) + cNI^{(\text{wt})}(0)) - d_V V_{(-)} \Rightarrow (\text{IIb-7})
\]

\[
\overline{V_{(-)}} = \frac{N}{d_V} \left( p^{(+)}(\overline{I^{(+)}(-)} + c\overline{I^{(0)}(0)}) + p^{(\text{wt})}(\overline{I^{(\text{wt})}(-)} + c\overline{I^{(\text{wt})}(0)}) \right) \quad (\text{IIb-8})
\]

\[
\frac{dV_{(+)}(t)}{dt} = (1 - p^{(+)})(NI^{(+)}(-) + cNI^{(+)}(0)) \Rightarrow (1 - p^{(+)})(NI^{(+)}(-) + cNI^{(+)}(0)) - d_V V_{(+)} \Rightarrow \overline{V_{(+)}} = \frac{N}{d_V} \left( (1 - p^{(+)})(\overline{I^{(+)}(-)} + c\overline{I^{(0)}(0)}) \right) \quad (\text{IIb-8})
\]
(AIIb-3) & (AIIb-7) → $\overline{I}^{(+)}_{(-)} = \alpha \left( p^{(+)} \overline{I}^{(+)}_{(-)} + c \overline{I}^{(w)}_{(-)} + p^{(w)} \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+)} \right)$

where $\alpha = \frac{kT^{(w)}N}{d^{(+)}_i d_y}$

(AIIb-4) & (AIIb-8) →

$\overline{I}^{(+)}_{(+)} = \alpha \left( 1 - p^{(w)} \right) \left( \overline{I}^{(+)}_{(-)} + c \overline{I}^{(w)}_{(+)} \right) + \left( 1 - p^{(w)} \right) \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)}$

(AIIb-5) & (AIIb-7) →

$\overline{I}^{(w)}_{(-)} = \beta \left( p^{(w)} \overline{I}^{(w)}_{(+)} + c \overline{I}^{(w)}_{(+)} \right) + \left( 1 - p^{(w)} \right) \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)}$

where $\beta = \frac{kT^{(w)}N}{d^{(w)}_i d_y}$

(AIIb-6) & (AIIb-8) →

$\overline{I}^{(w)}_{(+) - (w)} = \beta \left( 1 - p^{(w)} \right) \left( \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+)} \right) + \left( 1 - p^{(w)} \right) \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)}$

(AIIb-1) & (AIIb-2) → $\overline{T}^{(w)} / T^{(w)} = \frac{f}{1 - f}$

(AIIb-9) & (AIIb-11) → $\overline{I}^{(+)}_{(-)} / \overline{I}^{(w)}_{(-)} = \alpha / \beta = \frac{T^{(+)}_i d^{(w)}_i}{T^{(w)} d^{(w)}_i}$

(AIIb-10) & (AIIb-12) → $\overline{I}^{(+)}_{(+) - (w)} / \overline{I}^{(w)}_{(+) - (w)} = \alpha / \beta = \frac{T^{(+)}_i d^{(w)}_i}{T^{(w)} d^{(w)}_i}$

(AIIb-9) & (AIIb-10) & (AIIb-14) & (AIIb-15) →

$\overline{I}^{(+)}_{(-)} + c \overline{I}^{(+)}_{(+) - (w)}$

$= \alpha \left( p^{(+) + c(1 - p^{(w)})} \overline{I}^{(+)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)} \right) + \left( p^{(w)} + c(1 - p^{(w)}) \right) \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)}$

$= \alpha \left( p^{(+) + c(1 - p^{(w)})} \overline{I}^{(+)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)} \right) + \left( p^{(w)} + c(1 - p^{(w)}) \right) \overline{I}^{(w)}_{(+) - (w)} \overline{I}^{(w)}_{(-)} + c \overline{I}^{(w)}_{(+) - (w)}$
\[ 1 = \alpha \left( p^{(+)} + c(1 - p^{(+)} \right) + \frac{T^{(n)} d_i^{(+)} }{T^{(+)} d_i^{(n)}} \left( p^{(n)} + c(1 - p^{(n)} \right) \right) \]

\[ \Rightarrow \]

\[ = kN \left( \frac{T^{(n)} d_i}{d_i^{(n)}} \left( p^{(n)} + c(1 - p^{(n)} \right) + \frac{T^{(n)} d_i}{d_i^{(n)}} \left( p^{(n)} + c(1 - p^{(n)} \right) \right) \right) \]

(AIIb-13) & (AIIb-16) \Rightarrow

\[ \overline{T^{(n)}} = \frac{d_i^{(n)} d_i}{kN \left( \left( p^{(n)} + c(1 - p^{(n)} \right) + \frac{d_i^{(n)}}{d_i^{(n)}} \left( 1 - f \right) \left( p^{(n)} + c(1 - p^{(n)} \right) \right) - d_i^{(+)} d_i d_i \right)} \]

(AIIb-1) & (AIIb-17) \Rightarrow

\[ \overline{V_{(-)}} + \overline{V_{(+)}} = \frac{f \lambda}{kT^{(n)}} - \frac{d_i}{k} \]

\[ = \frac{f \lambda kN \left( \left( p^{(n)} + c(1 - p^{(n)} \right) + \frac{d_i^{(n)}}{d_i^{(n)}} \left( 1 - f \right) \left( p^{(n)} + c(1 - p^{(n)} \right) \right) - d_i^{(+)} d_i d_i \right)}{d_i^{(n)} d_i d_i} > 0 \]

\[ R_{2b} = \frac{f \lambda kN \left( \left( p^{(n)} + c(1 - p^{(n)} \right) + \frac{d_i^{(n)}}{d_i^{(n)}} \left( 1 - f \right) \left( p^{(n)} + c(1 - p^{(n)} \right) \right) }{d_i^{(n)} d_i d_i} \]

\[ = \frac{f \lambda kN \left( \left( p^{(n)} + c(1 - p^{(n)} \right) + \frac{1 - f}{f} \left( p^{(n)} + c(1 - p^{(n)} \right) \right) }{d_i^{(n)} d_i d_i} \]

\[ = R_0 \left( tf \left( p^{(n)} + c(1 - p^{(n)} \right) + (1 - f) \left( p^{(n)} + c(1 - p^{(n)} \right) \right) \right) > 1 \]

where \( t = \frac{d_i^{(n)}}{d_i^{(n)}} > 1 \)
4.7.5 Model IIc: the basic HIV model for WT and A3G-augmented cells with lower death rates for cells infected by A3G(+) viruses

\[
\frac{dT^{(+)}}{dt} = f\lambda - d_{T}T^{(+)} - k(V_{(-)} + V_{(+)}T^{(+)}) \Rightarrow f\lambda = \overline{T^{(+)}}(d_{T} + k(V_{(-)} + V_{(+)})) \tag{AIIc-1}
\]

\[
\frac{dT^{(wrt)}}{dt} = (1-f)\lambda - d_{T}T^{(wrt)} - k(V_{(-)} + V_{(+)}T^{(wrt)}) \Rightarrow (1-f)\lambda = \overline{T^{(wrt)}}(d_{T} + k(V_{(-)} + V_{(+)})) \tag{AIIc-2}
\]

\[
\frac{dI^{(+)}}{dt} = kV_{(-)}T^{(+)} - d_{I_{(-)}}I^{(+)} \Rightarrow I^{(+)} = \frac{kV_{(-)}\overline{T^{(+)}}}{d_{I_{(-)}}} \tag{AIIc-3}
\]

\[
\frac{dI^{(wrt)}}{dt} = kV_{(+)}T^{(wrt)} - d_{I_{(+)}I^{(wrt)} \Rightarrow I^{(wrt)} = \frac{kV_{(+)}\overline{T^{(wrt)}}}{d_{I_{(+)}}} \tag{AIIc-4}
\]

\[
\frac{dI^{(-)}}{dt} = kV_{(-)}T^{(wrt)} - d_{I_{(-)}}I^{(wrt)} \Rightarrow I^{(-)} = \frac{kV_{(-)}\overline{T^{(wrt)}}}{d_{I_{(-)}}} \tag{AIIc-5}
\]

\[
\frac{dI^{(wrt)}}{dt} = kV_{(+)}T^{(wrt)} - d_{I_{(+)}I^{(wrt)} \Rightarrow I^{(wrt)} = \frac{kV_{(+)}\overline{T^{(wrt)}}}{d_{I_{(+)}}} \tag{AIIc-6}
\]

\[
\frac{dV^{(-)}}{dt} = p^{(+))(NI^{(+)_{(-)}} + cNI^{(+)_{(+))}) + p^{(wrt)}(NI^{(wrt)}_{(-)} + cNI^{(wrt)}_{(+))} - d_{V}V_{(-)} \Rightarrow \tag{AIIc-7}
\]

\[
\overline{V_{(-)}} = \frac{N}{d_{V}} \left( p^{(+))(I^{(+)_{(-)}} + cI^{(+)_{(+))}) + p^{(wrt)}(I^{(wrt)}_{(-)} + cI^{(wrt)}_{(+))} \right) \tag{AIIc-7}
\]

\[
\frac{dV^{(+)}}{dt} = (1 - p^{(+)})(NI^{(+)_{(-)}} + cNI^{(+)_{(+))}) + (1 - p^{(wrt)})(NI^{(wrt)}_{(-)} + cNI^{(wrt)}_{(+))} - d_{V}V_{(+)} \Rightarrow \tag{AIIc-8}
\]

\[
\overline{V_{(+)}} = \frac{N}{d_{V}} \left( (1 - p^{(+)})(I^{(+)_{(-)}} + cI^{(+)_{(+))}) + (1 - p^{(wrt)})(I^{(wrt)}_{(-)} + cI^{(wrt)}_{(+))} \right) \tag{AIIc-8}
\]
(Allc-3) & (Allc-7) \rightarrow \overline{I}_{(−)}^{(+)} = \alpha \left( p^{(+)}(\overline{I}_{(−)}^{(+)} + c\overline{I}_{(+)}) + p^{(wr)}(\overline{I}_{(−)}^{(wr)} + c\overline{I}_{(+)}) \right) \\
\text{where } \alpha = \frac{kT^{(+)}N}{d_{I_{(−)}}d_{I_{(+)}}}

(Allc-4) & (Allc-8) \rightarrow \\
\overline{I}_{(−)}^{(+)} = \beta \left( (1 - p^{(+)})(\overline{I}_{(−)}^{(+)}) + (1 - p^{(wr)})(\overline{I}_{(−)}^{(wr)} + c\overline{I}_{(+)}) \right) \\
\text{where } \beta = \frac{kT^{(+)}N}{d_{I_{(−)}}d_{I_{(+)}}}

(Allc-5) & (Allc-7) \rightarrow \overline{I}_{(−)}^{(wr)} = \varepsilon \left( p^{(+)}(\overline{I}_{(−)}^{(+)}) + p^{(wr)}(\overline{I}_{(−)}^{(wr)} + c\overline{I}_{(+)}) \right) \\
\text{where } \varepsilon = \frac{kT^{(wr)}N}{d_{I_{(−)}}d_{I_{(+)}}}

(Allc-6) & (Allc-8) \rightarrow \\
\overline{I}_{(−)}^{(wr)} = \delta \left( (1 - p^{(+)})(\overline{I}_{(−)}^{(+)}) + (1 - p^{(wr)})(\overline{I}_{(−)}^{(wr)} + c\overline{I}_{(+)}) \right) \\
\text{where } \delta = \frac{kT^{(wr)}N}{d_{I_{(−)}}d_{I_{(+)}}}

(Allc-1) & (Allc-2) \rightarrow \frac{T^{(+)}}{T^{(wr)}} = \frac{f}{1 - f}

(Allc-9) & (Allc-11) \rightarrow \frac{\overline{I}_{(−)}^{(+)}}{\overline{I}_{(−)}^{(wr)}} = \frac{\alpha}{\varepsilon} = \frac{T^{(+)}}{T^{(wr)}}

(Allc-10) & (Allc-12) \rightarrow \frac{\overline{I}_{(−)}^{(+)}}{\overline{I}_{(−)}^{(wr)}} = \frac{\beta}{\delta} = \frac{T^{(+)}}{T^{(wr)}}

(Allc-9) & (Allc-10) & (Allc-14) & (Allc-15) \rightarrow 

(A11c-9)
\[ I_{(\rightarrow)}^{(+)} + c I_{(\rightarrow)}^{(\rightarrow)} = \alpha \left( \left( p_{(\rightarrow)} + \frac{\beta}{\alpha} c(1 - p_{(\rightarrow)}) \right) (I_{(\rightarrow)}^{(+)} + c I_{(\rightarrow)}^{(\rightarrow)}) + \left( p_{(\rightarrow)}^{(\rightarrow)} + \frac{\beta}{\alpha} c(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) (I_{(\rightarrow)}^{(\rightarrow)} + c I_{(\rightarrow)}^{(\rightarrow)}) \right) \]
\[ = \alpha \left( \left( p_{(\rightarrow)} + \frac{d_{l_{(\rightarrow)}}}{d_{t_{(\rightarrow)}}} c(1 - p_{(\rightarrow)}) \right) + \left( p_{(\rightarrow)}^{(\rightarrow)} + \frac{d_{l_{(\rightarrow)}}}{T} c(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \right) (I_{(\rightarrow)}^{(+)} + c I_{(\rightarrow)}^{(\rightarrow)}) \]
\[ 1 = \alpha \left( \left( p_{(\rightarrow)} + \frac{d_{l_{(\rightarrow)}}}{d_{t_{(\rightarrow)}}} c(1 - p_{(\rightarrow)}) \right) + \frac{T_{(\rightarrow)}}{T_{(\rightarrow)}} \left( p_{(\rightarrow)}^{(\rightarrow)} + \frac{d_{l_{(\rightarrow)}}}{d_{t_{(\rightarrow)}}} c(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \right) \]
\[ \Rightarrow \frac{kT_{(\rightarrow)} N}{d_{l_{(\rightarrow)}} d_{t_{(\rightarrow)}}} \left( p_{(\rightarrow)} + wc(1 - p_{(\rightarrow)}) \right) + \frac{1 - f}{f} \left( p_{(\rightarrow)}^{(\rightarrow)} + wc(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \]

(AIIc-13) & (AIIc-16) \Rightarrow \]
\[ \overline{T_{(\rightarrow)}} = \frac{d_{l_{(\rightarrow)}} d_{t_{(\rightarrow)}}}{k \left( \left( p_{(\rightarrow)} + wc(1 - p_{(\rightarrow)}) \right) + \frac{1 - f}{f} \left( p_{(\rightarrow)}^{(\rightarrow)} + wc(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \right)} \]

(AIIc-1) & (AIIc-17) \Rightarrow \]
\[ \overline{V_{(\rightarrow)}^{(\rightarrow)}} = \frac{f \lambda k N}{k \left( \left( p_{(\rightarrow)} + wc(1 - p_{(\rightarrow)}) \right) + \frac{1 - f}{f} \left( p_{(\rightarrow)}^{(\rightarrow)} + wc(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \right)} - d_{l_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}} \]
\[ = \frac{f \lambda k N}{d_{l_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}} k} \]
\[ \Rightarrow R_{sc} = \frac{f \lambda k N}{d_{l_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}} d_{t_{(\rightarrow)}}} \]
\[ = R_0 \left( f \left( p_{(\rightarrow)} + wc(1 - p_{(\rightarrow)}) \right) + (1 - f) \left( p_{(\rightarrow)}^{(\rightarrow)} + wc(1 - p_{(\rightarrow)}^{(\rightarrow)}) \right) \right) > 1 \]

where \( w = \frac{d_{l_{(\rightarrow)}}}{d_{t_{(\rightarrow)}}} > 1 \)
4.7.6 Model III: the basic HIV model for WT and A3G-augmented cells with auto-apoptosis capability

\[
\frac{dT^{(+)}}{dt} = f\lambda - d_T T^{(+)} - k(V_{(-)} + V_{(+)})T^{(+)} \Rightarrow f\lambda = \bar{t}^{(+)}(d_T + k(V_{(-)} + V_{(+)}) ) \tag{AIII-1}
\]

\[
\frac{dT^{(wt)}}{dt} = (1 - f)\lambda - d_T T^{(wt)} - k(V_{(-)} + V_{(+)})T^{(wt)} \Rightarrow
\]

\[(1 - f)\lambda = \bar{t}^{(wt)}(d_T + k(V_{(-)} + V_{(+)}) ) \tag{AIII-2}
\]

\[
\frac{dI^{(+)}}{dt} = rkV_{(-)} T^{(+)} - d_I I^{(+)} \Rightarrow \bar{I}^{(+)}/d_I = rkV_{(-)} T^{(+)} \tag{AIII-3}
\]

\[
\frac{dI^{(+)}}{dt} = rkV_{(+)} T^{(+)} - d_I I^{(+)} \Rightarrow \bar{I}^{(+)}/d_I = rkV_{(+)} T^{(+)} \tag{AIII-4}
\]

\[
\frac{dI^{(wt)}}{dt} = kV_{(-)} T^{(wt)} - d_I I^{(wt)} \Rightarrow \bar{I}^{(wt)}/d_I = kV_{(-)} T^{(wt)} \tag{AIII-5}
\]

\[
\frac{dI^{(wt)}}{dt} = kV_{(+)} T^{(wt)} - d_I I^{(wt)} \Rightarrow \bar{I}^{(wt)}/d_I = kV_{(+)} T^{(wt)} \tag{AIII-6}
\]

\[
\frac{dV_{(-)}}{dt} = p^{(+)}(NI^{(+)}) + cNI^{(+w)} + p^{(wt)}(NI^{(wt)} + cNI^{(wt)w}) - d_v V_{(-)} \Rightarrow
\]

\[
\bar{V}_{(-)} = \frac{N}{d_v} \left( p^{(+)}(\bar{I}^{(+)}) + c\bar{I}^{(+)w} + p^{(wt)}(\bar{I}^{(wt)} + c\bar{I}^{(wt)w}) \right) \tag{AIII-7}
\]

\[
\frac{dV_{(+)}}{dt} = (1 - p^{(+)})(NI^{(+)}) + cNI^{(+)w} + (1 - p^{(wt)})(NI^{(wt)} + cNI^{(wt)w}) - d_v V_{(+)} \Rightarrow
\]

\[
\bar{V}_{(+)} = \frac{N}{d_v} \left( (1 - p^{(+)})(\bar{I}^{(+)}) + c\bar{I}^{(+)w}) + (1 - p^{(wt)})(\bar{I}^{(wt)} + c\bar{I}^{(wt)w}) \right) \tag{AIII-8}
\]
(AIII-3) & (AIII-7) \Rightarrow \overline{I}_{(e)}^{(+)} = \alpha \left( p^{(+)} (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(\text{pr})}) + p^{(\text{pr})} (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})}) \right)

where \( \alpha = \frac{r k T^{(+)} N}{d_{i} d_{v}} \)  \hspace{1cm} (AIII-9)

(AIII-4) & (AIII-8) \Rightarrow \overline{I}_{(e)}^{(\text{pr})} = \alpha \left( 1 - p^{(+)} \right) (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)}) + (1 - p^{(\text{pr})}) (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})}) \hspace{1cm} (AIII-10)

(AIII-5) & (AIII-7) \Rightarrow \overline{I}_{(e)}^{(\text{pr})} = \beta \left( p^{(+)} (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)}) + p^{(\text{pr})} (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})}) \right)

where \( \beta = \frac{k T^{(\text{pr})} N}{d_{i} d_{v}} \)  \hspace{1cm} (AIII-11)

(AIII-6) & (AIII-8) \Rightarrow \overline{I}_{(e)}^{(\text{pr})} = \beta \left( 1 - p^{(+)} \right) (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)}) + (1 - p^{(\text{pr})}) (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})}) \hspace{1cm} (AIII-12)

(AIII-1) & (AIII-2) \Rightarrow \frac{T^{(+)}}{T^{(\text{pr})}} = \frac{f}{1 - f} \hspace{1cm} (AIII-13)

(AIII-9) & (AIII-11) \Rightarrow \frac{\overline{I}_{(e)}^{(\text{pr})}}{\overline{I}_{(e)}^{(+)} } = \frac{\alpha}{\beta} = \alpha \frac{T^{(+)}}{T^{(\text{pr})}} \hspace{1cm} (AIII-14)

(AIII-10) & (AIII-12) \Rightarrow \frac{\overline{I}_{(e)}^{(\text{pr})}}{\overline{I}_{(e)}^{(+)} } = \frac{\alpha}{\beta} = \alpha \frac{T^{(+)}}{T^{(\text{pr})}} \hspace{1cm} (AIII-15)

(AIII-9) & (AIII-10) & (AIII-14) & (AIII-15) \Rightarrow \overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)}

= \alpha \left( p^{(+)} + c (1 - p^{(+)}) \right) (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)} ) + \left( p^{(\text{pr})} + c (1 - p^{(\text{pr})}) \right) (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})} ) \hspace{1cm} (AIII-16)

= \alpha \left( p^{(+)} + c (1 - p^{(+)}) \right) (\overline{I}_{(e)}^{(+)} + c \overline{I}_{(e)}^{(+)} ) + \left( p^{(\text{pr})} + c (1 - p^{(\text{pr})}) \right) \frac{1}{r} \frac{T^{(\text{pr})}}{T^{(+)}} (\overline{I}_{(e)}^{(\text{pr})} + c \overline{I}_{(e)}^{(\text{pr})} )
1 = \alpha \left( p^{(+)} + c(1 - p^{(+)})) + \frac{1}{r} \frac{T^{(w)}}{T^{(s)}} (p^{(w)} + c(1 - p^{(w)})) \right)

\Rightarrow

\frac{kN}{d_idVy} \left( rT^{(s)} (p^{(+)} + c(1 - p^{(+)})) + T^{(w)} (p^{(w)} + c(1 - p^{(w)})) \right)

(AIII-13) & (AIII-16) ⇒

\frac{T^{(s)}}{kN} \left( r(p^{(+)} + c(1 - p^{(+)})) + \frac{1-f}{f} (p^{(w)} + c(1 - p^{(w)})) \right)

(AIII-17)

(AIII-1) & (AIII-17) ⇒

\frac{V^{(-)} + V^{(s)}}{kN} = \frac{f\lambda}{kT^{(s)}} - \frac{d_t}{k}

f\lambda kN \left( r(p^{(+)} + c(1 - p^{(+)})) + \frac{1-f}{f} (p^{(w)} + c(1 - p^{(w)})) \right) - d_idTd_v

\Rightarrow R_s = \frac{f\lambda kN \left( r(p^{(+)} + c(1 - p^{(+)})) + \frac{1-f}{f} (p^{(w)} + c(1 - p^{(w)})) \right)}{d_idTd_v}

R_0 \left( rf(p^{(+)} + c(1 - p^{(+)})) + (1 - f)(p^{(w)} + c(1 - p^{(w)})) \right) > 1
4.7.7 Model IV: the basic HIV model for A3G-augmented cells overexpressing A3G at low and high levels

\[
\frac{dI^{(hi)}}{dt} = f \lambda - d_{f}I^{(hi)} - k(V_{(-)} + V_{(+)})I^{(hi)} \Rightarrow f \lambda = T^{(hi)}(d_{f} + k(V_{(-)} + V_{(+)}) \quad \text{(AIV-1)}
\]

\[
\frac{dI^{(lo)}}{dt} = (1 - f) \lambda - d_{f}I^{(lo)} - k(V_{(-)} + V_{(+)})I^{(lo)} \Rightarrow (1 - f) \lambda = T^{(lo)}(d_{f} + k(V_{(-)} + V_{(+)}) \quad \text{(AIV-2)}
\]

\[
\frac{dI^{(+hi)}}{dt} = kV_{(-)}T^{(hi)} - d_{I}I^{(+hi)} \Rightarrow T^{(+hi)} = \frac{kV_{(-)}T^{(hi)}}{d_{I}} \quad \text{(AIV-3)}
\]

\[
\frac{dI^{(+lo)}}{dt} = kV_{(+)}T^{(lo)} - d_{I}I^{(+lo)} \Rightarrow T^{(+lo)} = \frac{kV_{(+)}T^{(lo)}}{d_{I}} \quad \text{(AIV-4)}
\]

\[
\frac{dI^{(-hi)}}{dt} = kV_{(-)}T^{(hi)} - d_{I}I^{(-hi)} \Rightarrow I^{(-hi)} = \frac{kV_{(-)}T^{(hi)}}{d_{I}} \quad \text{(AIV-5)}
\]

\[
\frac{dI^{(-lo)}}{dt} = kV_{(+)}T^{(lo)} - d_{I}I^{(-lo)} \Rightarrow I^{(-lo)} = \frac{kV_{(+)}T^{(lo)}}{d_{I}} \quad \text{(AIV-6)}
\]

\[
\frac{dV_{(-)}}{dt} = p^{(hi)}(N^{(hi)}_{(-)} + cN^{(hi)}_{(+)} + p^{(lo)}(N^{(lo)}_{(-)} + cN^{(lo)}_{(+)} - d_{V}V_{(-)} \Rightarrow \quad \text{(AIV-7)}
\]

\[
V_{(-)} = \frac{N}{d_{V}} \left( p^{(hi)}(I^{(hi)}_{(-)} + cI^{(hi)}_{(+)} + p^{(lo)}(I^{(lo)}_{(-)} + cI^{(lo)}_{(+)} \right) \quad \text{(AIV-7)}
\]

\[
\frac{dV_{(+)}}{dt} = (1 - p^{(hi)})(N^{(hi)}_{(-)} + cN^{(hi)}_{(+)} + (1 - p^{(lo)})(N^{(lo)}_{(-)} + cN^{(lo)}_{(+)} - d_{V}V_{(+)} \Rightarrow \quad \text{(AIV-8)}
\]

\[
V_{(+)} = \frac{N}{d_{V}} \left( (1 - p^{(hi)})(I^{(hi)}_{(-)} + cI^{(hi)}_{(+)} + (1 - p^{(lo)})(I^{(lo)}_{(-)} + cI^{(lo)}_{(+)} \right) \quad \text{(AIV-8)}
\]
(AIV-3) & (AIV-7) \Rightarrow I_{(-)}^{(+hi)} = \alpha \left( p^{(+hi)} (I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)}) + p^{(+lo)} (I_{(-)}^{(+lo)} + c I_{(+)}^{(+lo)}) \right)

where \( \alpha = \frac{kT^{(+hi)} N}{d_j d_v} \)

(AIV-4) & (AIV-8) \Rightarrow

I_{(+)}^{(+hi)} = \alpha \left( 1 - p^{(+hi)} \right) \left( I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)} \right) + \left( 1 - p^{(+lo)} \right) \left( I_{(-)}^{(+lo)} + c I_{(+)}^{(+lo)} \right)

(AIV-5) & (AIV-7) \Rightarrow I_{(-)}^{(+lo)} = \beta \left( p^{(+hi)} (I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)}) + p^{(+lo)} (I_{(-)}^{(+lo)} + c I_{(+)}^{(+lo)}) \right)

where \( \beta = \frac{kT^{(+lo)} N}{d_j d_v} \)

(AIV-6) & (AIV-8) \Rightarrow

I_{(+)}^{(+lo)} = \beta \left( 1 - p^{(+hi)} \right) \left( I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)} \right) + \left( 1 - p^{(+lo)} \right) \left( I_{(-)}^{(+lo)} + c I_{(+)}^{(+lo)} \right)

(AIV-1) & (AIV-2) \Rightarrow \frac{T^{(+hi)}}{T^{(+lo)}} = f

(AIV-9) & (AIV-11) \Rightarrow \frac{I_{(-)}^{(+hi)}}{I_{(-)}^{(+lo)}} = \alpha = \frac{T^{(+hi)}}{T^{(+lo)}}

(AIV-10) & (AIV-12) \Rightarrow \frac{I_{(+)}^{(+hi)}}{I_{(+)}^{(+lo)}} = \beta = \frac{T^{(+hi)}}{T^{(+lo)}}

(AIV-9) & (AIV-10) & (AIV-14) & (AIV-15) \Rightarrow

I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)}

= \alpha \left( p^{(+hi)} + c(1 - p^{(+hi)}) \right) \left( I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)} \right) + \left( 1 - p^{(+lo)} \right) \left( I_{(-)}^{(+lo)} + c I_{(+)}^{(+lo)} \right)

= \alpha \left( p^{(+hi)} + c(1 - p^{(+hi)}) \right) + \left( 1 - p^{(+lo)} \right) \frac{T^{(+lo)}}{T^{(+hi)}} \left( I_{(-)}^{(+hi)} + c I_{(+)}^{(+hi)} \right)
\[1 = \alpha \left( p^{(\text{hi})} + c(1 - p^{(\text{hi})}) \right) + \frac{T^{(\text{lo})}}{T^{(\text{hi})}} \left( p^{(\text{lo})} + c(1 - p^{(\text{lo})}) \right)\]

\[\Rightarrow \quad \frac{kN}{d_i d_v} \left( p^{(\text{hi})} + c(1 - p^{(\text{hi})}) \right) + \frac{1-f}{f} \left( p^{(\text{lo})} + c(1 - p^{(\text{lo})}) \right)\]

(AIV-13) & (AIV-16) \Rightarrow

\[\frac{T^{(\text{hi})}}{T^{(\text{hi})}} = \frac{d_i d_v}{kN \left( p^{(\text{hi})} + c(1 - p^{(\text{hi})}) \right) + \frac{1-f}{f} \left( p^{(\text{lo})} + c(1 - p^{(\text{lo})}) \right)}\]  

(AIV-17)

(AIV-1) & (AIV-17) \Rightarrow

\[\frac{V^{(\text{lo})}}{V^{(\text{hi})}} = \frac{f \lambda kN}{kT^{(\text{hi})}} - \frac{d_T}{k}\]

\[= \frac{f \lambda kN \left( p^{(\text{hi})} + c(1 - p^{(\text{hi})}) \right) + \frac{1-f}{f} \left( p^{(\text{lo})} + c(1 - p^{(\text{lo})}) \right) - d_i d_r d_v}{d_i d_r d_v} > 0\]  

(AIV-18)

\[\Rightarrow R_i = \frac{f \lambda kN \left( p^{(\text{hi})} + c(1 - p^{(\text{hi})}) \right) + \frac{1-f}{f} \left( p^{(\text{lo})} + c(1 - p^{(\text{lo})}) \right)}{d_i d_r d_v} > 1\]
Chapter 5. Mechanistic Models Predict Efficacy of CCR5-Deficient Stem Cell Transplants in HIV Patient Populations
5.1 Summary

Combination anti-retroviral therapy (cART) effectively suppresses viral load in HIV-infected individuals, but it does not eradicate the virus from the body and hence does not lead to cure. Stem cell therapy represents a potential promising alternative to cART. In 2008, the Berlin patient underwent a bone marrow transplant from a CCR5−/− donor. Since then, he has shown no signs of active HIV-1 replication in the absence of cART. However, finding such a matched donor for each HIV patient is challenging; instead, ex vivo knockout of CCR5 in patients’ or matched donors’ stem cells pre-transplantation could provide HIV-resistance to progeny target cells and lead to cure. In this chapter, we investigate key logistics of successful treatments, e.g. percentage of augmented stem cells in the post-transplant chimeric immune system. We built and validated a mechanistic model of HIV/AIDS that includes major players of infection, reproduces the complete course of the disease and can simulate crucial components of clinical treatments such as cART, irradiation, host recovery, gene augmentation, and donor chimerism. Using clinical data from more than 170 cART-naïve HIV-infected individuals, we created virtual populations of HIV patients to predict performance of CCR5-deficient stem-cell therapies in virtual clinical trials and to explore interpatient variability in response to these therapies. We successfully validated our model against a published clinical study of CCR5-modified T cell therapy. Our model predicted that donor chimerism must be at least 75%, i.e. CCR5-modified stem cells must be > 75% to achieve 90% probability of cure across the patient populations.
5.2 Methods

5.2.1 Description of the model

Our model is built from previous well-established models of HIV dynamics by others [146-149,177,178,187,188,205-208] and models of host-pathogen molecular interaction and therapy from our lab [2,3,156]. Previous models of HIV dynamics could predict some features of in vivo HIV/AIDS, such as the biphasic viral decay after starting cART. However they also have shortcomings including that most models assume that the disease is at a steady state, do not reproduce the late-stage viremia, focus only on acute or chronic infection, fail to reproduce the clinical results for both treated and untreated patients, have unrealistic and/or time-dependent parameter values, and are trained with limited viral load and/or CD4+ data. To address these issues, we built a mechanistic model of HIV/AIDS that captures the complete course of the disease, reproduces both treated and untreated patients, has constant parameters that take biologically relevant values, is trained with viral load, CD4+ and CD8+ data, uses cohort data to capture interpatient variability and has the ability to test anti-HIV CCR5-based therapies. The model includes key components of the infection: the virus, and multiple immune cell types: CD4+ T-cells, monocytes/macrophages, latently infected CD4+ T-cells, and CD8+ Cytotoxic T lymphocytes (CTLs). We also track wild type (WT) cells and augmented cells, in which CCR5 has been rendered dysfunctional, e.g. has been knocked out by zinc fingers. The model was developed in the Matlab SimBiology software package. We formulated the model as a system of coupled ordinary differential equations as follows:
\[\frac{dT^W_U}{dt} = (1 - f_T) s_T + \frac{P_T}{c_T + V} V T^W_U + r_T T^W_U \left(1 - \frac{T}{T_{\text{max}}^U}\right) - k_{VT} V T^W_U - k_{MT} M_I T^W_U - d_{T^W_U}, \]

\[\frac{dT^A_U}{dt} = f_T s_T + \frac{P_T}{c_T + V} V T^A_U + r_T T^A_U \left(1 - \frac{T}{T_{\text{max}}^U}\right) - \alpha_{CCR5} k_{VT} V T^A_U - \epsilon_{CCR5} k_{MT} M_I T^A_U - d_{T^A_U}, \]

\[\frac{dT^W_I}{dt} = (1 - k_L)(k_{VT} V T^W_U + k_{MT} M_I T^W_U) + a_L T^W_I - k_{ET} E T^W_I - d_{T^W_I}, \]

\[\frac{dT^A_I}{dt} = (1 - k_L)(\alpha_{CCR5} k_{VT} V T^A_U + \epsilon_{CCR5} k_{MT} M_I T^A_U) + a_L T^A_I - k_{ET} E T^A_I - d_{T^A_I}, \]

\[\frac{dT^W_L}{dt} = k_L(k_{VT} V T^W_U + k_{MT} M_I T^W_U) - a_L T^W_L - d_{T^W_L}, \]

\[\frac{dT^A_L}{dt} = k_L(\alpha_{CCR5} k_{VT} V T^A_U + \epsilon_{CCR5} k_{MT} M_I T^A_U) - a_L T^A_L - d_{T^A_L}, \]

\[\frac{dM^W_U}{dt} = (1 - f_M)s_M + \frac{P_M}{c_M + V} V M^W_U - k_{VM} V M^W_U - d_{M^W_U}, \]

\[\frac{dM^A_U}{dt} = f_M s_M + \frac{P_M}{c_M + V} V M^A_U - \beta_{CCR5} k_{VM} V M^A_U - d_{M^A_U}, \]

\[\frac{dM^W_I}{dt} = k_{VM} V M^W_U - k_{EM} E M^W_I - d_{M^W_I}, \]

\[\frac{dM^A_I}{dt} = \beta_{CCR5} k_{VM} V M^A_U - k_{EM} E M^A_I - d_{M^A_I}, \]

\[\frac{dE}{dt} = s_E + P_E \frac{T}{c_E + T} T_I E - d_{E}, \]

\[\frac{dV}{dt} = n_T(T^W_I + T^A_I) + n_M(M^W_I + M^A_I) - d_v V. \]

All the variables are capitalized and represent concentrations of virus or different cell types and all the parameters are lower-cased. Here, \( T \) represents the CD4+ T cell population and it can take two superscripts and three subscripts. The superscripts \( W \) and \( A \) represent the WT and augmented cell populations and the superscripts \( U, I, \) and \( L \) denote different cell states: uninfected, productively infected and latently infected, respectively. For example, \( T^W_I \) represents WT latently infected CD4+ T cells; \( T^A_I \)
represents augmented CD4+ T cells in all states including uninfected, productively infected and latently infected; and $T_I$ represents all the WT and augmented infected CD4+ T cells. Similarly, $M$ represents the macrophage population (monocytes, tissue macrophages, and monocyte-derived macrophages that have been reported as susceptible to SIV and HIV infection) and it can take two superscripts, WT and augmented, and two subscripts, uninfected and infected. In our model, $E$ represents the CD8+ T cell population or cytotoxic T lymphocytes (CTLs) to simulate the role of the immune response and $V$ denotes free virus. The change in the concentration of uninfected WT CD4+ T cell population, as described in Equation (1), is a function of six terms: 1) production from the bone marrow in which $s_T$ and $f_T$ denote the production rate of CD4+ T cells and the percentage of augmented stem cells, which produce CCR5$^{-/-}$ CD4+ T cells; 2) proliferation term due to an immune response to HIV infection, described by a Michaelis–Menten equation; 3) homeostatic proliferation term described by a logistic growth function with a carrying capacity of $\tau_{\text{max}}$; 4) infection by HIV, which is proportional to the concentration of free virus and WT CD4+ T cells and hence the infectivity rate constant is denoted by $k_{VT}$. 5) cell-cell infection of WT CD4+ T cells by infected macrophages, described by a term proportional to the concentration of infected macrophages and WT CD4+ T cells and hence the infectivity rate constant is denoted by $k_{MT}$; 6) death of cells at a specific rate, denoted by $d_T$. Similarly, Equation (2) describes the change in the concentration of uninfected CCR5$^{-/-}$ CD4+ T cell population, in which $\alpha_{\text{CCR5}}$ and $\epsilon_{\text{CCR5}}$ denote the reduction in infection of CCR5$^{-/-}$ CD4+ T cells by free virus and infected macrophages, respectively.
After HIV entry into CD4+ T cells and integration of the viral DNA into the host genome, most of the cells become productively infected and produce viruses. However, a small fraction of infected cells does not produce viral proteins and becomes latently infected; this fraction is denoted by $k_L$ in Equations (3-6), which describe the rate of change in concentrations of productively infected or latently infected cells in both WT and augmented CD4+ T cells populations. The first and second terms in Equations (3) and (4) represent the production of infected CD4+ T cells from uninfected cells and activation of latently infected cells to productively infected cells, respectively. The last two terms represent the killing of infected CD4+ T cells by CTLs at a rate of $k_{ET}$ and the natural death of infected cells at a rate of $d_{TI}$. In Equations (5) and (6), the first term represents the production of latently infected CD4+ T cells, whereas the second and third terms model the loss of these cells due to activation at a rate of $a_L$ or death at a rate of $d_L$.

Uninfected WT and augmented macrophages are described by Equations (7) and (8), where production is represented by a zero-order rate constant $s_M$, proliferation due to HIV infection is reflected by a Michaelis–Menten equation, infection by free virus is controlled at an infectivity rate constant $k_{VM}$, and finally death of macrophages is modeled by a first-order term at a rate of $d_M$. Infected macrophages are described by Equations (9) and (10), where the first term represents the production of infected cells, the second term models killing of infected macrophages by CTLs at a rate of $k_{EM}$ and the third terms represents natural death of infected macrophages at a rate of $d_{MI}$. In Equations (8) and (10), $\beta_{CCR5}$ denotes the reduction in infection of CCR5-/- macrophages by HIV.

Equation (11) describes the change in the concentration of CD8+ T cell population as a function of three terms: production at a rate of $s_E$, proliferation due to infection, and
death at a rate of $d_E$. The proliferation term is proportional to the number of infected CD4+ T cells but also includes a Michaelis–Menten term to describe the observation that CD8+ cells expand after HIV infection until the middle of chronic stage of HIV disease and tend to decline afterwards (Figure 5-1, I-L). The concentration of virus in plasma in Equation (12) changes as HIV is produced from infected T cells and macrophages at $n_T$ and $n_M$ viruses/cell, respectively, and is cleared at a rate of $d_V$. Note that for both T cells and macrophages, the underlying parameters including proliferation rates, rates of killing by CTLs and death rates are assumed to be the same for CCR5+/- and WT populations.

5.2.2 Model calibration

We used scatter search methodology [209] to calibrate the model parameters to match the average profiles of clinical measurements for each subpopulation. We calculated the 2-norm of the differences between simulation results and the clinical data for viral load, CD4+ and CD8+ T cell counts and defined the objective function as the sum of the calculated 2-norms. In our optimization, we set the size of the reference set = 20, the size of the initial set = 200, the number of sub-ranges per parameter = 8, and the maximum number of regenerations = 3. During the optimization, for each generated vector of parameter values, we calculated $s_T$, $s_M$, $s_E$ and $\tau_{\text{max}}$ as follows:

\[
\tau_{\text{max}} = T_0,
\]

\[
s_T = T_0 \times d_T,
\]

\[
s_M = M_0 \times d_M,
\]

\[
s_E = E_0 \times d_E,
\]

where $T_0$, $M_0$, and $E_0$ take their values for each population from Table 5-1.
Table 5-1. Initial values for immune cells in each subpopulation.

<table>
<thead>
<tr>
<th></th>
<th>$\text{AIDS} \leq 3.5$</th>
<th>$3.5 &lt; \text{AIDS} \leq 7$</th>
<th>$7 &lt; \text{AIDS} \leq 9$</th>
<th>$9 &lt; \text{AIDS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_0$ (Ave. CD4+ T cells at $t = 0$)</td>
<td>907.8</td>
<td>826.6</td>
<td>1019.9</td>
<td>994.9</td>
</tr>
<tr>
<td>$M_0$ (Ave. Macrophages at $t = 0$)</td>
<td>426.3</td>
<td>381.9</td>
<td>398.6</td>
<td>483.0</td>
</tr>
<tr>
<td>$E_0$ (Ave. CD8+ T cells at $t = 0$)</td>
<td>739.5</td>
<td>735.6</td>
<td>693.7</td>
<td>782.2</td>
</tr>
</tbody>
</table>

We ran the optimization algorithm 200 times and when the algorithm converged, we included the vector with the lowest objective function in our “optimized” set. The values of objective function for this set are shown in Figure 5-2. Then, we selected the 100 vectors with the lowest objective functions as the 100 best parameter sets and used them for the rest of our analysis.

5.2.3 Virtual population

Using the 5th and 95th percentiles of the clinical data for viral load, CD4+ and CD8+ T cell counts in each subpopulation, we created 4 virtual patient populations. To do so, we used a virtual population development methodology introduced in [210-212], which includes three main steps:

1) Patient generation: We used the 100 best parameter sets and randomly varied parameters around their values by maximum of ±5%. Then, for each generated vector of parameter values, we calculated $s_T$, $s_M$, $s_E$ and $\tau_{\text{max}}$ as follows:

$$
\tau_{\text{max}} = T_0 \times (1 + 1.8 \times (u_T - 0.5)),
$$

$$
s_T = T_0 \times (1 + 1.8 \times (u_T - 0.5)) \times d_T,
$$

$$
s_M = M_0 \times (1 + 1.8 \times (u_M - 0.5)) \times d_M,
$$

198
\[ s_E = E_0 \times (1 + 1.8 \times (u_E - 0.5)) \times d_E, \]

where \( T_0, M_0, \) and \( E_0 \) take their values for each population form Table 5-1 and \( u_T, u_M, \) and \( u_E \) are uniform random variables in the range [0, 1]. Random variables are multiplied by 1.8 to make sure the entire range of clinical baseline values of CD4+ T cells, macrophages and CD8+ T cells at \( t = 0 \) are covered.

2) Patient selection: Then we ran the mathematical model and only selected those virtual patients whose simulated profiles lie within the 5th and 95th percentiles of clinical data for the corresponding subpopulation; At the end of this step, we had a 1000 virtual patients in each population.

3) Patient validation: We also validated the virtual populations against another clinical dataset for decay of viremia (which was not used for training) to increase the confidence in the predicative ability of the model.

5.3 Results

5.3.1 Clinical data (patient cohorts)

To train and validate our model, we used the publicly available dataset (release P20 at http://www.ntis.gov/) of the Multicenter AIDS Cohort Study (MACS) \[213,214\], which includes longitudinal semi-annual clinical measurements of patients from 1984-2007. The dataset has information from 6972 individuals including seronegatives, seropositives, and seroconverters. In this study, we were interested in cART-naïve HIV-infected individuals, whose date of HIV seroconversion and date of initial AIDS diagnosis were (approximately) known. In the MACS cohort, we began our analysis with seroconverters (585 patients) and seroprevalents with known date of seroconversion (59 patients) and we
selected patients whose first visit with seropositive status was less than a year apart from the last visit with seronegative status (486 patients). The date of seroconversion was calculated as the midpoint between the two visits. cART was broadly introduced in the US in 1996, and therefore we removed all the patients whose seroconversion date or initial AIDS diagnosis date was after 1996. The patients with unknown AIDS diagnosis date were also removed from our study. We then categorized patients based on their progression time to AIDS into 4 subgroups: AIDS occurring before 3.5 years (rapid progressors, \( N = 32 \)), between 3.5 to 7 years (\( N = 61 \)), between 7 to 9 years (\( N = 39 \)), and after 9 years since seroconversion (slow progressors, \( N = 40 \)). See Table 5-2 for more information on patient characteristics and their clinical measurements.

**Table 5-2.** Subject characteristics and clinical measurements.

<table>
<thead>
<tr>
<th></th>
<th>All ((N = 172))</th>
<th>AIDS (\leq 3.5) ((N = 32))</th>
<th>3.5 &lt; AIDS (\leq 7) ((N = 61))</th>
<th>7 &lt; AIDS (\leq 9) ((N = 39))</th>
<th>9 &lt; AIDS ((N = 40))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at seroconversion, median (range), (yrs)</td>
<td>32.5 (18-55.5)</td>
<td>35.5 (19-54.5)</td>
<td>35.5 (23.5-55)</td>
<td>31 (21-53)</td>
<td>31.25 (18-55.5)</td>
</tr>
<tr>
<td>Time from seroconversion to AIDS, median (range),</td>
<td>6.5* (1.5-12)</td>
<td>3 (1.5-3.5)</td>
<td>6 (4-7)</td>
<td>8.5 (7.5-9)</td>
<td>10.5* (9.5-12)</td>
</tr>
<tr>
<td>CD4+ counts (cells/(\mu)L) at seroconversion, average ± SD</td>
<td>915.1 ± 448.1</td>
<td>907.8 ± 462.6</td>
<td>826.6 ± 382.8</td>
<td>1019.9 ± 477.4</td>
<td>994.9 ± 507.0</td>
</tr>
<tr>
<td>CD4± counts (cells/(\mu)L) 1 yr post-seroconversion, average ± SD</td>
<td>615.1 ± 270.4</td>
<td>499.9 ± 293.1</td>
<td>576.3 ± 269.9</td>
<td>768.8 ± 198.4</td>
<td>647.2 ± 252.7</td>
</tr>
<tr>
<td>Viremia (log(_{10}) copies/mL) 1 yr post-seroconversion, average ± SD</td>
<td>4.62 ± 0.60</td>
<td>4.99 ± 0.32</td>
<td>4.78 ± 0.48</td>
<td>4.39 ± 0.61</td>
<td>4.24 ± 0.69</td>
</tr>
</tbody>
</table>

* excluding patients who did not progress to AIDS
There are significant variations in the clinical measurements, e.g. plasma viremia, CD3+ CD4+ and CD8+ T cell counts among individuals within each subpopulation of HIV patients (Figure 5-1); e.g. one year post-seroconversion, CD4+ T cell count ranges from 100 to 1200 cells/μL in rapid progressors (Figure 5-1E). There is also substantial variability in the baseline levels of CD4+ T cells before infection (day 0 in Figure 5-1, E-H). Between subpopulations, there is a high degree of overlap, however, rapid progressors have the highest viral loads in the shortest period of time with rapid decline in the CD4+ T cell counts, whereas slow progressors show relatively stable viral load and CD4+ T cell counts for a longer time compared to other populations. We also observed that CD8+ T cell count goes up until the middle of chronic stage of HIV disease and it tends to drop after that (Figure 5-1, I-L), suggesting homeostasis failure before the onset of AIDS [215]. Most published models of HIV dynamics only reproduce the average profiles of viral load and CD4+ T cell counts and use that model to predict the efficacy of treatments. This gives a limited view of treatment efficacy, as the huge interpatient variability is neglected. In this chapter, we build models to capture the range of variability observed in patients and use that information to run virtual clinical trials and predict the interpatient variability of treatment performances.

Figure 5-1. Clinical data from cART-naïve HIV-infected individuals in the multicenter AIDS cohort study (MACS).
From the MACS cohort, we collected viral load, CD3+ CD4+ and CD3+ CD8+ T cell counts from cART-naïve, HIV-infected individuals, whose date of HIV seroconversion and date of initial AIDS diagnosis were (approximately) known. We then categorized patients based on their progression time to AIDS into 4 subgroups: (A, E, I) AIDS occurring before 3.5 years, (B, F, J) between 3.5 to 7 years, (C, G, K) between 7 to 9 years, (D, H, L) and after 9 years since seroconversion; light colored dashed lines: clinical data from individual patients; dark colored dashed lines: 5th and 95th percentiles of the aggregated data; dark colored solid lines: average and the standard deviation of the data.
5.3.2 The “average” clinical profiles can be reproduced using parameter sets with biologically relevant, yet vastly different values

To train our model, we used average profiles of viral loads, CD4+ and CD8+ T cell counts in each subpopulation. We used the scatter-search based optimization method [209] to find the parameter values that generate curves with the best fit (the lowest error) to the clinical data (Figure 5-2). Parameter values were chosen to be in the biologically relevant ranges and consistent with previously-published measurements (Table 5-3). Parameter optimization is discussed in detail in the Methods section. The 100 best curves matching to the average viral load, CD4+ and CD8+ T-cell counts for rapid progressors are shown in Figure 5-3. These 100 best curves produce good fits to the clinical data and are tightly placed on top of each other. Inspecting the values of the 23 parameters across the 100 best parameter sets, we observe that some parameters (e.g. \( p_M \) and \( kVT \)
Figure 5-2. The values of objective function for the “optimized” set in each subpopulation.

We selected the 100 parameter vectors with the lowest objective functions as the 100 best parameter sets in each subpopulation. Model calibration is discussed in the Methods section.

Have values in a relatively tight range and therefore the model is more sensitive to these and we have greater confidence in their values, whereas others (e.g. $k_{EM}$) take values in a wider range (see Figure 5-4 for parameter values generating the 100 best fits in each subpopulation). Although none of these curves represents real patients (only the population average), this observation suggests that there might be patients with very similar longitudinal viral loads and T cell counts who might have very different immune systems and HIV infections (as represented by the different parameter values) and hence respond differently to treatments. We performed paired t-tests to identify parameters taking statistically different values across populations (used the Bonferroni method to correct for multiple comparisons) and calculated spearman correlations to test for trend across populations (Figure 5-4).
We calibrated the model to match the average viral load, CD4+ and CD8+ T cell counts in each subpopulation; symbols: average clinical profiles; light colored curves: 100 best fits to the data.

For example, we found that $n_T$, the burst size of infected T cells, takes statistically smaller values as we transition from rapid progressors to slow progressors, whereas $a_L$, activation rate of latently infected T cells, does not take significantly different values across populations. We also studied the correlation coefficients between the system parameters in all populations (Figure 5-5). In most cases, there is a weak correlation between parameters of the system, suggesting that these parameters act relatively independently. However, in all populations, we observed strong negative correlations between $n_T$ and $k_{VT}$ and between $n_M$ and $k_{VM}$, suggesting that the burst size of infected T cells and macrophages are inversely associated with the infection rate of T cells and macrophages by free virus, respectively.
Figure 5-4. Comparison of the 100 best parameter sets across subpopulations.
The boxplots represent parameter values corresponding to the 100 best curves in each subpopulation and asterisks indicate whether parameters take statistically different values in any pairs of populations. We used the Bonferroni method to correct for multiple comparisons and performed paired t-tests with $\alpha = 0.05/(23 \times 6)$. We also calculated spearman correlations to test for trend across populations ($\rho$ denotes the correlation coefficient and $p$ represents the $p$-value for testing the hypothesis of no correlation against the alternative that there is a non-zero correlation). For example, we found that $n_T$, the burst size of infected T cells, takes statistically smaller values as we move from rapid progressors to slow progressors, whereas $a_L$, activation rate of latently infected T cells, does not take statistically different values across populations.

We also looked at other variables during HIV/AIDS to qualitatively compare them with our current knowledge of the disease (Figure 5-6 and Figure 5-7). For example, the model suggests that in the first 500 days, the percentage of infected CD4+ T cells remains under 10 percent in rapid progressors and this percentage goes down as we transition to slow progressors (Figure 5-6). This is consistent with the low values for clinical measurements of infected CD4+ T cells during the chronic infection [171,216,217]. The model also suggests that percentage of infected macrophages remains less than 5 percent.
for most of the infection and the number of macrophages rises at the late stage of the
disease, explaining the rise in viral load during AIDS (Figure 5-7). The model indicated
that the majority of viremia is due to infected T cells during the acute/early chronic
infection and as the disease progresses, viruses released from infected macrophages play
a more important role and eventually constitute the majority of viral load during AIDS
(Figure 5-7), consistent with the experimental evidence that macrophages produce large
amounts of SIV even after CD4+ T cells are depleted in macaques [78].

Figure 5-5. Correlation analysis of the 100 best parameter sets in all populations.
Pearson correlation was performed to calculate correlation coefficients between parameters. In
most cases, we observed a weak correlation between parameters of the system, suggesting that
these parameters act relatively independently. However, in all populations, we observed strong
negative correlations between $n_T$ and $k_{VT}$ and between $n_M$ and $k_{VM}$, suggesting the burst size of
infected CD4+ T cells and macrophages are inversely associated with the rate constant for
infection of CD4+ T cells and macrophages by the virus, respectively.
Figure 5-6. Dynamics of productive and latently infected T cells in the 100 “average” fits. The light colored curves represent system variables corresponding to the 100 best fits to the average clinical profiles in each subpopulation; dark colored curves: average of simulation curves.

Figure 5-7. Dynamics of infected and uninfected macrophages and viremia produced by CD4+ T cells and macrophages in the 100 best “average” fits. The light colored curves represent system variables corresponding to the 100 best fits to the average clinical profiles in each subpopulation. The dark colored curves in the first and second rows represent the average of simulation curves. The total viral load consists of two components: viruses released from infected CD4+ T cells ($V_T$; light color) and viruses released from infected macrophages ($V_M$; dark color). In the third row, we separated these two components to show the role of infected CD4+ T cells and infected macrophages during the course of HIV/AIDS.
Table 5-3. Parameters of the model, their ranges for optimization, and definitions.

<table>
<thead>
<tr>
<th>Parameter (units)</th>
<th>Scale</th>
<th>Range</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T (d^{-1})$</td>
<td>Log</td>
<td>0.01-0.1</td>
<td>Maximum proliferation rate of uninfected CD4+ T cells due to infection</td>
</tr>
<tr>
<td>$c_T (\mu L^{-1})$</td>
<td>Log</td>
<td>40-4000</td>
<td>Concentration of virus, at which the proliferation rate of uninfected CD4+ T cells due to infection is half of $p_T$</td>
</tr>
<tr>
<td>$r_T (d^{-1})$</td>
<td>Linear</td>
<td>0.01-0.05</td>
<td>Proliferation rate of uninfected CD4+ T cells in the absence of infection</td>
</tr>
<tr>
<td>$p_M (d^{-1})$</td>
<td>Log</td>
<td>0.003-0.03</td>
<td>Maximum proliferation rate of uninfected macrophages due to infection</td>
</tr>
<tr>
<td>$c_M (\mu L^{-1})$</td>
<td>Log</td>
<td>1-100</td>
<td>Concentration of virus, at which the proliferation rate of uninfected macrophages due to infection is half of $p_M$</td>
</tr>
<tr>
<td>$k_{VT} (d^{-1} \mu L)$</td>
<td>Log</td>
<td>0.0001-0.005</td>
<td>Rate constant for infection of CD4+ T cells by HIV</td>
</tr>
<tr>
<td>$k_{MT} (d^{-1} \mu L)$</td>
<td>Log</td>
<td>$10^{-6}$-$10^{-2}$</td>
<td>Rate constant for infection of CD4+ T cells by infected macrophages</td>
</tr>
<tr>
<td>$k_{ET} (d^{-1} \mu L)$</td>
<td>Log</td>
<td>$10^{-5}$-$2 \times 10^{-4}$</td>
<td>Rate constant for killing of infected CD4+ T cells by CD8+ T cells</td>
</tr>
<tr>
<td>$k_{VM} (d^{-1} \mu L)$</td>
<td>Log</td>
<td>$10^{-6}$-$10^{-4}$</td>
<td>Rate constant for infection of macrophages by HIV</td>
</tr>
<tr>
<td>$k_{EM} (d^{-1} \mu L)$</td>
<td>Log</td>
<td>$10^{-8}$-$10^{-5}$</td>
<td>Rate constant for killing of infected macrophages by CD8+ T cells</td>
</tr>
<tr>
<td>$p_E (d^{-1} \mu L)$</td>
<td>Log</td>
<td>$10^{-5}$-$10^{-3}$</td>
<td>Maximum proliferation rate of CD8+ T cells due to infection</td>
</tr>
<tr>
<td>$c_E (\mu L^{-1})$</td>
<td>Linear</td>
<td>100-700</td>
<td>Conc. of CD4+ T cells, at which the proliferation rate of CD8+ T cells due to infection is half of $p_E$</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Unit</td>
<td>Value</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>$n_T$ (d⁻¹)</td>
<td>Linear</td>
<td>20-100</td>
<td>Burst size of infected CD4+ T cells</td>
</tr>
<tr>
<td>$n_M$ (d⁻¹)</td>
<td>Linear</td>
<td>1-20</td>
<td>Burst size of infected macrophages</td>
</tr>
<tr>
<td>$d_T$ (d⁻¹)</td>
<td>Linear</td>
<td>0.01-0.025</td>
<td>Death rate of uninfected CD4+ T cells</td>
</tr>
<tr>
<td>$d_{TT}$ (d⁻¹)</td>
<td>Linear</td>
<td>0.37-0.65</td>
<td>Death rate of infected CD4+ T cells</td>
</tr>
<tr>
<td>$d_L$ (d⁻¹)</td>
<td>Log</td>
<td>0.0002-0.001</td>
<td>Death rate of latently infected CD4+ T cells</td>
</tr>
<tr>
<td>$d_M$ (d⁻¹)</td>
<td>Log</td>
<td>0.0055-0.008</td>
<td>Death rate of uninfected macrophages</td>
</tr>
<tr>
<td>$d_{Mf}$ (d⁻¹)</td>
<td>Log</td>
<td>0.039-0.059</td>
<td>Death rate of infected macrophages</td>
</tr>
<tr>
<td>$d_E$ (d⁻¹)</td>
<td>Log</td>
<td>0.005-0.05</td>
<td>Death rate of CD3+ CD8+ T cells</td>
</tr>
<tr>
<td>$d_V$ (d⁻¹)</td>
<td>Linear</td>
<td>21-25</td>
<td>Clearance rate of free virus</td>
</tr>
<tr>
<td>$a_L$ (d⁻¹)</td>
<td>Log</td>
<td>$10^{-4}$-$10^{-3}$</td>
<td>Activation rate of latently infected CD4+ T cells</td>
</tr>
<tr>
<td>$k_L$</td>
<td>Log</td>
<td>$5\times10^{-5}$-$5\times10^{-4}$</td>
<td>Fraction of infected CD4+ T cells that become latently infected.</td>
</tr>
<tr>
<td>$\alpha_{CCR5}$</td>
<td>-</td>
<td>0.075 (Fixed)</td>
<td>Reduction in $k_{VT}$ for CCR5⁻/⁻ CD4+ T cells</td>
</tr>
<tr>
<td>$\beta_{CCR5}$</td>
<td>-</td>
<td>0.075 (Fixed)</td>
<td>Reduction in $k_{VM}$ for CCR5⁻/⁻ macrophages</td>
</tr>
<tr>
<td>$\epsilon_{CCR5}$</td>
<td>-</td>
<td>0.075 (Fixed)</td>
<td>Reduction in $k_{MT}$ for CCR5⁻/⁻ CD4+ T cells</td>
</tr>
</tbody>
</table>

5.3.3 Virtual patient populations capture the range of variability observed in the clinical data

To facilitate capturing the observed clinical variability in measurements from HIV-infected individuals, we created 4 virtual patient populations using the 5th and 95th
percentiles of the clinical data for each subpopulation. To do so, we used a virtual population development methodology introduced in [210-212], which includes three main steps: 1) patient generation: to explore the uncertainty in the model parameter values by creating locally randomized parameter sets around the 100 best fits, which generate potential virtual patients; 2) patient selection: to select those virtual patients whose simulated profiles lie within the range of clinical data for the corresponding subpopulation; 3) patient validation: to validate the virtual populations against another clinical dataset (which was not used for training) to increase the confidence in the predicative ability of the model. The details of virtual population development are discussed in the Methods section. Figure 5-8 shows the viral load, CD4+ and CD8+ T cell profiles of the 1000 virtual patients in all populations, completely capturing the range of clinical measurements. For validation, we used the viral decay data (represented by black symbols in Figure 5-8) from cART-treated patients, who started taking cART approximately 208 days after infection and were monitored for 48 weeks [216]. After receiving cART, all the virtual patient populations display biphasic decay, consistent with the viral decay from cART-treated patients (Figure 5-8). The decay is followed by a plateau in viral load, which is due to occasional activation of latently infected CD4+ T cells and is below the detection level of clinical assays. The temporal histogram of virtual patient profiles also demonstrates a good match between average clinical patterns and the high intensity regions, where a higher proportion of the patient profiles lies (Figure 5-9). The parameter values for the 1000 virtual patients in each population and the analysis of trends across populations are shown in Figure 5-10. Comparing Figure 5-4 and Figure 5-10, we observed that all the parameter trends have remained the same across
populations. Also, the correlation analysis shown in Figure 5-11 revealed that the associations between parameters in virtual populations have only slightly changed compared to those of parameters for the average profiles. Figure 5-12 and Figure 5-13 show other variables of the virtual patients during HIV/AIDS. For example, the average percentage of infected CD4+ T cells in the first 500 days remains under 10 percent and is lower for slower progressors, consistent with our current understanding that infected CD4+ T cells constitute a small fraction of cells during the chronic infection. In cART-treated patients, the ratio of latently infected CD4+ T cells is in the range (10^{-6}-10^{-3}), which is consistent with our current knowledge of the size of the latent reservoir [218].

**Figure 5-8. Generation of virtual populations.**
We built a virtual patient population to capture the observed clinical variability in measurements of viremia, CD4+ and CD8+ T cells in cART-naïve patients and validated it against clinical data from decay of plasma viremia in another group of cART-treated patients who received cART at \( t = 208 \) [216]; red symbols: average clinical profiles; dashed lines: 5th and 95th percentiles of clinical data; black symbols: clinical data for viral load decay; light colored curves: cART-naïve virtual patients; dark colored curves: cART-treated virtual patients.
Figure 5-9. Temporal histograms of virtual patient profiles.
The intensity increases from light gray to black. There is a good match between average clinical patterns and the high intensity regions, where a higher proportion of the patient profiles lies.

Figure 5-10. Comparison of the 1000 “virtual patient” parameter sets across subpopulations.
The boxplots represent parameter values corresponding to the 1000 virtual patients in each subpopulation and asterisks indicate whether parameters take statistically different values in any pairs of populations. We used the Bonferroni method to correct for multiple comparisons and performed paired t-tests with $\alpha = 0.05/(23\times6)$. We also calculated spearman correlations to test for trend across populations ($\rho$ denotes the correlation coefficient and $p$ represents the $p$-value for testing the hypothesis of no correlation against the alternative that there is a nonzero correlation).
Figure 5-11. Correlation analysis of the “virtual patient” parameter sets in all populations. Pearson correlation was performed to calculate correlation coefficients between parameters. Comparing Figure 5-5 and Figure 5-11 suggests that associations between parameters in virtual populations have only slightly changed compared to those of parameters for the average profiles.

Figure 5-12. Dynamics of productive and latently infected cells in virtual patients.
Figure 5-13. Dynamics of infected and uninfected macrophages and viremia produced by CD4+ T cells and macrophages in virtual patients.

The light colored curves represent cART-naïve virtual patients. The dark colored curves in the first and second rows represent the average of simulation curves. The total viral load consists of two components: viruses released from infected CD4+ T cells ($V_T$; light color) and viruses released from infected macrophages ($V_M$; dark color). In the third row, we separated these two components to show the role of infected CD4+ T cells and infected macrophages during the course of HIV/AIDS.

5.3.4 Virtual clinical trials predict the efficacy of CCR5-based stem cell therapy

CCR5-deficient stem cell therapy has appeared to functionally cure the Berlin patient of HIV, but one key question is whether CCR5-based therapy would be successful at stopping the disease in other HIV-infected individuals. To predict the efficacy of CCR5-based stem cell therapy, we simulated the complex clinical procedure for the Berlin patient as closely as possible and ran a virtual clinical trial using our validated virtual populations. To do so, we picked our subpopulation of the 1000 rapid progressors (Figure 5-14) and we assumed that they undergo the following scenario: 1) Infection at day 0: CD4+ T cells quickly drop while plasma viremia rapidly increases in these patients. 2)
cART 208 days post-infection (dpi): patients start taking cART, which stops the infection of CD4+ T cells and macrophages, and hence reduces the viral load in a biphasic way and returns T cell counts to normal levels. Eventually, the viral load plateaus and becomes undetectable. 3) Bone marrow transplant at 500 dpi: this procedure consists of bone marrow irradiation for a short period of time (7 days), followed by the engraftment of new stem cells (6 months). During the irradiation, stem cells are killed and hence no more immune cells are produced. Also, CTLs, uninfected and infected CD4+ T cells are killed, and the size of the latent reservoir is reduced by three orders of magnitude [219,220]. However, macrophages will remain intact as they mostly reside in tissues. Note that even after the irradiation is complete, some residual WT stem cells are left behind. This leads to donor chimerism after infusion of CCR5-modified stem cells. During the engraftment (507-687 dpi), we assume that a certain percentage of stem cells are CCR5-deficient and the rest are WT (donor chimerism = 50% in Figure 5-14). Also, the production levels of immune cells return to pre-transplant values. Note that all the patients remain on cART during the irradiation and engraftment periods. 4) No cART (687 dpi): six months after engraftment, CD4+ T cells are back to normal levels (similar to the cART phase, except that now 50% of the CD4+ T cells are CCR5-deficient) and viremia is below the detection level. And hence, these patients cease taking cART (infection of CD4+ T cells and macrophages is the same as the pre-cART phase).

We observe that some patients, e.g. the individual shown in green, maintain their CD4+ T cell levels and keep their viremia below the level of detection, whereas for other patients, e.g. the individual shown in red, viral load rapidly increases and CD4+ T cells start to decline, suggesting that CCR5-deficient cells slowed the infection but were not
able to completely stop it. If we only had one model with one parameter set, our answer to the question of whether CCR5-based therapy is successful at stopping HIV infection would be an all-or-none response. In reality, there is huge variability among patients and the significance of using virtual populations is that, similar to running a clinical trial, we are able to look at the inter-patient variability in response to new anti-HIV therapies in a virtual clinical trial.

**Figure 5-14. Performance prediction of CCR5-deficient stem cell therapy in rapid progressors.**

We ran a virtual clinical trial and monitored (A) viral load and (B) CD4+ T cell counts in rapid progressors to predict the efficacy of CCT5-modified stem cell transplants in this population. The virtual clinical trial includes 1) infection at day 0, 2) taking cART at 208 dpi, 3) bone marrow transplant at 500 dpi, which consists of irradiation for 7 days followed by six months of engraftment, and 4) ceasing cART at 687 dpi. A patient is considered functionally cured if one year post therapy, normalized CD4 level > 95% and viral load < 50 copies/mL and decaying; gray curves: individual patients; green curves: an example of a cured patient; red curves: an example of a patient not cured after therapy.

To calculate the probability of cure, we defined an HIV-infected patient as functionally cured if one year post-therapy, 1) 95% < CD4+ level normalized by the baseline level and 2) viral load < 50 copies/mL and decaying; otherwise the patient would not be cured. Figure 5-15A and Figure 5-15E show one year post therapy viral load and normalized CD4+ T cell counts, respectively, for any level of donor chimerism.
(percentage of CCR5-modified stem cells) in rapid progressors. The model predicts that if augmented stem cells are less than 10%, post therapy viral load is at about $10^5$ copies/mL for almost all patients and normalized CD4+ levels are between 10%-80%. As the donor chimerism increases, the post therapy viral load drops and CD4+ T cell count rises, however, we observe a wide spectrum. For augmented stem cells > 80%, viral load is below the level of detection and CD4+ T cell counts are back to normal levels for almost all patients. Using the post-therapy viral load and CD4 levels and based of the two conditions that we defined, we calculated the probability of cure for rapid progressors (Figure 5-15I). Surprisingly, the model predicts that if the donor chimerism is less than 20% (current clinical ranges), the probability of cure is zero, suggesting that high levels of chimerism and transfection efficiencies are needed to achieve meaningful ranges for the probability of cure in patients. Comparing rapid progressors with the other populations, we observe that the post-therapy viral load decreases faster and CD4+ T cell counts rise to normal levels more quickly (Figure 5-15), and hence the probability of cure is higher in these populations for a given percentage of CCR5-modified stem cells (Figure 5-15, I-L). For example, if the percentage of stem cells transfected is 60%, the probability of cure will be 51.5% and 83.2% in rapid progressors and slow progressors, respectively. The model also crucially suggested that most of the difference is between rapid progressors and the rest of the HIV-infected individuals (Figure 5-15, I-L). To achieve 90% probability of cure in all patient populations, our model predicted that the level of donor chimerism must be at least 75%. In our model, we assume a partial inhibition of HIV infection for CCR5-modified CD4+ T cells and macrophages, however these cells can still get infected. The level of inhibition is at 92.5%, consistent with the
reported experimental data [221]. To study the impact of this inhibition of infection (level of anti-HIV activity needed) on the probability of cure, we ran the virtual clinical trials for two other cases: less inhibition at 85% and complete inhibition at 100% (Figure 5-16). Using the curves in Figure 5-16, one could design clinical trials and calculate the minimum level of donor chimerism and anti-HIV activity to achieve a certain probability of cure in HIV-infected patient populations. For example, the model suggests that with 60% donor chimerism in slow progressors, 91% probability of cure can be achieved if CCR5-modified cells are completely resistant to infection.

Figure 5-15. Calculating probability of cure for CCR5-deficient stem cell therapy in all populations.
(A-D) Post-therapy viral loads and (E-H) normalized CD4+ levels are shown for any given percentage of stem cells transfected. Using the one-year post-therapy simulation results and the two conditions for a functional cure, we calculated (I-L) the probability of cure in all patient populations.
Figure 5-16. Impact of variations in $\alpha_{CCR5}$, $\beta_{CCR5}$, and $\epsilon_{CCR5}$ on probability of cure for CCR5-deficient stem cell therapy in all populations.

Parameters $\alpha_{CCR5}$, $\beta_{CCR5}$, and $\epsilon_{CCR5}$ are defined as the reduction in infection rate of CCR5$^{-/-}$ CD4$^+$ T cells by HIV, infection rate of CCR5$^{-/-}$ macrophages by HIV, and infection rate of CCR5$^{-/-}$ CD4$^+$ T cells by infected macrophages, respectively. We calculated the probability of cure for complete inhibition ($\alpha = \beta = \epsilon = 0$; dark colors) or partial inhibitions ($\alpha = \beta = \epsilon = 7.5\%$ or 15%; medium and light colors) in all patient populations.

5.3.5 Validation of the virtual populations against clinical data from CCR5-based autologous T cell therapy in HIV-infected patients

Recently, Tebas et al. investigated the infusion of CCR5-modified autologous CD4$^+$ T cells to 12 patients, who were receiving cART and had chronic aviremic HIV infection [99]. Six of these patients, who had baseline CD4$^+$ T cell counts > 450 cells/µL (546-1123 cells/µL) and a documented nadir of not lower than 300 cells/µL, underwent a 12-week cART interruption 4 weeks after the single dose infusion of $10^{10}$ cells (~2000 cells/µL), with 20% of those cells being CCR5-modified on average. Plasma viremia, CCR5-modified and total CD4$^+$ T cell counts were monitored in these patients for 36 weeks (symbols represent the median results in Figure 5-17). For two of these patients, the treatment interruption was terminated prematurely at week 8 of the interruption period. To validate our model against this data, we selected all the virtual patients who had baseline CD4$^+$ T cells levels < 1200 cells/µL from all subpopulations. In our model, these patients began cART 208 dpi, and had undetectable viral load by 400 dpi, at which they received a single dose infusion of 2000 cells/µL (WT: 1600 cells/µL; CCR5-modified: 400 cells/µL). Four weeks after the infusion (428 dpi), the virtual patients
ceased taking cART for 12 weeks (512 dpi), at which they resumed their antiretroviral treatment. Figure 5-17 shows the individual virtual patients (gray curves) and the median of the simulation results (purple curves), which predicts the clinical data (symbols) with low RMSPE (root-mean square prediction error) values. The details of how each subpopulation reacted to the infusion of CCR5-modified T cells are shown in Figure 5-18.

Figure 5-17. Validation of combined virtual populations against clinical data from CCR5-modified autologous T cell therapy in HIV-infected patients.
We ran a virtual clinical trial and monitored (A) viral load, (B) total CD4+ T cell count and (C) CCR5-modified CD4+ T cell count in virtual patients to validate our model against the clinical data from an infusion of CCR5-modified CD4+ T cells to HIV infected individuals. To match the patients in the clinical trial, we selected all the virtual patients who had baseline CD4+ T cells levels < 1200 cells/μL from all subpopulations for this study. The virtual clinical trial includes 1) infection at day 0, 2) taking cART at 208 dpi, 3) a single dose infusion of 2000 cells/μL (WT: 1600 cells/μL; CCR5-modified: 400 cells/μL) at 400 dpi, 4) taking off cART at 428 dpi and 5) restarting cART at 512 dpi; gray curves: individual patients; purple curves: median of virtual patients; symbols: median of clinical data; RMSPE: root mean square prediction error.

5.4 Conclusions and discussion

In this chapter, we introduced a novel mechanistic model of HIV/AIDS that includes key players of HIV infection and reproduces the complete course of the disease from the acute infection to AIDS. Unlike most models in the literature, our model is not running at a steady state, which challenges the HIV RNA setpoint theory, assuming a stable viral load during chronic infection [222]. Using the model, we can successfully create virtual patient populations to capture the huge variations in clinical measurements in cART-
 naïve HIV-infected patients, who progressed to AIDS 1.5-12 years post-seroconversion in the MACS cohort (Table 5-2). Although CD4+ T cells are thought to be the driving force in HIV infection and the role of macrophages is still hotly debated in the field [78,223-226], the model crucially suggested that macrophages are required to reproduce the clinical data in cART-naïve patients (Figure 5-3), as without macrophages, the calibration failed and the model could not fit to the data. In addition, the model suggested that in the late stage of the disease when the viral load increases and CD4+ T cells are almost depleted, infected macrophages play a major role in producing viruses. The immune response to HIV infection is represented by CD8+ CTLs in our model, as there is a strong negative association between the strength of the immune response and progression to AIDS [227]. We also included latently infected CD4+ T cells, as they present the major obstacle to achieve HIV cure [228]. The model has been validated against clinical data for the biphasic viral decay in cART-treated patients (Figure 5-8).

Figure 5-18. Validation of each virtual population against clinical data from CCR5-modified autologous T cell therapy in HIV-infected patients.
To test the efficacy of CCR5-modified stem cell therapies in blocking HIV, we considered two separate populations of WT and CCR5-modified cells for CD4+ T cells and macrophages. The model has the capability to closely simulate the clinical events for a HIV clinical trial such as taking patients on and off cART, bone marrow and total body irradiation, engraftment and host recovery, gene modifications, and donor chimerism. Using virtual populations, our model successfully predicted the results for a clinical study of infusion of autologous CCR5-modified CD4+ T cells to HIV infected individuals (Figure 5-17 and Figure 5-18). However, both the simulation and clinical results indicated that augmented CD4+ T cell therapies do not lead to cure, as there is no source for constant production of these cells in patients. Instead, we used the virtual patients to predict whether the CCR5-deficient stem cell therapy performed on the Berlin patient can be replicated in other HIV-infected individuals (Figure 5-15 and Figure 5-16). The simulations suggested some key results: 1) For donor chimerism < 10% (current clinical ranges), the probability of cure is zero; 2) high levels of donor chimerism is required to achieve meaningful success rates in patients (at least 75% of stem cells must be CCR5 modified to achieve cure in 90% of patients); 3) most of the difference in the success rate of the stem cell therapy lies between rapid progressors (AIDS ≤ 3.5 yrs) and the rest of HIV-infected individuals (AIDS > 3.5 yrs). A major barrier in stem cell therapies has been low efficiency of gene transfections into stem cells, leading to low levels of augmented cells \textit{in vivo} [229-231]. To make anti-HIV stem cell therapies more successful in clinical trials, the results indicate that more efforts should be undertaken to increase \textit{ex vivo} transfection efficiencies and \textit{in vivo} selection of augmented cells, and to deal with graft-vs-host effects for achieving high levels of donor chimerism.
One major issue with blocking or knocking out CCR5 alone is that it could lead to selection of CXCR4 tropic virus [8,232,233]. However, a recent study demonstrated that although the Berlin patient had a minor population of CXCR4 tropic virus prior to transplantation, this population was not able to reestablish HIV infection due to its dependence on CCR5 for replication and high genetic barrier toward CXCR4 usage [234]. Therefore, inclusion of CXCR4 tropic virus would be a useful extension to the model. Regardless of the role of CXCR4 tropic virus, multiple layers of protection against HIV are likely required for complete inhibition of HIV and safe stem cell therapies. Therefore, in addition to knocking out CCR5, other restriction factors such as members of the APOBEC family [22] and SAMHD1 [84] or gene circuits that induce apoptosis in HIV-infected cells [3,235,236] could be used [203] (Figure 1-1). Our model can be extended to design and test the performance of new CD4+ T cell or stem cell therapies with multiple genes involved.
Chapter 6.  Computational Modeling of HIV Infection and Gene-Augmented Stem Cell-Based Cures
6.1 Summary

In the previous chapter, we introduced a mechanistic model of HIV/AIDS that includes major players of infection and reproduces the complete course of the diseases. We created virtual populations to capture huge variability observed in clinical measurements in HIV-infected individuals. Using virtual populations, we tested the performance of CCR5-modified T cell and stem cell therapies. One major issue with blocking CCR5 alone is that it could lead to selection of CXCR4 tropic virus. Therefore, multiple layers of protection against HIV are likely required for complete inhibition of HIV and safe stem cell therapies. In this chapter, we extend the model to include other restriction factors such as APOBEC3G and SAMHD1 or gene circuits that induce apoptosis in HIV-infected cells. We test the efficacy of stem cell therapies involving these genes individually and in combination. The model suggested that APOBEC3G-, CCR5- and apoptosis-based therapies have similar performances, although they target different pathways. Also, we observed additive effects for dual therapies. The results indicated that SAMHD1 alone does not lead to cure in HIV-infected individuals as it only protects macrophages and does not provide protection to CD4+ T cells.

6.2 Description of the extended model

In the model introduced in the previous chapter, we included the effect of CCR5 knockout in CD4+ T cells and macrophages. In the extended model, we include the effects of A3G overexpression and apoptosis-inducing circuits in CD4+ T cells and the impact of SAMHD1 overexpression in macrophages. The following equations describe the model:
\[
\frac{dT^W_U}{dt} = (1 - f_T)S_T + \frac{p_T}{c_T + V}VT^W_U + r_T T^W_U \left(1 - \frac{T}{\tau_{\text{max}}} \right) - k_{VT} VT^W_U - k_{MT} M^I T^W_U - d_T T^W_U, \\
\frac{dT^A_U}{dt} = f_T S_T + \frac{p_T}{c_T + V}VT^A_U + r_T T^A_U \left(1 - \frac{T}{\tau_{\text{max}}} \right) - \alpha_{cCR5} k_{VT} VT^A_U - \epsilon_{cCR5} k_{MT} M^I T^A_U - d_T T^A_U, \\
\frac{dT^W_I}{dt} = (1 - k_{T})(k_{VT} T^W_U + k_{MT} M^I T^W_U) + a_T T^W_U - k_{ET} ET^W_I - d_T T^W_I, \\
\frac{dT^W_I}{dt} = (1 - k_{T})(k_{VT} T^W_U + k_{MT} M^I T^W_U) + a_T T^W_U - k_{ET} ET^W_I - d_T T^W_I, \\
\frac{dT^W_I}{dt} = (1 - k_{T})(k_{VT} T^W_U + k_{MT} M^I T^W_U) + a_T T^W_U - k_{ET} ET^W_I - d_T T^W_I, \\
\frac{dT^A_I}{dt} = (1 - k_{T})\alpha_{APOP}(\alpha_{cCR5} k_{VT} T^A_U + \epsilon_{cCR5} k_{MT} M^I T^A_U) + a_T T^A_U - k_{ET} ET^A_I - d_T T^A_I, \\
\frac{dT^W_L}{dt} = k_T(k_{VT} T^W_U + k_{MT} M^I T^W_U) - a_T T^W_U - d_T T^W_I, \\
\frac{dT^W_L}{dt} = k_T(k_{VT} T^W_U + k_{MT} M^I T^W_U) - a_T T^W_U - d_T T^W_I, \\
\frac{dT^W_L}{dt} = k_T(k_{VT} T^W_U + k_{MT} M^I T^W_U) - a_T T^W_U - d_T T^W_I, \\
\frac{dT^A_L}{dt} = k_T\alpha_{APOP}(\alpha_{cCR5} k_{VT} T^A_U + \epsilon_{cCR5} k_{MT} M^I T^A_U) - a_T T^A_U - d_T T^A_I, \\
\frac{dT^W_M}{dt} = (1 - f_M)S_M + \frac{p_M}{c_M + V}VM^W_M - k_{VM} VM^W_M - d_T M^W_M, \\
\frac{dT^A_M}{dt} = f_M S_M + \frac{p_M}{c_M + V}VM^A_M - \beta_{cCR5} k_{VM} VM^A_M - d_T M^A_M, \\
\frac{dM^W_I}{dt} = k_{VM} V^I_M - k_{EM} EM^W_I - d_T M^W_I, \\
\frac{dM^W_I}{dt} = k_{VM} V^I_M - k_{EM} EM^W_I - d_T M^W_I, \\
\frac{dM^A_I}{dt} = \beta_{cCR5} k_{VM} V^A_I - k_{EM} EM^A_I - d_T M^A_I. \\
\]
\[
\frac{dM_{i(+)}}{dt} = \beta_{CRS} k_{YM} V_{i(+) - k_{EM}} E M_{i(+) - d_M} M_{i(+)}, \tag{16}
\]

\[
\frac{dE}{dt} = s_E + p_E \frac{T}{c_E + T_L} E - d_E E, \tag{17}
\]

\[
\frac{dV_{(-)}}{dt} = n_T (T_{i(-)} + cT_{i(-)} + pT_{i(-)} + pcT_{i(-)}) + n_M (M_{i(-)} + cM_{i(-)} + mM_{i(-)} + cmM_{i(-)}) - d_V V_{(-)}, \tag{18}
\]

\[
\frac{dV_{(+)}}{dt} = n_T ((1 - p)T_{i(+) + (1 - p)cT_{i(+) + d_V V_{(+)}}}, \tag{19}
\]

When A3G-overexpressing CD4+ T cells are infected by viruses, A3G overexpressed in these cells overcomes Vif and gets encapsulated in some of the budded viruses, meaning that infected cells produce two types of virus: those that carry A3G, hence called A3G(+) viruses, and those viruses that do not, dubbed A3G(-) viruses; the ratio of A3G(-) to total released viruses is denoted by \( p \) in Equations (18) and (19). Since macrophages do not overexpress APOBEC3G and we assume all viruses released by infected macrophages are A3G(-) viruses. Also, we assume that infected macrophages act the same as A3G(-) viruses when they infect CD4+ T cells. Macrophages and CD4+ T cells infected by A3G(+) viruses produce fewer virions compared to cells infected by A3G(-) viruses, because the A3G carried in A3G(+) viruses affects viral production by inhibiting several steps of the HIV life cycle inside the infected cell. This reduction in production and release of virus is denoted by \( c \) in Equations (18) and (19). Note that in the extended model, there are 4 populations of productively infected CD4+ T cells because we need to track WT and augmented cells that are infected by A3G(-) and A3G(+) viruses. Similarly, latently infected T cells and infected macrophages are each represented by four populations.
CD4+ T cells augmented with apoptosis inducing circuits have a normal lifespan until virus entry is detected, and only then is the circuit activated. The activation of the apoptosis pathway upon infection causes the infected cell to die, significantly reducing viral production. We assume that some inefficiencies may be involved with this process, suggesting that some augmented cells can survive after infection. This is captured by the inefficacy rate $\alpha_{APOP}$, representing the fraction of infected augmented cells that have escaped auto-apoptosis in Equations (5), (6), (9), and (10).

Overexpression of SAMHD1, induced by IL-12 and IL-18, has shown to boost resistance of MDMs to HIV-1 infection [85]. In the extended model, the number of viruses produced by SAMHD1-overexpressing macrophages is smaller than that of WT macrophages. This reduction in burst size is denoted by $m$ in Equation (18).

6.3 Results

6.3.1 Virtual clinical trials predict the efficacy of augmented stem cell therapies involving a single gene

To predict the efficacy of augmented stem cell therapies, we used the virtual populations developed in the previous chapter along with the extended model. We simulated the complex clinical procedure for stem cell therapies as closely as possible and ran virtual clinical trials for APOBEC3G-, SAMHD1-, apoptosis- and CCR5-based stem cell therapies at different potencies. The virtual clinical trial includes 1) infection at day 0, 2) taking cART at 208 dpi, 3) bone marrow transplant at 500 dpi, which consists of irradiation for 7 days followed by six months of engraftment, and 4) ceasing cART at 687 dpi (the details are discussed in Chapter 5; see Figure 5-14). As mentioned before, during
the irradiation, stem cells are killed and hence no more immune cells are produced. However, after the irradiation is complete, some residual WT stem cells are still left behind. This leads to donor chimerism after infusion of augmented stem cells, and hence we assume that a certain percentage of stem cells are augmented and the rest are WT. We monitored viral load (Figure 6-1) and CD4+ T cell counts (Figure 6-2) in all patients and defined a patient as functionally cured if one year post therapy, normalized CD4+ level > 95% and viral load < 50 copies/mL and decaying (Figure 6-3). For A3G-based therapies, the model suggests that the viral load is lower, CD4+ levels are higher and hence the probability of cure increases, as the potency of A3G increases, i.e. $p$ and $c$ take smaller values. The difference is more pronounced for levels of donor chimerism above 20% for rapid progressors and between 20%-80% for the rest of HIV-infected patient populations. Note that $p$ denotes the percentage of A3G-free viruses released from augmented infected cells and $c$ represents the reduction in the burst size of cells infected by A3G-carrying viruses. Also, the decay in viral load and the return of CD4+ T cells to normal levels are faster, as we transition from rapid progressors to slow progressors. Therefore the probability of cure is higher in slow progressors compared with rapid progressors.

For SAMHD1-based therapies, the model suggests that although CD4+ T cells slightly increase for high levels of donor chimerism, the impact on post-therapy viral load is minimal. And the therapy does not lead to cure even for high potency SAMHD1 (small values for $m$) and at complete chimerism, i.e. when 100% of the stem cells are augmented after engraftment. Note that $m$ represents the reduction in infectivity rate of the augmented macrophages by virus. This is because SAMHD1 only protects macrophages but does not provide any protection for CD4+ T cells; and that protection is not nearly
enough to block HIV infection. For apoptosis-based therapies, we also observed that the probability of cure increases as the apoptosis-inducing circuit becomes more potent, i.e. the failure rate of the circuit, $\alpha_{\text{APOPOP}}$, decreases. We included the results for CCR5-based stem cell therapies (Figure 5-16) for the sake of comparison with the performance of other therapies. The results indicated that APOBEC3G-, CCR5- and apoptosis-based therapies have similar performances at medium to high potencies, although they target different pathways. Also, for all the therapies except SAMHD1-based ones, the probability of cure is lower in rapid progressors compared with the rest of the HIV-infected populations. Regardless of the gene modification, the model predicts that if the donor chimerism is less than 20% (current clinical ranges), the probability of cure is close to zero, suggesting that high levels of chimerism and transfection efficiencies are needed to achieve meaningful ranges for the probability of cure in patients.

**Figure 6-1. Viral load one year post-therapy.**
We ran virtual clinical trials and monitored one-year post-therapy viral load for A3G-, SAMHD1-, apoptosis-, and CCR5-based stem cell therapies; light colored curves: low potency; medium colored curves: medium potency; dark colored curves: high potency.
Figure 6-2. Normalized CD4+ T cell counts one year post-therapy.
We ran virtual clinical trials and monitored one-year post-therapy CD4+ T cell counts for A3G-, SAMHD1-, apoptosis-, and CCR5-based stem cell therapies; light colored curves: low potency; medium colored curves: medium potency; dark colored curves: high potency.

Figure 6-3. Probability of cure for augmented stem cell therapies with a single gene.
Using the one-year post-therapy viral load and CD4+ T cell counts and the two conditions for a functional cure, we calculated the probability of cure for A3G-, SAMHD1-, Apoptosis-, and CCR5-based stem cell therapies; light colored curves: low potency; medium colored curves: medium potency; dark colored curves: high potency.
6.3.2 Virtual clinical trials predict the efficacy of dual stem cell therapies

The rationale of cART is to use multiple anti-retroviral drugs with different mechanisms of action to completely inhibit viral replication and reduce the chance of viral escape. Similarly, several anti-HIV genes should likely be combined in stem cell therapies to provide independent layers of protection against HIV, and hopefully to increase the performance of these therapies. Our model has the capability to simulate the combination of multiple genes and predict the performance of the combined therapies. We selected the top performing single gene strategies, i.e. A3G overexpression, CCR5 knockouts, and apoptosis-inducing circuits, and ran virtual clinical trials to test the performance of stem cell therapies using pairs of these gene strategies. Figure 6-4 shows the performances of individual stem cell therapies with a single gene and that of combined therapies using our model. Assuming $f_A$ and $f_B$ represent the performance of therapy A and B, we used theoretical models [237] to calculate the performance of combined therapies in two cases: 1) individual therapies have similar mechanisms of action (the Loewe additivity model [238] in Equation 20) and 2) individual therapies act independently (the Bliss independence model [239] in Equation 21). If the performance of combined therapies predicted by the model exceeds that predicted by the Bliss model, we observe synergistic effects; otherwise, it is antagonistic. Also, if the simulated performance matches that of the Loewe additivity model, we observe additive effects.

\[
f_{A*B} = \left(1 - (1 - f_A)(1 - f_B)\right)
\]

\[
\frac{f_{A+B}}{1-f_{A+B}} = \frac{f_A}{1-f_A} + \frac{f_B}{1-f_B}
\]
The model suggests that the combined therapies outperform the individual therapies in all cases and exhibit antagonistic but additive effects for donor chimerism > 50% in most cases. The effects are more pronounced in rapid progressors for the A3G-containing therapies, for which we observe performances achieving that predicted by the Bliss independence model. Finally, the model suggests that none of the combinations offers significant improvement over the other combination therapies.

![Figure 6-4. Probability of cure for augmented stem cell therapies with two genes.](image)

Using the one-year post-therapy viral load and CD4+ T cell counts and the two conditions for a functional cure, we calculated the probability of cure for dual stem cell therapies; light and medium colored curves: performance of stem cell therapies with a single gene; dark colored curves: performance of therapies with two genes; light gray dashed curves: predicted performance of dual therapies using the Loewe additivity model; dark gray dashed curves: predicted performance of dual therapies using the Bliss independence model.

### 6.4 Discussion

Here, we extended the mechanistic model of HIV infection introduced in the previous chapter to include new anti-HIV gene-based therapies, including the effects of A3G overexpression and apoptosis-inducing circuits in CD4+ T cells and the impact of SAMHD1 overexpression in macrophages. We used our virtual populations to predict the performance of augmented stem cell therapies individually and in combination. The
simulations suggested some key results: 1) A3G-, CCR5- and apoptosis-based therapies have similar performances, although they target different pathways; 2) SAMHD1 does not lead to cure even at complete chimerism with high levels of potency; 3) the probability of cure is lower in rapid progressors (AIDS ≤ 3.5 yrs) compared with the rest of HIV-infected individuals (AIDS > 3.5 yrs); 4) When combined, we observed additive effects between therapies involving A3G overexpression, CCR5 knockouts and apoptosis-inducing circuits. The results further indicated that although combination therapies help lessen the burden on each individual gene to work at high efficiency and reduce the chance of viral escape, we cannot rely on combining genes to compensate for low levels of donor chimerism. Also, the individual therapies that we used in our simulations performed at medium to high levels of potency and hence might have already reached a level that is very close to the maximum achievable performance. Therefore, we hypothesize that if we simulate the individual therapies at lower levels of potency, we may get more synergistic effects in combination therapies.

In addition to A3G and SAMHD1, other anti-HIV genes and restriction factors such as TRIM5α, BST-2, Ribozymes, and mc46 have been tested in vitro and in vivo (reviewed in [240,241,242]). Our model can be extended to include those genes and design therapies with multiple genes. Our current model includes HIV-1 latency in CD4+ T cells, because HIV appears to mainly establish latent infection in resting CD4+ T cells; however, there is evidence that latency can also be established in the cells of monocyte/macrophage lineage [243] and hence these cells could also be added to the model. For simplicity, the model only consists of the blood compartment with different types of immune cells including CD4+ and CD8+ T cells and the macrophage population.
representing monocytes, tissue macrophages, and monocyte-derived macrophages that are susceptible to HIV infection. Ideally, the model should include other key compartments such as lymph nodes, spleen and CNS [244] for a more complete representation of the disease in vivo. One key question in the field of HIV cure is to predict the median time to viral rebound for different cure strategies [218]. To do so for stem cell therapies, we can use a hybrid version of our model, in which a stochastic model of latency replaces the deterministic ODE equation for latently infected CD4+ T cells.
Chapter 7. Future Directions
In this work, we introduced a multiscale mechanistic model of the *in vitro* HIV lifecycle inside and outside of cells, using differential equations. In the model, we included molecular interactions between the host restriction factor A3G and viral protein Vif to study the impact of A3G-based treatments. Experimental data were used to establish system parameters and to validate our simulation results against CD4+ T cell culture experiments. Finally, we used the model to provide insights into HIV infection and generate testable hypotheses by predicting the effects of several A3G-based treatments in blocking *in vitro* HIV infection. For example, the model predicted that A3G with a mutated Vif binding site [185,186] is significantly more effective than other molecules. To verify predictions of the computational model, one possible future direction is to perform a series of experiments. These experiments will measure system parameters such as intracellular A3G and Vif concentrations (by ELISA) and the incorporation rate of A3G into HIV particles (by northwestern blot). Also, using standard gene delivery protocols, one can transfecct cell lines to constitutively express mutated forms of A3G. These cells will be infected with HIV using standard protocols. Then, HIV levels (p24 assay) and cellular infection (e.g. luciferase assay) will be compared to the model predictions and used to refine the model.

In addition to A3G, other endogenous restriction factors have been discovered that confer protection against retroviral infections. In this thesis, we extensively discussed the APOBEC3 family of proteins, which are incorporated into HIV-1 virions and deaminate cytosines to uracils, causing G-to-A mutations. In our *in vivo* model, we also implicitly included SAMHD1, which inhibits HIV and SIV infection in macrophages and myeloid cells by depleting the intracellular pool of dNTPs required for viral reverse transcription.
Other restriction factors include the tripartite-motif-containing (TRIM) family members (which target the viral capsid and inhibit viral transcription) [247-251], and bone marrow stromal cell antigen 2 (BST2) or tetherin (which inhibits the release of budding virions from infected cells) [252-256]. As these anti-HIV genes have therapeutic effects and could potentially be used in augmented stem cell therapies, their mechanisms of action should be included in the in vitro single-cell model. The results then will be integrated into the in vivo model to predict the efficacy of augmented stem cell therapies using these genes individually or in combination with other anti-HIV genes.

In our mechanistic in vivo model, we included multiple immune cell types such as CD4+ T cells, macrophages, latently infected T-cells, and CTLs. In the model, we track WT CD4+ cells and augmented cells, in which A3G is overexpressed, CCR5 is rendered dysfunctional, or an apoptosis-inducing circuit is included. We also track WT macrophages or augmented macrophages, in which SAMHD1 is overexpressed or CCR5 is rendered dysfunctional. In the model, macrophages represent a mixed population of monocytes, tissue macrophages, and monocyte-derived macrophages that are susceptible to HIV infection. Also, we have not separately included resting naïve, activated, and resting memory CD4+ T cells in the model. Therefore, a possible future direction would be to add multiple compartments to the model including the spleen, lymph nodes, CNS, and blood [244], to include monocytes and macrophages separately, and to include CD4+ T cells at different stages of maturation. We have only included latently infected CD4+ T cells as HIV appears to mainly establish latent infection in resting CD4+ T cells. The model should be extended to include latently infected monocytes/macrophages, as there is evidence that latency can also be established in these cells [243]. Although we include
WT and augmented cell populations in the model, there is no direct interaction between these cells to represent graft-vs-host effects, which can be thought of as a step-by-step escalation in immune activation, leading to massive target tissue apoptosis [257]. Fortunately, rational approaches have been designed to better manage patients after bone marrow transplants. As our knowledge of graft-vs-host effects increases, the model should be augmented to include the biology of these effects and the clinical treatments to simulate a more accurate representation of anti-HIV stem cell transplants in patients. Also, as the clinical results of anti-HIV T cell and stem cell therapies become available, the model should be validated against the results and used to provide insight into the results. For our models, we have used the aggregated clinical data from untreated patients to create virtual populations. The virtual population can be potentially built using clinical data from individual patients if the data has been collected regularly and with sufficient resolution. Gene-augmented stem cell therapies have recently been tested in nonhuman primate models of AIDS [15,19,21], and hence the model should be translated to a mathematical model of non-human primates to be able to use pre-clinical data, validate against it and then apply the refined model to predict the results in humans. Finally, collaborations with companies such as Sangamo BioSciences and research centers that perform pre-clinical anti-HIV stem cell therapies are recommended as it could provide essential preclinical and clinical data for the model development and benefit the field of HIV cure with the insights from the mechanistic models.
Bibliography


of Human APOBEC3G and APOBEC3F. Journal of Infectious Diseases 198: 486-492.


89. Cohen OJ, Vaccarezza M, Lam GK, Baird BF, Wildt K, Murphy PM, et al. (1997) Heterozygosity for a defective gene for CC chemokine receptor 5 is not the sole
determinant for the immunologic and virologic phenotype of HIV-infected long-term nonprogressors. Journal of Clinical Investigation 100: 1581.


92. Muntinghe FL, Carrero JJ, Navis G, Stenvinkel P (2011) TNF-α levels are not increased in inflamed patients carrying the CCR5 deletion 32. Cytokine 53: 16-18.


governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem
279: 14481 - 14483.


Decay characteristics of HIV-1-infected compartments during combination


equation models of HIV-1 infection. Mathematical Biosciences 179: 73-94.

Proceedings of the National Academy of Sciences of the United States of America
102: 8198-8203.

193. Sedaghat AR, Siliciano RF, Wilke CO (2008) Low-level HIV-1 replication and the
dynamics of the resting CD4+ T cell reservoir for HIV-1 in the setting of HAART.
BMC Infect Dis 8: 2.

Plasma Virus Concentration during Primary HIV Infection. Journal of Theoretical
Biology 203: 285-301.


infection. JAIDS Journal of Acquired Immune Deficiency Syndromes 41: 259-265.


Curriculum Vitae

Iraj Hosseini was born in Isfahan, Iran, on August 26, 1984. Iraj received his B.Sc. with honors in electrical and computer engineering (ECE) from Isfahan University of Technology, Iran in 2007. During his undergraduate studies, Iraj participated in international RoboCup competitions, where he won four awards in competitions held in Italy, Germany, and Portugal. He was awarded the university gold medal, as the top undergraduate student among 140 students in the ECE department. In 2009, Iraj obtained his M.Sc. in ECE from University of Alberta (UofA), Canada. At UofA, he received several prestigious awards including the iCORE Scholarship for International Students and the Alberta Ingenuity Graduate Student Scholarship. Iraj enrolled at the Johns Hopkins School of Medicine in 2009 to pursue his Ph.D. in biomedical engineering, focusing his research on building molecular-detailed computational models of HIV/AIDS to better understand HIV pathogenesis and design stem cell-based anti-HIV therapies. Beyond his research, Iraj led novel efforts to reinvigorate PhD education and prepare students for alternative career opportunities. Iraj co-founded BME EDGE, obtaining $150,000 in competitive funding from the Provost's Office and matching funds from the BME department. At EDGE, Iraj co-chaired a team of 17 students and staff members to develop an integrated career pathway for graduate students through internships, professional development workshops, alumni speaker series, networking events and PhD-focused job fairs. At Hopkins, Iraj was awarded numerous awards including the William and Mary Drescher Endowment Fund for Graduate Medical Research, the Ruth H. Aranow Fellowship in Computational Medicine and the Siebel Scholarship. Starting in August 2015, Iraj will be relocating to San Francisco, CA, where he will join Genentech.