HLA-mediated control and CD8+ T cell response mechanisms in persistent viral infections

A thesis submitted to
Imperial College London
for the degree of Doctor of Philosophy

by

Nafisa-Katrin Seich al Basatena

Department of Medicine
Imperial College London
St Mary’s Campus
September 2012
Acknowledgements

I would like to most cordially thank my supervisor Becca Asquith for her constant help, guidance and support. She truly created a great environment for advancing during the course of my PhD research. I have certainly learned a lot from her that I will always value. I would also like to express my gratitude to the Wellcome Trust for funding this work. Special thanks should also go to my second supervisor Charles R. B. Bangham for many useful discussions and vital input to my research. I am deeply grateful to all my colleagues in the Department of Immunology and most of all to U. Kadolsky and M. Elemans for encouraging me and sharing their insightful advice on many aspects of my work.

This project would not have been feasible without the data provided by our collaborators Alison M. Vine, Koichiro Usuku, Mitsuhiro Osame, Chloe L. Thio,, Jacqie Astemborski, Gregory D. Kirk, Sharyne M. Donfield, James J. Goedert, Salim Khakoo, Mary Carrington, Nicole Klatt, Guido Silvestri and Emma Thomson; for this I am very grateful. Additionally, I would like to especially extend my appreciation to S. Khakoo, M. Carrington, J. Trowsdale, J. Traherne, F. Graw, R. Regoes, G. O’Connor and D. McVicar for very useful comments and productive collaborations.

Finally and very importantly, I want to express my whole-hearted appreciation for the genuine, very altruistic and all-around support of my partner, K. Chatzimichalis and warmly thank my lovely parents and all my great friends for always being there for me during these demanding three years.
To my family and friends ... 

for always being there for me ...

και πάνω απ' όλα στην αξέχαστη γιαγιά μου ...

... To my family and friends ...

... for always being there for me ...

και πάνω απ' όλα στην αξέχαστη γιαγιά μου ...
Declaration of Originality

All the work presented in this thesis is the author’s own unless clearly stated in the statement of collaboration and the relevant sections of the thesis.

Signature ..................

Date .....................
Statement of Collaboration

The work presented in Chapters 2 and 3 would not have been possible without the experimental work and the datasets collected by our collaborators; HTLV-1 cohort: Alison M. Vine, Koichiro Usuku, Mitsuhiro Osame and Charles R. M. Bangham and HCV cohort: Chloe L. Thio, Jacquie Astemborski, Gregory D. Kirk, Sharyne M. Donfield, James J. Goedert, Mary Carrington and Salim I. Khakoo. The cohorts were HLA and KIR genotyped for other studies and were generously provided for further analyses. Additionally, Dr Aidan Macnamara contributed a part of the epitope prediction results presented in Chapter 3. The findings of this project were published in [1].

For the part of Chapter 3 that refers to the role of KIR3DS1 in HTLV-1 infection, the study was lead by Dr Geraldine O’Connor and Dr Daniel McVicar and I contributed the statistical analysis for the Japanese HTLV-1 cohort. The latter study was published in [2].

The study-design presented in Chapter 4 involves the analysis of experimental data (a longitudinal HCV cohort) gathered by our collaborator Dr Emma Thomson. The KIR-typing was performed by Dr Susanne Knapp and the preparation of the samples for next-generation sequencing was done by Dr Heather Niederer and Aviva Witkover. The sequencing was performed by Niall Gormley and his team at Illumina. The study design was part of a grant proposal submitted by my supervisor Dr Becca Asquith to the MRC. The grant has now received full funding in order to be implemented by our group.

The work presented in Chapter 5 was a collaborative effort with Dr Marjet Elemans. The R code for fitting the models presented in this chapter was written by myself and implemented by both myself and my collaborator. Other parts of
this study were done only by Dr Marjet Elemans and Dr Becca Asquith and are
not presented in this thesis. The experimental data obtained from SIV-infected
macaques were provided by Dr Nicole Klatt and Dr Guido Silvestri. All the
study results (including the part presented here) were published in [3].

Konstantinos Chatzimichalis provided valuable advice on coding in the C++
programming language for the 3D Cellular Automaton computational model pre-
presented in Chapters 6 and 7. Additionally, the rules for the motility of T cells
were provided in text format by Dr Frederik Graw and I implemented them from
scratch in C++. This work is currently submitted for publication.
Dissemination

Publications

1. **Seich al Basatena N.K.** and Asquith B., Is the probability of target recognition by CD8+ T cells low? (In preparation).


Presentations

1. Poster Presentation in Viruses, Genes and Cancer Workshop in Venice (Travel award) - September 2010

2. Poster Presentation in British Society of Immunonology Congress in Liverpool - December 2010

3. Oral Presentation in International Workshop: T lymphocyte dynamics in acute and chronic viral infection in London - January 2011

4. Oral Presentation at Division of Infectious Diseases 'Away-Day' in Imperial College London - January 2011

5. Poster Presentation at Keystone meeting on HIV Evolution, Genomics and Pathogenesis in Vancouver - March 2011

6. Poster Presentation at the European Mathematical Genetics Meeting at Kings College London - April 2011

7. Poster Presentation at the 19th HIV Dynamics & Evolution International Conference at Asheville, North Carolina - April 2012
Outreach activities

1. Project management team for the ‘Beautiful Science’ discussion and exhibition events and leader of the outreach event, Imperial College and Wellcome Trust People’s Award - Jul. 2011 - Jul. 2012

2. Project management team for the British Society of Immunology stand at the Big Bang Science Fair, London Excel - Nov. 2010 - Mar. 2011

3. Co-author of the article ‘Immunology... Do the Maths?’ article featured at the British Society of Immunology Newsletter - Jan. 2012

4. Volunteer at the Royal Society Summer Science Exhibition, Imperial College and Exscitec - Jul. 2011

5. Soapbox Scientist at the Imperial College Festival discussing the topic: ‘Is science too advanced to explain to the public?’ - June 2012

6. Volunteer at the ‘Greying Matters’ event at the Dana Center discussing with the public what happens to our brains as we grow older - June 2011

7. Member of the ‘Reaching Further’ scheme that communicates science to students. Developed together with other colleagues an activity that introduces PCR to A-level students
Abstract

Background  There are many viruses that result in persistent infections affecting millions of people worldwide. Although our immune system deploys different strategies to eliminate them, many times they prove unsuccessful and call for a better understanding of the host-virus interplay. One important weapon of the immune system is CD8+ T cells which identify infected cells and limit the spread of infection using different effector mechanisms.

Aim  The aim of this study is two-fold; the investigation of 1) the impact of immunogenetic factors such as HLA class I molecules and Killer cell immunoglobulin-like receptors (KIRs) on CD8+ T cell responses and 2) the efficiency of lytic and non-lytic CD8+ T cell responses and how they shape viral escape dynamics.

Methods  The methods used to address the aims include statistical models, high-throughput sequence analysis, ordinary differential equation models and agent-based models.

Results  We find that HLA class I molecules explain a small percentage of the heterogeneity observed in the outcome of HCV, HTLV-1 and HIV infections. However, we show that an inhibitory KIR, namely *KIR2DL2*, can enhance both protective and detrimental HLA class I-restricted anti-viral immunity, for both HCV and HTLV-1 infections and in a manner compatible with the modulation of CD8+ T cell downstream responses. Furthermore, for HIV/SIV infection, we show that the CD8+ T cell control of the infection can be consistent with a non-lytic mechanism. Additionally, we find that lytic CD8+ T cell responses are more efficient than non-lytic responses which can lead to slower and less frequent viral escape explained by spatial factors.

Conclusions  We conclude that KIRs can play an important role in shaping HLA class-I mediated immunity and suggest that this occurs in synergy with CD8+ T cells whose lytic and non-lytic effector functions can differ in efficiency and lead to variable viral escape rates.
Contents

Acknowledgements

Declaration of Originality

Statement of Collaboration

Dissemination

Abstract

List of Abbreviations

List of Figures

List of Tables

1 Introduction

1.1 HLA class I molecules ........................................................................... 27

1.1.1 Antigen presentation ....................................................................... 28

1.1.2 Polygenism .................................................................................. 30

1.1.3 Polymorphism ................................................................................ 30

1.1.4 Nomenclature ................................................................................ 32

1.1.5 Associations with viral infection .................................................... 32

1.2 CD8+ T cell function in viral infection .................................................. 33

1.2.1 Background .................................................................................. 33

1.2.2 CD8+ T cell effector mechanisms .................................................. 34

1.2.3 Evading CD8+ T cell responses ....................................................... 36

1.3 Killer Cell Immunoglobulin-like receptors ......................................... 37

1.3.1 Background .................................................................................. 37
1.3.2 Associations with disease ............................................. 38
1.4 Persistent viral infections ................................................. 39
  1.4.1 Hepatitis C ............................................................... 39
  1.4.2 Human T-Lymphotropic Virus 1 ................................... 42
  1.4.3 Human Immunodeficiency Virus 1 ................................. 45
1.5 Thesis Outline ............................................................... 49

2 HLA class I impact on HCV and other viral infections ........ 52
  2.1 Aim ................................................................. 52
  2.2 Introduction ............................................................ 52
  2.3 Methods ................................................................. 56
    2.3.1 Data ............................................................... 56
    2.3.2 Statistical analysis .................................................. 57
    2.3.3 Epitope prediction ................................................... 61
  2.4 HLA class I and HCV infection ....................................... 66
    2.4.1 Heterozygotes’ advantage ......................................... 66
    2.4.2 Rare allele advantage ............................................... 73
    2.4.3 Individual alleles ................................................... 73
    2.4.4 HLA class I breadth ................................................. 77
    2.4.5 Protein specificity ................................................... 79
    2.4.6 Epitope specificity ................................................... 83
  2.5 Quantification of HLA class I impact on viral infection ....... 85
  2.6 Discussion ............................................................... 89

3 Impact of KIRs on HLA class I-mediated immunity ................ 93
  3.1 Aim ................................................................. 93
  3.2 Introduction ............................................................ 93
  3.3 Methods ................................................................. 95
    3.3.1 Data ............................................................... 95
3.3.2 Statistical analysis ................................................. 96
3.3.3 Epitope prediction .................................................. 98
3.4 KIR2DL2 in HTLV-1 infection ........................................ 99
   3.4.1 Disease status .................................................. 99
   3.4.2 Proviral load ................................................... 100
   3.4.3 HLA class I specificity ......................................... 102
3.5 KIR2DL2 in HCV infection ........................................... 103
   3.5.1 Spontaneous viral clearance .................................... 103
   3.5.2 Viral load in chronic infection .................................. 103
3.6 No KIR2DL2 main effect on outcome ................................. 107
3.7 Linkage disequilibrium .............................................. 107
   3.7.1 Linkage between HLA class I alleles .......................... 111
   3.7.2 Linkage between KIR genes .................................... 113
   3.7.3 Conclusions on LD .............................................. 116
3.8 Canonical KIR-HLA binding .......................................... 116
3.9 The role of KIR2DL2 ligands ........................................ 118
3.10 The role of other KIRs .............................................. 118
   3.10.1 Other Inhibitory KIRs ........................................ 118
   3.10.2 Activatory KIRs ................................................ 120
   3.10.3 KIR haplotypes ................................................ 122
3.11 Potential mechanism ................................................ 124
3.12 Discussion .......................................................... 128

4 Do KIRs affect CD8+ T cell selection pressure? - Study design 131
   4.1 Aim ................................................................. 131
   4.2 Introduction ........................................................ 131
   4.3 Description of study design ....................................... 133
   4.4 Data ................................................................. 136
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABMs</td>
<td>Agent-based Models</td>
</tr>
<tr>
<td>ACs</td>
<td>Asymptomatic Carriers</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>AICc</td>
<td>bias-adjusted Akaike Information Criterion</td>
</tr>
<tr>
<td>AICD</td>
<td>Activation-Induced Cell Death</td>
</tr>
<tr>
<td>ARF</td>
<td>Alternative Reading Frame</td>
</tr>
<tr>
<td>ARVs</td>
<td>Antiretroviral Drugs</td>
</tr>
<tr>
<td>ATL</td>
<td>Adult-T cell Leukemia</td>
</tr>
<tr>
<td>CA</td>
<td>Cellular Automata</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>dpi</td>
<td>days post infection</td>
</tr>
<tr>
<td>ELISPOT</td>
<td>Enzyme-linked immunosorbent spot</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-Wide Association Studies</td>
</tr>
<tr>
<td>HAART</td>
<td>Highly Active Antiretroviral Therapy</td>
</tr>
<tr>
<td>HAM/TSP</td>
<td>HTLV-I associated myelopathy/tropical spastic paraparesis</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C Virus</td>
</tr>
<tr>
<td>HIV-1</td>
<td>Human Immunodeficiency Virus 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>HLA</td>
<td>Human Leukocyte Antigen</td>
</tr>
<tr>
<td>HTLV-1</td>
<td>Human T-Lymphotropic Virus 1</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IS</td>
<td>Immunological Synapse</td>
</tr>
<tr>
<td>KIRs</td>
<td>Killer Cell Immunoglobulin-like Receptors</td>
</tr>
<tr>
<td>LIF</td>
<td>Leukemia Inhibitory Factor</td>
</tr>
<tr>
<td>MΦs</td>
<td>Macrophages</td>
</tr>
<tr>
<td>MDC</td>
<td>Macrophage Derived Chemokine</td>
</tr>
<tr>
<td>MHC</td>
<td>Major Histocompatibility Complex</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MTOC</td>
<td>Microtubule-organising Center</td>
</tr>
<tr>
<td>NKs</td>
<td>Natural Killer Cells</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PBMC</td>
<td>Peripheral Blood Mononuclear Cell</td>
</tr>
<tr>
<td>pMHC</td>
<td>peptide MHC complex</td>
</tr>
<tr>
<td>PVL</td>
<td>Proviral Load</td>
</tr>
<tr>
<td>R₀</td>
<td>Basic Reproductive Ratio</td>
</tr>
<tr>
<td>RBV</td>
<td>Ribavirin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RN</td>
<td>Reticular Network</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>SDF-1</td>
<td>Stromal Cell-Derived Factor-1</td>
</tr>
<tr>
<td>SIV-1</td>
<td>Simian Immunodeficiency Virus 1</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>STLV-1</td>
<td>Simian T-Lymphotropic Virus 1</td>
</tr>
<tr>
<td>SVR</td>
<td>Sustained Virological Response</td>
</tr>
<tr>
<td>TAP</td>
<td>Transporter associated with Antigen Processing</td>
</tr>
<tr>
<td>TCR</td>
<td>T Cell Receptor</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour Necrosis Factor</td>
</tr>
<tr>
<td>VL</td>
<td>Viral Load</td>
</tr>
<tr>
<td>VPA</td>
<td>Valporic Acid</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Antigen presentation by HLA class I molecules to CD8+ cytotoxic T lymphocytes (CTLs). ......................................................... 29

1.2 Polymorphism of HLA class I genes ............................................. 31

1.3 Schematic outline of the thesis ....................................................... 51

2.1 Consistency of predictions between Metaserver and NetMHCPan ............. 65

2.2 Predicted HLA class I breadth in HCV (genotype 1a) infection .................. 71

2.3 Predicted HLA class I breadth in HTLV-1 infection .......................... 72

2.4 HLA class I associations with HCV infection outcome in UK, USA and pooled cohorts. ......................................................... 76

2.5 Binding preference of HCV proteins for protective and detrimental HLA class I molecules ............................ 81

2.6 Quantification of the HLA class I impact on persistent viral infections ......... 88

3.1 The protective effect of binding HBZ in HTLV-1 infection is enhanced by KIR2DL2 ................................................................. 104

3.2 The impact of HLA-B*57 and KIR2DL2 on HCV viral load ...................... 108

3.3 KIR2DL2 enhances HLA class I-mediated antiviral immunity .................. 109

3.4 Potential mechanisms of KIR2DL2 impact on HLA class I-mediated immunity ................................................................. 127

3.5 KIR2DL2 enhances HLA class I-mediated antiviral immunity .................. 109

4.1 Experimental setup for obtaining the NGS HCV data ........................... 134

4.2 The HCV amplicons sequenced in this study ..................................... 135

4.3 Pipeline for post-processing the HCV NGS data in silico .................... 135

4.4 A longitudinal cohort of HCV/HIV co-infected individuals ..................... 137

4.5 Predicted HCV epitopes for viral proteins NS3, NS5A and NS5B ............. 138

4.6 Errors arising during the NGS data generation .................................. 140
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>Base calling quality control (Filter a)</td>
<td>142</td>
</tr>
<tr>
<td>4.8</td>
<td>Base calling quality control (Filter b)</td>
<td>143</td>
</tr>
<tr>
<td>4.9</td>
<td>Optimal base calling quality filter</td>
<td>144</td>
</tr>
<tr>
<td>4.10</td>
<td>Adapter sequence bias</td>
<td>145</td>
</tr>
<tr>
<td>4.11</td>
<td>Sequence variation of 78 HCV genotype 1a strains</td>
<td>147</td>
</tr>
<tr>
<td>4.12</td>
<td>Coverage of HCV amplicons</td>
<td>149</td>
</tr>
<tr>
<td>4.13</td>
<td>Coverage per HCV amplicon</td>
<td>150</td>
</tr>
<tr>
<td>5.1</td>
<td>Best fitting lytic and non-lytic models to experimental data</td>
<td>163</td>
</tr>
<tr>
<td>5.2</td>
<td>Model comparison using AICc</td>
<td>164</td>
</tr>
<tr>
<td>6.1</td>
<td>A snapshot of the 3D cellular automaton model</td>
<td>169</td>
</tr>
<tr>
<td>6.2</td>
<td>Cell populations considered in the 3D cellular automaton model</td>
<td>169</td>
</tr>
<tr>
<td>6.3</td>
<td>The speed of CD8+ T cells in the 3D cellular automaton model</td>
<td>172</td>
</tr>
<tr>
<td>6.4</td>
<td>The mean displacement of CD8+ T cells in the 3D cellular automaton model</td>
<td>173</td>
</tr>
<tr>
<td>6.5</td>
<td>The ‘probability of successful scan’ and CD8+ T cell speed</td>
<td>174</td>
</tr>
<tr>
<td>6.6</td>
<td>The CD4+ T cell influx model</td>
<td>176</td>
</tr>
<tr>
<td>6.7</td>
<td>The lifespan of wild-type infected cells</td>
<td>179</td>
</tr>
<tr>
<td>6.8</td>
<td>Possible CD8+ T cell responses after the <em>scanning time</em> of the infected target is completed</td>
<td>182</td>
</tr>
<tr>
<td>6.9</td>
<td>Non-lytic models simulated with the 3D cellular automaton model</td>
<td>184</td>
</tr>
<tr>
<td>6.10</td>
<td>Simulated secretion patterns of soluble factors</td>
<td>185</td>
</tr>
<tr>
<td>6.11</td>
<td>Schematic of a model which includes saturated epitope-specific CD8+ T cell killing</td>
<td>189</td>
</tr>
<tr>
<td>7.1</td>
<td>Dynamics of wild-type infected cells over a course of 150 days</td>
<td>194</td>
</tr>
<tr>
<td>7.2</td>
<td>Probability of successful scan by specific CD8+ T cells and killing rate</td>
<td>197</td>
</tr>
<tr>
<td>7.3</td>
<td>Duration of CD8+ T cell lysis and killing rate</td>
<td>198</td>
</tr>
</tbody>
</table>
7.4 Dynamics of wild-type and variant infected cells over a course of 150 days
7.5 Estimation of the variant escape rate
7.6 Higher CD8+ T cell killings leads to faster viral escape
7.7 CD8+ T cell killing rate is higher during ‘escape phase’
7.8 The absolute number of wild-type infected cells killed per day is constant
7.9 Estimation of the CD8+ T cell killing rate using a saturated killing term
7.10 The AICc of the fitted models over different simulated datasets
7.11 Correlation of simulated and estimated killing rate using a mass-action or a saturated killing term.
7.12 Non-lytic suppression leads to slower escape dynamics
7.13 Number of uninfected cells protected’ from infection under a non-lytic CD8+ T cell response
7.14 New infections prevented under a non-lytic CD8+ T cell response that blocks infection
7.15 Set-point of productively infected cells
7.16 Formation of clusters of infected cells during infection spread
7.17 New infections prevented under a non-lytic CD8+ T cell response that blocks infection
7.18 Number of uninfected cells protected’ from infection under a non-lytic CD8+ T cell response
7.19 New infections prevented under a non-lytic CD8+ T cell response that blocks infection
A.1 The HCV genome and gene products
A.2 Dependence of Explained Fraction on factor frequency
B.1 A*02 is a strong epitope binder - Experimental data
B.2 A*02 is a strong epitope binder - Predicted data
C.1 Sequence alignment using ambiguous reference strain
List of Tables

1.1 Groups of HLA class I molecules which are known ligands for KIRs 38

2.1 Evaluation of epitope predictors 63
2.2 Impact of zygosity on HCV infection outcome 69
2.3 Impact of supertype zygosity on HCV infection outcome 70
2.4 Allelic distribution per HLA class I locus 70
2.5 Impact of HLA-A and B supertype frequency on HCV infection outcome 74
2.6 The HLA alleles that are consistently associated with the outcome of the HCV infection 75
2.7 Overall breadth of HLA-A, B, C molecules in HCV infection 78
2.8 Breadth per protein of HLA-A, B, C molecules in HCV infection 80
2.9 No predicted epitopes were consistently associated with the outcome of HCV infection 84
2.10 Factors explaining part of the heterogeneity in HCV infection outcome 86
2.11 Factors explaining part of the heterogeneity in HTLV-1 infection outcome 87
2.12 HLA class I alleles that explain part of the heterogeneity in HIV-1 infection outcome 87

3.1 KIR2DL2 enhances the protective effect of C*08 and the detrimental effect of B*54 on HAM/TSP risk and, independently, on proviral load 101
3.2 KIR2DL2 enhances the protective effect of HLA-B*57 on HCV status and, independently, on HCV viral load 105
3.3 KIR2DL2 enhancement of the B*57 protective effect in HCV infection stratified by race 106
3.4 No KIR2DL2 main effect on status or viral burden 110
3.5 Linkage between the KIRs in HTLV-1 and HCV cohorts 114
3.6 No role for KIR3DS1 in HTLV-1 disease outcome . . . . . . . . . . . . . 121
3.7 No role for KIR3DS1 in HTLV-1 proviral load . . . . . . . . . . . . . 121
3.8 The role of KIR haplotypes . . . . . . . . . . . . . . . . . . . . . . . . . 123
3.9 KIR-B haplotype enhancement is not seen in the absence of KIR2DL2 . 123
5.1 List of ODE models fitted to the experimental data . . . . . . . . . . . . 161
5.2 Comparison of ODE models of lytic and non-lytic control . . . . . . . 161

A.1 List of HLA class I associations with HCV outcome reported in the
literature. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 243
A.2 Confounding factors in the HCV cohort . . . . . . . . . . . . . . . . . . . 244
B.1 False Discovery Rate of cohort stratifications . . . . . . . . . . . . . . . . 246
B.2 Impact of HLA-C1 alleles on HTLV-1 disease status and proviral load . 248
B.3 Only the effect of B*57 and not of alleles with similar binding is en-
hanced by KIR2DL2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 250
B.4 KIRs which are known to bind B*57 do not enhance the B*57 protective
effect as much as KIR2DL2 does . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 250
E.1 CA model initialisation values and motility parameters . . . . . . . . . . 261
E.2 CA model parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 262
F.1 Percentage of simulations not resulting in viral escape . . . . . . . . . . . 266
Chapter 1

Introduction

The present thesis studies the CD8+ T cell response to persistent viral infections. We investigate the immunogenetics of persistent viral infections, focusing on the effect of Human Leukocyte Antigen (HLA) molecules and Killer-cell Immunoglobulin-like Receptors (KIRs) on the disease outcome. We also study the CD8+ T cell effector mechanisms, focusing on the dynamics of lytic and non-lytic CD8+ T cell immune pressure and how these relate to ‘viral escape’ dynamics. The persistent viral infections discussed in the thesis are HTLV-1, HCV and HIV. The methods used include statistical models, high-throughput sequence analysis, ordinary differential equation models and agent-based models.

1.1 HLA class I molecules

Human Leukocyte Antigen complex (HLA) is the name of the major histocompatibility complex (MHC) in humans\(^1\). It is a set of genes located at chromosome 6 and constitutes the most gene dense region of the human genome. They are expressed on all the cells of the body that have a nucleus.

HLA class I molecules are important in shaping the response of two different types of immune cells: CD8+ T cells and Natural Killer (NK) cells. Specifically, they are responsible for presenting antigen (e.g. viral peptides) mainly from the inside of the cell to CD8+ T cells as opposed to MHC class II molecules which present extracellular antigens via Antigen Presenting Cells (APCs) to helper CD4+ T cells. Additionally,

\(^1\)In this thesis, since we predominantly focus on human disease, we use the terms HLA and MHC interchangeably.
they tune Natural Killer cell responses by controlling their cytotoxic activity. NK cells have both activating and inhibitory receptors on their surface and it is the balance between activating and inhibiting signals that determines the response of NK cells to infected cell targets. Many of these receptors are ligated by HLA class I molecules.

In this study, we primarily focus on the effect of HLA class I on CD8+ T cell responses.

1.1.1 Antigen presentation

MHC class I molecules present endogenous antigens such as viral peptides (9-12 amino acids long) on the surface of infected cell via an extensive process. During the synthesis of the viral material, some of the peptides: 1) undergo cleavage by the proteasome in the cytoplasm, 2) are transported by the Transporter associated with Antigen Processing (TAP) heterodimer to the endoplasmic reticulum (ER), 3) are loaded on partially folded MHC class I molecules and 4) are exported in a peptide:MHC (pMHC) complex to the cell surface via the Golgi apparatus. The binding strength of the pMHC complex is defined by the binding affinity (measured in the IC50 scale).

The viral peptides presented by the MHC class I molecules act as a signal to inform the immune system that the cell has become infected (see Figure 1.1). CD8+ cytotoxic T lymphocytes (CTLs) can receive this signal and eliminate the pathogen infected cell. CTLs recognise both a part of the MHC class I molecule and the viral peptide bound on it; a peptide that is recognised by the immune system is also called an epitope. If the self-MHC is not recognised then the viral peptide is also not recognised and this characteristic of the antigen presentation process is referred to as an MHC-restricted immune response. Furthermore, an affinity threshold of approximately 500nM (preferably 50nM or less) has been shown to determine the capacity of the pMHC complex to elicit a CD8+ T cell response [4]. Interestingly, significant functional differences have been reported between CD8+ T cells recognising identical peptides but restricted by
different, albeit closely related MHC class I molecules [5], suggesting that the response to antigen presentation is a complicated process controlled by multiple factors.

![Figure 1.1: Presentation of a viral peptide from HLA class I molecules to virus-specific CD8+ T cells. The image is adapted from [6].](image)

**Epitope binding prediction**

Peptide binding to MHC class I molecules is a key element in CD8+ T cell mediated immunity. There are many available *in silico* models that attempt to predict which viral peptides can bind specific HLA class I molecules and they are referred to as epitope prediction algorithms. These algorithms exploit properties of the pMHC binding complexes in order to obtain reliable predictions. The first epitope prediction models were based on sequence motifs shared by experimentally defined complexes [7, 8]. As the availability of data grew and quantitative measurements such as the affinity of the pMHC complex became possible more complex algorithms were developed. These include position-specific matrices which assign different probabilities that specific amino acids are found in different peptide positions [9], Hidden Markov Models [10] and machine learning approaches such as Artificial Neural Networks (ANNs) [11] and Support Vector Machines (SVMs) [12]. The latter algorithms can capture the influence of the sequence context on the binding contribution of a given amino acid in the binding peptide [13]. More elaborate algorithms integrate several steps of the antigen presentation
process in their models such as proteasomal cleavage and TAP-transportation [14,15]; generated peptides need to be properly ‘chopped’ at the proteasome, loaded and carried by the TAP complex before they can be displayed by the MHC class I molecules on the surface. This process may result in a limited number of peptides that can actually trigger a downstream response.

1.1.2 Polygenism

The MHC class I genetic locus consists of several different genes. Specifically, its human version, the HLA class I gene family includes, amongst others, three highly polymorphic bi-allelic loci: HLA-A, HLA-B and HLA-C\(^2\). The expression of these genes is co-dominant. Every individual possesses a set of different HLA class I molecules (minimum 3, maximum 6) with different ranges of peptide binding specificities.

1.1.3 Polymorphism

HLA genes are the most polymorphic genes in humans. For each of the different HLA class I molecules there are multiple variants of each gene within the whole population (see Figure 1.2). The fact that different people have different shapes (alleles) of the HLA class I molecules and thus present different parts of the pathogen (peptides) to the immune cells impacts on the effectiveness of an individual’s immune response.

Three models have been proposed to explain maintenance of HLA polymorphism in the population [16]: (1) balancing selection, where alleles considered protective against one disease can confer susceptibility to another, (2) heterozygote advantage, where a higher HLA repertoire increases the breadth of peptide recognition and immune defense against pathogens and (3) frequency-dependent selection, where a pathogen has evolved to escape an efficient immune response driven by alleles commonly found in the population but remains prone to responses mediated by rare, i.e. low-frequency, \(^2\)In this study when we refer to HLA class I genes we only consider the A, B and C loci.
alleles. Linkage association of neutrally selected elements to positively or negatively selected ones also plays a role [16]. However, establishing the relative importance of the three proposed mechanisms has been a challenging task. Interestingly, using a mathematical model, it has been shown that the heterozygote advantage on its own is insufficient to explain the high population diversity of the MHC molecules, even in a very large host population, and a high degree of polymorphism is only achieved under unrealistically similar allelic fitness contribution [17]. Using again a theoretical approach, another analysis has shown that provided a sufficiently large host population, selection for rare MHC alleles driven by host-pathogen co-evolution can account for realistic MHC polymorphism [18]. In a very recent study, the latter model was supported by experimental data demonstrating that antagonistic co-evolution is a viable mechanism explaining the evolution and maintenance of MHC polymorphism in vertebrate populations given fitness trade-offs associated with pathogen adaptation [19]. This mechanism predicts that MHC alleles that confer resistance to a subset of diseases can lead to susceptibility to another subset as a natural consequence. Nevertheless, this explanation is not excluding an additive effect from a heterozygote advantage-driven
1.1.4 Nomenclature

The nomenclature of HLA alleles has recently been revised [20][21]. Currently, an allele name may be composed of four, six or eight digits depending on its sequence. The different pairs of digits are separated by colons. The convention is to use a four-digit code to distinguish HLA alleles that differ in the proteins they encode. Alleles whose numbers differ in the first four digits must differ in one or more nucleotide substitutions that change the amino-acid sequence of the encoded protein. Based on hla.alleles.org the following rules are used for naming HLA molecules:

- The **1st and 2nd digits** describe the allele group.
- The **3rd and 4th** are assigned to specific HLA proteins.
- The **5th and 6th digits** distinguish alleles that differ only by synonymous nucleotide substitutions within the coding sequence.
- The **7th and 8th digits** discriminate between alleles that only differ by sequence polymorphisms in non-coding regions.

1.1.5 Associations with viral infection

The important role of HLA class I molecules in viral infection has made them the focus of many studies. Different HLA alleles have been significantly associated with different disease outcomes in a range of infections including HIV-1, HTLV-1 and HCV (the ones studied in this project). Importantly, many of the reported associations between HLA class I genes and the outcome of HCV, HIV-1 and HTLV-1 infection have been observed in independent cohorts. Both the impact of individual alleles and their homozygosity/heterozygosity status has been investigated. Several alleles have been suggested by these studies to have a protective (e.g. A*02 in HTLV-1, B*57 in
HIV and HCV) or detrimental (e.g. B*35 in HIV, B*54 in HTLV, C*04 in HCV) role in the progress of the infection [22–28]. HLA class I associations with disease outcome have also been shown in other viral infections including malaria, HBV and dengue virus infection [29–31] as well as autoimmune diseases such as ankylosing spondylitis [32].

1.2 CD8+ T cell function in viral infection

1.2.1 Background

CD8+ T cells can recognise the presence of antigen, in the context of MHC class I molecules, via their T Cell Receptors (TCRs). Antigen presentation can trigger a cascade of coordinated molecular interactions that may result in the limiting the infection. CD8+ T cells primarily recognise endogenous antigen which is directly presented by infected cells but they can also recognise exogenous antigen which is ‘cross-presented’ by other cells such as dendritic cells (DCs). Specifically, cross-presentation of antigen by DCs can stimulate naive CD8+ T cells and induce their proliferation; however, under normal circumstances it is probably less efficient than direct presentation, since it requires the additional step of transfer from one cell to another [33].

There are many lines of evidence, whilst not all equally strong, demonstrating the importance of CD8+ T cells in antiviral immunity. First, there are statistically significant associations between specific MHC class I molecules and the outcome of viral infection [22, 28, 30, 31, 34]. Second, the existence of escape mutants associated with CD8+ T cell pressure strongly implies their role in viral control [35, 36]. Third, there are in vitro experiments indicating that CD8+ T cells can inhibit replication [37, 38]. Fourth, there is a temporal correlation between CD8+ T cell appearance and viremic control [39, 40]. Fifth, and perhaps most convincingly, the in vivo depletion of CD8+ T cells, mainly in animal models, leads to higher viral burden in many viral infections [40, 41]. Sixth, the infusion of infected subjects with CD8+ T cells leads to
the dramatic, but transient, reduction of virus-positive cells [45].

1.2.2 CD8+ T cell effector mechanisms

CD8+ T cells can elicit two main effector mechanisms upon specific antigen recognition. Although the manifestation of the two effector functions is different, they both can lead to the limitation of infection.

Lytic function

The lytic function is the one that gives CD8+ T cells the name cytotoxic T cells. Under a lytic mechanism the extracellular secretion of perforin [46], a pore-forming protein, allows the breach of the membrane of the infected cell followed by the entrance of granzyme A and B proteins [47] in the target cell cytoplasm which eventually prompt the degradation of the cells DNA. Cytotoxic CD8+ T cells can also kill through the Fas/CD95 pathway which requires neither calcium nor perforin [48]. Binding Fas, on the infected cell, with its ligand (FasL), on the T-cell, activates the caspase cascade. In culture, the granule pathway is the one that mainly destroys target cells and only when this pathway is compromised there are significant levels of Fas-mediated killing; however, the situation is more complicated in vivo [49]. Under both pathways, once the infected cells undergo apoptosis, phagocytic cells can identify and ingest them.

Non-lytic function

The non-lytic CD8+ T cell function involves the secretion of non-cytolytic soluble suppressor factors such as cytokines and chemokines. Several studies on different viral infections have provided evidence supporting this effect. It has been shown that certain immunoregulatory cytokines, including type I and II interferons (IFN-α,β and IFN-γ respectively) and tumour necrosis factor (TNF) can activate a number of intracellular pathways that directly suppress viral replication without killing the host cell [50–53].
Interferons have been shown to suppress hepatitis B virus (HBV), hepatitis C virus (HCV), herpes simplex virus, vesicular stomatitis virus, vaccinia virus, picornaviruses, retroviruses, influenza viruses and other types of viruses in vitro in a non-cytolytic manner [50,54]. Interestingly, clinical studies have shown that HBV replication can be suppressed by acute hepatitis A virus-induced production of soluble factors including IFN-γ [55]. This suggests a non-specific ‘bystander’ suppression effect which can occur in certain tissues and viral systems [50] and might act on a short but not necessarily strictly adjacent distance.

In this thesis we particularly study the non-lytic CD8+ T cell response in the context of HIV/SIV infection. It has been shown that in HIV infection, suppressor activity can be mediated by diverse soluble factors [50]. Many studies have explored the role of these soluble factors. 1) The ‘mystery’ suppressor named CD8 Antiviral Factor (CAF) [56] was the first reported to be released by primary CD8+ T cells upon activation in vitro. Studies indicated that an unknown factor is responsible for CAF activity, but did not eliminate the possibility that it is a collection of known antiviral cytokines with redundant functions [50]. 2) Secretion of IFN-γ by activated T cells has been shown to inhibit HIV in macrophages and was associated with viral suppression and a lack of disease progression [57]. 3) Activated CD8+ T cells or PBMC obtained from HIV-infected individuals produced increased levels of MIP-1α, MIP-1β and RANTES [58,59]. These are CC-chemokines that bind to and activate the chemokine receptor CCR5 blocking the entry of HIV-1 strains that use it as a co-receptor (R5). 4) Other agents that have been studied as potential viral suppressors are: IFN-α, Macrophage Derived Chemokine (MDC), IL-13 and Leukemia Inhibitory Factor (LIF) [60]. Depending on the secreted factor, CD8+ T cells can either block viral production from infected cells or ‘protect’ uninfected cells from viral entry.
1.2.3 Evading CD8+ T cell responses

CD8+ T cells target epitopes derived from viral proteins and also cryptic epitopes encoded by viral alternative reading frames (ARF). On the other hand, viruses ‘exploit’ opportunities and ‘formulate’ mechanisms in an attempt to avoid recognition and elimination.

The main evasion mechanism employed to thwart CD8+ T cell activity is mutational escape from antigen-specific responses [61][62]. If a genetic mutation in a viral peptide occurs during the antigen presentation process and abrogates its recognition by a specific CD8+ T cell response then the variant viral strain might acquire a survival advantage over the wild-type strain. However, the same mutation might also involve a fitness cost for the virus. This cost usually occurs because of structural constraints influencing the kinetics of viral replication; specifically, conserved viral regions are often considered vital to viral replication. The balance between the survival advantage and the fitness cost will define the outgrowth rate of the variant strain over the wild-type. A mutation that does not offer a survival benefit is expected to eventually die out unless it is neutral and in linkage disequilibrium to one that does.

Interestingly, the loss of recognition of a specific epitope due to a viral mutation can result in a new response against subdominant epitopes [63]. On the other side however, if a mutation bares a fitness cost for the viral replication, compensatory viral mutations that restore the viral replicative capacity can arise [64][65].

Importantly, the mutations of the viral peptides are frequently associated with the presence of specific MHC class I molecules. Identical MHC class I genotypes are associated with very similar escape variants suggesting viral constraints [66]. The importance of MHC-restriction in viral escape evolution is also supported by the reversion to the wild-type strain in the absence of the restricting MHC class I molecule [67][68] albeit not always [69].

Apart from mutations that affect epitope presentation, there are other strategies
that viruses employ in order to avoid immune control. For example, in the case of HIV-1, the viral accessory protein Nef has been shown to downregulate MHC class I expression. Nef can downmodulate all HLA-A and HLA-B but not HLA-C and HLA-E allotypes [70]. This selective downregulation may lead to decreased antigen presentation to CD8+ T cells, whilst simultaneously avoiding NK-mediated killing by retaining HLA-C and HLA-E expression [71].

1.3 Killer Cell Immunoglobulin-like receptors

1.3.1 Background

Killer cell immunoglobulin-like receptors (KIRs) are a family of transmembrane proteins that are expressed on natural killer (NK) cells and subsets of T cells [72,74]. They are both polymorphic and polygenic and are found at chromosome 19. KIRs bind HLA class I molecules and have both activatory (DS) and inhibitory (DL) isoforms [75]. For example, KIR2DL2 binds group C1 HLA-C molecules which have asparagine at residue 80, and with a weaker affinity, group C2 alleles which have a lysine at position 80 [76] while KIR3DL1 binds Bw4 alleles which are distinguished from Bw6 alleles by the motif spanning amino acid positions 77-83 [77]. Although for most of the inhibitory KIRs their ligands are known, this is not true for activatory KIRs [78] (Table 1.1).

KIRs contribute both directly and indirectly to antiviral immunity. Directly, KIRs on NK cells sense the loss of HLA class I molecules from the cell surface and trigger NK-mediated cytolysis. Additionally, inhibitory KIRs expressed directly on T cells have been suggested to increase cell survival by reducing activation-induced cell death [79,80]. Indirectly, NK cells can regulate adaptive immunity via crosstalk with dendritic cells and by the production of chemokines and cytokines [81,82].
Table 1.1: Groups of HLA class I molecules which are known ligands for KIRs. In this study we denote HLA-C<sup>Asp80</sup> as HLA-C1 and HLA-C<sup>Lys80</sup> as HLA-C2.

### 1.3.2 Associations with disease

Early research on KIRs investigated protection and/or susceptibility for disease by studying associations with KIRs in the context of their HLA class I ligands. In HCV infection, homozygosity of KIR2DL3 and its HLA-C1 ligand has been associated with viral clearance [83] while the HLA-Bw4I80/KIR3DS1 compound has been shown to have a protective effect against the development of HCV-associated hepatocellular carcinoma [84]. In HIV, the epistatic interaction of KIR3DS1 and HLA-B delays the progression to AIDS [85] while various distinct allelic combinations of the KIR3DL1 and HLA-B loci have significant and variably strong influence both on protection from progression to AIDS and plasma HIV RNA abundance [86]. In agreement with the latter result, the presence of the inhibitory allele KIR3DL1 in combination with the HLA-B*57 allele had a highly protective effect against progression to AIDS in Zambian patients [84]. In chronic myeloid leukemia, KIR2DL2 and/or KIR2DS2 in the presence of its ligand was found to be protective [87]. In [88], susceptibility to Crohn’s disease is shown to be mediated by KIR2DL2/KIR2DL3 heterozygosity and their HLA-C ligand. Also in psoriatic arthritis, individuals with activating KIR2DS1 and/or
*KIR2DS2* genes were found to be susceptible to developing disease, but only when HLA ligands for their homologous inhibitory receptors, KIR2DL1 and KIR2DL2/3, were missing [89]. An extended review about the role of KIRs in disease is given in [90].

Although the associations of KIR genes (with or without their ligands) with disease outcome are many, they are not always confirmed across different cohorts and they can be puzzling and pointing to contradictory suggestions. In the case of HIV for example, both the inhibitory receptor *KIR3DL1* and its activatory counterpart, *KIR3DS1* are associated with protection from disease progression making the underlying mechanism hard to decipher.

### 1.4 Persistent viral infections

#### 1.4.1 Hepatitis C

**Epidemiology**

Hepatitis C is among the most frequent viral infections in humans with 170 million infected people worldwide. It is caused by the Hepatitis C virus (HCV) which is a member of the Flaviviridae family. Infected individuals show considerable heterogeneity in the outcome of infection. HCV persists in approximately 70% of infected individuals while the rest spontaneously clear the infection (usually within the first 6 months). Chronic HCV infection can cause serious liver damage such as cirrhosis, liver failure and hepatocellular carcinoma [91].

**Virology**

The hepatitis C virus (HCV) is a small positive-stranded RNA virus (≈ 9.6Kb). It encodes a single large polyprotein which is 3000 amino acids long (see Appendix Figure
During the post-translational stage the polyprotein is chopped in 10 distinct mature viral proteins unless there is a frame-shift and then 11 proteins are produced (with the inclusion of F). There are 3 structural proteins (Core, E1, E2) and 7 non-structural proteins (P7, NS3, NS4A, NS4B, NS5A, NS5B). A description of the structure and function of the HCV proteins is given in [92, 93]. HCV, like other hepatitis viruses, is predominantly a hepatotropic virus.

**Treatment**

The available treatment is based on pegylated-interferon-α (PEG-IFN-α) plus ribavirin (RBV) for patients with chronic hepatitis C infection and achieves sustained virologic response (SVR) in 40%-52% of treated patients infected with HCV genotype 1 [94]. The SVR can be higher for genotype 2 and 3 [95]. However, side effects are common and sometimes serious, leading to premature termination of treatment in many patients.

**Vaccine**

Currently, there is not an available preventative or therapeutic HCV vaccine although promising attempts are being made [96, 97].

**Immunogenetics**

In HCV infection, similar to other viral -or even non-viral infections- the origins of the observed heterogeneity in disease outcome are not yet understood. Several host genetic factors are suggested as key players in HCV spontaneous clearance. Some of the most important factors are: 1) several HLA class I (see Appendix Table A.1) and II alleles; B*57 and C*01 which can have a protective effect and most importantly have been replicated in several independent cohorts [24, 26, 98, 99], 2) compounds of KIRs with their relevant HLA class I ligands which are associated with disease outcome; the HLA-C1/KIR2DL3+KIR2DL3+ compound has been linked to HCV clearance [83, 100].
while the HLA-Bw4l80/KIR3DS1 compound has been shown to have a protective effect against the development of HCV-associated hepatocellular carcinoma [84], 3) the single nucleotide polymorphism (SNP) rs12979860 which is found 3kbs upstream of the IL28B gene (encodes the type III interferon, IFN-λ3) has also been linked with both natural and treatment-associated control of HCV in multiple studies [101, 102]. Other SNPs that influence hepatitis C outcome [103, 105] include polymorphisms in genes such as interleukins, chemokines and interferon-stimulated genes and have been linked to both spontaneous and treatment-controlled infection [106, 109].

**CD8+ T cell responses**

Strong and maintained virus-specific CD4+ and CD8+ T cell responses are thought to be required for spontaneous viral clearance and can be detected in resolved patients for more than 20 years after successful elimination of HCV [110]. There are several lines of evidence supporting a strong role for specific CD8+ T cell in HCV infection including immunogenetic data of HLA class I associations with disease outcome. Many studies report enriched mutational changes in experimentally described or predicted epitopes in patients with the restricting-HLA genotype [62, 111–115] suggesting the exertion of a strong CD8+ T cell response. Additionally, clearers have been shown to mount significantly broader CD8+ T cell responses of higher functional avidity and with wider variant cross-recognition capacity than non-clearers [116]. In chimpanzees, antibody-mediated depletion of CD8+ T cells before re-infection with HCV led to prolonged virus replication until HCV-specific CD8+ T cells recovered in the liver [40]. Furthermore, suppression of acute viremia in vaccinated chimpanzees occurred as a result of massive expansion of peripheral and intrahepatic HCV-specific CD8+ T lymphocytes that cross-reacted with vaccine and virus epitopes [117].
Animal models

The only existing HCV animal model is the chimpanzee model which involves a lot of limitations since the experimentally infected chimpanzees may clear HCV infection more readily than humans, and those that develop persistent infection typically show mild disease characteristics [91]. Of course, the ethics of animal experimentation limit the use of animal models in scientific research.

Challenges

There are many challenges posed in the study of the hepatitis C virus such as the high mutation rate of the virus - $1.5 - 2 \times 10^{-3}$ base substitutions per genome site per day [118], the high replication rate - $10^{12}$ virions per day [119] and the high diversity of HCV sequence - 7 major genotypes and more than 80 subtypes [110]. Additionally, the existence of quasispecies - different but closely related viral genomes in the same host [120] and the lack of a small animal model burden the discovery of a potential vaccine.

1.4.2 Human T-Lymphotrophic Virus 1

Epidemiology

Human T-Lymphotrophic Virus 1 (HTLV-1) is a persistent retrovirus that infects 10-20 million people worldwide. Most infected individuals remain lifelong asymptomatic carriers (ACs). However, approximately 1% of infected individuals develop virus-associated diseases including HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP), an inflammatory disease of the central nervous system that results in progressive paralysis. In addition, another 2-3% develop Adult-T cell Leukemia (ATL) and a small number of other less well-defined inflammatory disorders. It is poorly understood why some individuals remain asymptomatic whereas others develop disease,
but one strong correlate of disease is the proviral load, which is significantly higher in HAM/TSP patients than in ACs \cite{121}.

**Virology**

HTLV-1 is a human retrovirus. Once the diploid genome of HTLV-1 is copied into a double stranded-DNA ($\approx 8.5$kb) form, it is integrated in the genome of the host. Then the virus is referred to as provirus and the viral load of the host is usually quantified as proviral load (PVL). The HTLV-1 proteome consists of two structural proteins (Gag, Env) and 10 non-structural proteins (Pro, Pol, Rof, P12, Tof, P13, Rex, P21, Tax, HBZ). About 90-95% of the HTLV-1 proviral load is carried by CD4+ T cells and only 5-10% is carried by CD8+ T cells \cite{122}.

**Treatment**

Different therapies, mainly anti-inflammatory agents, have been considered for the treatment of HAM/TSP. These include corticosteroids \cite{123}, nucleoside analogues such as zidovudine and lamivudine \cite{124} and valporic acid (VPA) \cite{125}. Corticosteroids may help ease the disease symptoms -although not rigorously tested- but treatments with nucleoside analogues and VPA have shown no significant decrease of proviral load or improvement of HAM-TSP related symptoms. However, in \cite{126} the combined treatment with valproate and azidothymidine is shown to be a safe and effective means to decrease PVL *in vivo* in Simian T-Lymphotropic Virus 1 (STLV-1) infection.

**Vaccine**

There is no available vaccine for HTLV-1 related diseases.
Immunogenetics

HLA class I molecules have been reported to influence both disease status and viral load. Specifically, \textit{HLA-A*02} and \textit{C*08} are associated with a reduced risk of HAM/TSP and a reduced proviral load in ACs, \textit{HLA-B*54} is associated with an increased prevalence of HAM/TSP and an increased proviral load in HAM/TSP patients \cite{23,127}. Other genetic factors that influence the risk of HAM/TSP include a polymorphism in the TNF-\(\alpha\) promoter, the cytokine gene IL-15 and the chemokine gene SDF-1\(\alpha\) \cite{128}.

CD8+ T cell responses

Apart from the association of specific HLA class I molecules with disease outcome, additional evidence for the ability of CD8+ T cells to control the infection include the association of high mRNA levels of cytolytic genes in CD8+ T cells with low HTLV-1 proviral load \cite{129}, the spontaneous killing of HTLV-1 expressing cells by autologous CD8+ T cells \textit{in vitro} \cite{130} and \textit{ex vivo} \cite{131} and the higher variation of Tax coding sequences in ACs compared to HAMs that suggests a higher CD8+ T cell selection pressure exerted in ACs \cite{132}.

Animal models

There are no animal models that can mimic human HTLV-1 infection reliably. However, the experimental infection of rabbits, rats and non-human primates has been reported in the literature \cite{133,135} but none of these animal models develops inflammatory disease of the central nervous system that is similar to HAM/TSP in humans.
Challenges

The lack of a reliable small animal model and the absence of samples derived during acute infection are key challenges in the understanding of the reasons differentiating ACs from HAM/TSP and ATL patients.

1.4.3 Human Immunodeficiency Virus 1

Epidemiology

Based on the 2010 Joint United Nations Programme on HIV-1/AIDS (UNAIDS) there are worldwide approximately 33 million individuals infected with Human Immunodeficiency Virus 1 (HIV-1). The number of new incidents is declining, mainly due to prevention efforts, and there is a higher survival period for HIV-1-infected individuals thanks to effective treatments.

Virology

HIV-1 is a retrovirus composed of two copies of single-stranded RNA (≈ 9.8kb) enclosed by a capsid. The HIV-1 genome consists of Gag, Pol, Env, the transactivators Tat, Rev, Vpr, other regulators, Vif, Nef, Vpu and rarely Tev. Pol codes for the viral enzymes reverse transcriptase, integrase, and HIV protease which are vital for the infection of new targets. The tropism of the virus depends on the Env protein defining the viral envelope which enables the virus to attach to and infect target cells. Viruses that bind the CCR5 chemokine receptor are macrophage tropic (M-tropic or R5) and can infect CD4+ T cells, macrophages and dendritic cells. Conversely, the lymphotropic strains (T-tropic or X4) can enter only CD4+ T cells and use a different chemokine receptor known as CXCR4, which is not bound by CCR5 ligands.
Treatment

The treatment of HIV-1 mainly consists of a variety of Antiretroviral Drugs (ARVs) which target different viral proteins and therefore different phases of the viral cycle. The drugs include: 1) entry inhibitors, 2) nucleoside/nucleotide reverse transcriptase inhibitors, 3) non-nucleoside reverse transcriptase inhibitors, 4) integrase inhibitors and 5) protease inhibitors. When taken in combination, the treatment is known as Highly Active Antiretroviral Therapy (HAART) and although effective can have serious side-effects. The ARV therapy can fail, amongst other reasons, because of low adherence to the treatment regimen or the emergence of viral strains resistant to the drugs.

Vaccine

Unfortunately, there is no available preventative or therapeutic vaccine although vaccines based on both humoral and cellular responses are being considered. T-cell based vaccines might only be able to control the infection by lowering the viral load but not provide sterilising immunity [136]. However, containment of viral load would not only enhance survival but also limit the risk of transmission [137]. Many vaccines are in pre-clinical or clinical trials but the high viral diversity and the challenges associated with generating broadly reactive neutralizing antibodies and cellular immune responses are key obstacles to be overcome [138]. For both nucleic acid- or protein-based vaccines, another difficulty will be to achieve a high effector/target ratio during the viral expansion phase [139]. Two big T cell based vaccine studies, the STEP [140] and the RV144 [141], showed none to limited protective effect. Many researchers now believe that vaccine candidates need to induce both sustained broadly neutralizing antibodies and a strong cell-mediated response [142].
Immunogenetics

In many but not all genome-wide association studies (GWAS) in multi-ethnic cohorts of HIV-1 infected individuals, HLA class I molecules are found to be the major genetic determinants of HIV-1 control. The HLA class I allele B*57 is perhaps the one most consistently associated with both slower disease progression and better control of viral replication. HLA-B*27 has been shown to have a protective effect on progression while HLA-B*35 (specifically, a subset of HLA-B*35 molecules) has a detrimental effect on outcome. Furthermore, the haplotype B*35/Cw*04 is associated with rapid disease development in Caucasians. In macaques, MHC class I molecules also influence the outcome of SIV infection.

Additionally, more rapid disease progression is observed in HLA class I homozygote individuals and in line with this, maximum HLA class I heterozygosity (A, B, and C) delays disease progression. In a very well-defined study in SIV-infected macaques, a clear protective effect of a MHC class I heterozygote advantage is demonstrated; the animals are all infected by the same strain and they share a low number of MHC alleles forming as such the best system to study this effect.

Apart from HLA class I genotype, KIR/HLA compounds are also associated with disease outcome. In a sample size of over 1,500 HIV-infected individuals, distinct allelic combinations of the KIR3DL1 receptor and the HLA-B loci (including B*57) were associated with slow progression to AIDS and decreased plasma HIV RNA abundance. In another study, an epistatic interaction between KIR3DS1 and HLA-B alleles delays the progression to AIDS. However, the ligand for the KIR3DS1 receptor is not well-defined and HLA-B Bw4-80I alleles that encode molecules with isoleucine at position 80 are thus far only a putative binder; only a rare KIR3DS1 allotype (KIR3DS1*014) has been experimentally observed to bind HLA-Bw4 molecules. These findings have been suggestive of a critical role for NK cells in the natural history of HIV infection and although there are some functional data consistent with
this hypothesis, the underlying mechanism of the KIR/HLA effect on disease outcome remains unclear.

**CD8+ T cell responses**

Indirect evidence for the role of CD8+ T cell responses in controlling HIV infection include specific HLA class I genes that are consistently associated with disease outcome and selection of viral escape mutants which avoid CD8+ T cell recognition and lead to the loss of immune control. Direct evidence come from human studies, where infusion of HIV-1 infected subjects with CD8+ T cells led to the dramatic, but transient, reduction of HIV RNA-positive cells and animal studies where SIV-macaques depleted of their CD8+ T cells exhibited a high rise in plasma viremia. Furthermore, in a cohort of 578 untreated HIV-infected individuals from KwaZulu-Natal, increasing breadth of Gag-specific responses has been associated with decreasing viremia while increasing Env breadth with increasing viremia. However, it is important to note that the impact of T-cell responses on the control of viral replication cannot be explained by the mere quantification of the magnitude and breadth of the CD8+ T-cell response and the quality of the response (polyfunctional, avid etc.) can also be crucial for viral control. Finally, a temporal correlation between CD8+ T cell appearance and viremic control has also been reported.

**Animal models**

There are many animal counterparts of HIV with the main one being the Simian Immunodeficiency Virus 1 (SIV-1) which infects many African non-human primate species. Chimpanzees were initially studied but then there was a switch to studying monkeys, in which AIDS develops sooner, in order to expedite the testing of new hypothesis. In contrast to HIV-infected humans, the natural SIV hosts (for example, sooty mangabeys, African green monkeys and mandrills) typically do not develop
AIDS despite chronic infection with a highly replicating virus [172]. Limited immune activation and preserved mucosal immunity are the two main mechanisms which may explain why SIV infection of natural hosts remains non-pathogenic [173]. However, in many species such as rhesus macaques, SIV infection resembles many features of HIV infection [173] and can ultimately generate new approaches that can lead to the development of a successful HIV vaccine [172, 173]. Humanised mouse models [174] and feline models [175] are also being considered.

Challenges

There are many obstacles to surpass before obtaining a successful HIV vaccine. Many of these challenges are similar to the ones involved in the study of HCV infection. The virus has extensive clade and sequence diversity, it can evade both humoral and cellular immune responses and it can establish latent viral reservoirs quite early in infection [176]. Additionally, no small animal model exists.

1.5 Thesis Outline

The present thesis focuses on the study of the CD8+ T cell response to persistent viral infections (see Figure 1.3). In Chapters 2-4 we study the immunogenetics of persistent viral infections, focusing on the effect of Human Leukocyte Antigen (HLA) molecules and Killer-cell Immunoglobulin-like Receptors (KIRs) on disease outcome and viral burden. In Chapters 5-7, we investigate CD8+ T cell effector mechanisms and particularly the dynamics of lytic and non-lytic CD8+ T cell responses and how these relate to 'viral escape'. The persistent viral infections considered in this thesis are HTLV-1, HCV and HIV. The methods applied include statistical models, high-throughput sequence analysis, ordinary differential equation models and agent-based models. Specifically,

In chapter 2, HLA class I-mediated determinants of outcome are investigated
in the context of HCV infection. These are: 1) the heterozygote advantage, 2) the rare allele advantage, 3) individual alleles, 4) allelic breadth, 5) protein specificity and 6) epitope specificity. Furthermore, the impact of HLA class I on disease outcome is quantified for several molecules and for three different viral infections using the Explained Fraction.

In chapter 3, known HLA class I associations with disease outcome are studied in different KIR genetic backgrounds for both HCV and HTLV-1 infections. The effect of KIRs on HLA class I-mediated immunity is measured in terms of both viral burden and disease status. Furthermore, potential effector mechanisms that can explain the findings are discussed.

In chapter 4, a preliminary study design for testing the hypothesis that KIRs can influence the immune pressure exerted by CD8+ T cells during HCV infection is presented and discussed.

In chapter 5, an extensive set of differential equation models that describe lytic and non-lytic CD8+ T cell responses to SIV/HIV infection are fitted to experimental viral load, CD4+ and CD8+ T cell data of SIV-infected macaques in order to investigate the different CD8+ T cell response mechanisms.

In chapter 6, the development and implementation of a 3D Cellular Automaton (CA) model of acute and chronic HIV/SIV infection is described. A CA is an individual-based computer simulation of a system that evolves in time on a multidimensional (usually 2D or 3D) lattice of nodes and simulates both spatial and temporal characteristics of a system.

In chapter 7, the CA model is used in order to study the efficiency of lytic and non-lytic CD8+ T cell effector mechanisms and how they shape the dynamics of viral escape. Using an extensive set of simulations, the hypothesis that non-lytic CD8+ T cell effector function can lead to the outgrowth of the variant strain over the wild-type is tested and compared to the outgrowth observed under a lytic effect. In addition, the quantification of CD8+ T cell lytic killing rate is discussed and two different models:
mass action killing and saturated killing are explored.

Finally, in CHAPTER 8, the main conclusions of the thesis are discussed along with suggestions for future avenues of research originating from our findings.

![Figure 1.3: Schematic outline of the thesis.](image)

Many of the results presented in the following chapters have been published in peer-reviewed journal articles listed in the ‘Dissemination’ section.
Chapter 2

HLA class I impact on HCV and other viral infections

Part of the analysis and findings presented in section 2.5 have been published in Elemans M., Seich al Basatena N.-K. and Asquith B., PLoS Comput. Biol. 8(2), 2012 [177].

2.1 Aim

Our aim was to investigate the impact of HLA class I-related factors on the outcome of HCV infection and quantify the influence of HLA class I molecules on the outcome of HCV, HTLV-1 and HIV-1 infections.

2.2 Introduction

An effective CD8+ T cell response is considered crucial in controlling many viral infections. Polyfunctionality, frequency, breadth, avidity, specificity as well as location and persistence of the response are likely to be important factors in shaping its effectiveness. However, measurements of CD8+ T cells such as frequency, phenotype and function are informative but readily influenced by antigen load; hence, it can be difficult to ascertain if a particular observation is the cause or the effect of the immune control. An alternative approach is to examine CD8+ T cell mediating host genotype factors such as HLA class I genes, where the direction of causality is unequivocal. HLA class I alleles and other related factors such as zygosity, frequency, allelic breadth and
specificity can potentially explain part of the heterogeneity observed in the outcome of viral infection. In this study, we focus on the impact of individual alleles, heterozygote advantage, rare allele advantage and binding peptide:HLA properties such as allelic breadth, protein specificity and epitope specificity on disease outcome, mainly in HCV infection and partly in HIV and HTLV-1 infection.

In HCV infection, many HLA class I molecules (A, B and C), have been linked to spontaneous viral clearance but also to viral persistence that can lead to cirrhosis, liver failure or even hepatocellular carcinoma. Although many HLA class I alleles have been reported in the literature as significant determinants of outcome (see Appendix Table A.1 for an extensive list), very few have been replicated in independent studies (e.g. \textit{B*57} \cite{24, 25, 98, 178}). This can be the result of a high false positive discovery rate but it could also be explained by ethnic and geographical differences between the cohorts studied. Additionally, the outcomes of infection explored are not always the same; some studies investigate healthy individuals versus infected individuals while others compare cases of different disease progression (e.g. spontaneous clearance versus viral persistence). Furthermore, since HCV is characterised by the existence of many quasispecies, different cohorts might be infected with various viral strains leading to differences in HLA class I-mediated immune control; a clear example is the study of a large cohort of HCV genotype 1b-infected Irish women with a single strain \cite{26} which failed to detect a protective effect of \textit{HLA-B*57}. The virus in the Irish outbreak is suggested to contain polymorphism that are different from the general genotype 1b consensus sequence and would not allow recognition by the \textit{HLA-B*57}-restricted responses characterized in the other reports \cite{178}. However, when specific HLA class I alleles are repetitively associated with outcome across independent studies they are highly likely genuine determinants of disease course.

Apart from individual HLA class I alleles, HLA class I zygosity can also influence disease status. In the context of viral infections, under the ‘heterozygote advantage’ hypothesis (also referred to as overdominant selection) MHC heterozygous individuals
can present a more diverse repertoire of viral peptides to the immune system than homozygous individuals resulting in a potentially beneficial effect that may even delay the emergence of viral escape mutants. In HIV infection, maximum heterozygosity of all HLA class I loci delayed progression to AIDS among HIV-1 infected individuals whereas individuals who were homozygous for one or more loci progressed rapidly to AIDS and death [22]. These findings were confirmed by another study relating HLA class I homozygosity and rapid progression to late-stage HIV-1-related conditions [159]. An MHC heterozygote advantage has also been shown in SIV-infected macaques [160]. In HTLV-1 infection, HLA class I heterozygosity has an effect on lowering HTLV-1 proviral load which is a known important factor in the risk of developing HTLV-1 associated disease. In HCV infection, evidence for heterozygote advantage has been reported for HLA class II alleles of the HLA-DRB1 supertype [179] and similarly, heterozygotes for HLA class II alleles have exhibited protection against hepatitis B virus infection [180].

Another HLA-mediated factor that can potentially influence the epidemic of infectious diseases is the ‘rare-allele advantage’, also termed as frequency-dependent selection. Individuals who express rare MHC molecules might have a selective advantage over those that express common alleles to which the pathogen has adapted. However, as the selected allele increases in frequency, the virus will adapt to it and the advantage will be lost. In HIV infection, there is a positive correlation of the frequency of HLA supertypes with viral load suggesting a selective advantage for individuals with a rare supertype [181]; HLA class I supertypes can be considered as a plausible classification scheme for alleles with similar peptide binding specificities. In the case of malaria infection in the Gambian population, a frequency-dependent selection may have caused the increase of the rare in other racial groups and protective allele, HLA-B*53 [29,182].

Although specific HLA class I alleles are associated with a protective or detrimental effect on outcome, the reason underlying this effect remains in most cases unknown. Importantly, it is both the HLA class I molecule and the viral peptide bound by it that
trigger CD8+ T cell recognition and activation. Therefore, the study of pMHC binding properties such as breadth (number of binding peptides), protein and epitope specificity (preference for binding a specific viral protein or a specific viral peptide) can help in that direction. A positive correlation between pMHC binding strength, although for a limited amount of complexes, and eliciting of CD8+ T cell responses has been shown in [4, 183], indicating an impact of MHC binding affinity on immunogenicity. Crucially, the constantly improving epitope prediction software that can identify in silico which viral peptides can form complexes with specific MHC molecules facilitates such an analysis. Using epitope prediction software and focusing on pMHC binding properties, a study showed that in HIV-1 infection, HLA class I alleles associated with slow progression to AIDS prefer to present the p24 Gag viral protein while non-protective HLA alleles preferentially target the viral protein Nef [184]. In HTLV-1, the same approach revealed that HLA Class I binding of the viral protein HBZ determines the outcome of the infection [185]. Here, using epitope prediction software, we also explore whether pMHC binding properties can influence disease outcome in the context of HCV infection.

In the first part of this Chapter, using two large HCV cohorts, we study three HLA class I-mediated factors which have been associated with the outcome of viral infection: 1) the heterozygote advantage, 2) the rare allele advantage and 3) individual alleles. In the second part we use epitope prediction software to obtain a list of binding peptides for each HLA class I molecule present in the cohorts and study whether 1) allelic breadth, 2) protein specificity or 3) epitope specificity can explain the variability in the outcome of HCV infection (viral clearance v persistence). We, also quantify the role of individual HLA class I alleles in explaining heterogeneity of disease outcome in three different viral infections, namely HCV, HTLV-1 and HIV.
2.3 Methods

2.3.1 Data

Cohorts The HCV cohort consists of three sub-cohorts from the US: AIDS Link to Intravenous Experience (ALIVE, N=262) [186], Multicenter Hemophilia Cohort Study (MHCS, N=320) [187], Hemophilia Growth and Development Study (HGDS, N=110) [188] and a UK cohort (N=341) [83]. 251 individuals were excluded due to incomplete information. The cohort had 257 resolved and 525 chronic patients. The HTLV-1 cohort (N=431) [23] consists of individuals recruited in Japan. All individuals were of Japanese ethnic origin and resided in Kagoshima Prefecture, Japan. The cohort includes 229 HAM/TSP patients and 202 asymptomatic carriers (ACs). The HLA and KIR genotyping, as well as the viral load measurements were obtained by collaborators.

HLA genotyping For the HCV cohort, genomic DNA was amplified using locus specific primers as described in [83]. The resulting PCR products were blotted onto nitrocellulose membranes and hybridized with sequence specific oligonucleotide probes (SSO). US: alleles were assigned according to the reaction patterns of the SSO probes, ambiguities were resolved by sequencing analysis. UK: PCR products were typed by direct sequencing. HLA types that were not resolved by sequencing or which gave unusual results were also tested by SSO typing using commercial kits (Dynal, RELI SSO, Wirral, UK). For the HTLV-I cohort, 96 PCR-sequence-specific primer reactions were performed to detect all known HLA-A, B, and C specificities in an allele or group specific manner [189].
2.3.2 Statistical analysis

Multiple logistic regression

For studying disease status the multiple logistic regression model is used (see Equation 2.1). The outcome of the infection (response) which is a binary variable (HCV:1-Cleared and 0-Persisted and HTLV-1:1-ACs and 0-HAM/TSPs) is studied based on potential explanatory variables, $x_i$, such as HLA class I mediated factors (e.g. zygosity, breadth etc.) and potential confounders (see below). Some of the explanatory variables are categorical (e.g. HBV status, mode of infection etc.) and some others are continuous (e.g. breadth and specificity). The available cohorts (in HCV) are studied separately and also pooled in order to gain more statistical power. The Odds Ratio (OR) is given by the exponent of the coefficient, $b_i$, of the relevant covariate, $x_i$. The OR is a descriptive statistic that is a measure of effect size. It describes the strength of association between two binary data values (e.g. having a specific HLA class I molecule and the outcome of infection). Based on the logistic regression model we also obtain the two-tailed p-values for the statistical significance of the explanatory variables. The model is fitted using the statistical software R v2.9.2.

$$\text{logit}(p) = \ln \left( \frac{p}{1-p} \right) = b_0 + b_1 x_1 + ... + b_k x_k \quad (2.1)$$

where $p$ is the probability of the outcome of interest, e.g. spontaneous clearance, $b_0$ is the intercept and $b_i$ is the change in $\text{logit}(p)$ for a one unit change in the predictor $x_i$.

Multiple linear regression

For studying viral burden the multiple linear regression model is used (see Equation 2.2). The logarithms of the viral load (HCV) or the proviral load (HTLV-1) and not the raw measurements are considered in the model so that the data are normalised. The
viral burden (response), \( y \), which is a continuous variable is studied based on potential explanatory variables, \( x_i \), such as HLA class I mediated factors (e.g. zygosity, breadth etc.) and potential confounders (see below). Some of the explanatory variables are categorical (e.g. HBV status, mode of infection etc.) and some others are continuous (e.g. breadth and specificity). The coefficients of the continuous explanatory variables, \( b_i \), indicate the increase or decrease in log-(pro)viral load when this variable, \( x_i \), changes by 1 unit. If the explanatory variable is binary then the coefficient indicates the difference in mean log-viral load between the categories. Based on the linear regression model, we also obtain the two-tailed p-values for the statistical significance of the covariates. The available cohorts (in HCV) are studied separately and also pooled in order to gain more statistical power. The model is fitted using the statistical software R v2.9.2.

\[
y = b_0 + b_1 \vec{x}_1 + ... + b_k \vec{x}_k
\]  

(2.2)

Confounding factors

In the literature, there are examples which suggest that HLA class I associations studies can be confounded by several other factors. To account for potential confounding in our models we first examine which variables have an impact on disease status or viral burden and then we include them in the model used to explain the variation in the outcome of infection or in viral burden. For HCV infection the confounding factors included in the model are (Appendix Table A.2): mode of infection, HBV status, the HLA-C1C1/KIR2DL3++ compound, the SNP rs12979860 and the cohort. For the HTLV-1 infection the confounding factors are age and gender [23]. For the HCV cohort, we first consider the UK and USA cohorts separately to investigate whether the results can be replicated and then we pool the cohorts to gain more statistical power.
Linkage Disequilibrium

Linkage disequilibrium (LD) occurs when genotypes at two loci are not independent of each other. Similarly, we say that two alleles are in positive LD when their occurrence together happens significantly more often than expected by chance while we refer to negative LD when they segregate significantly more often than expected by chance. The LD for HLA class I alleles is calculated using the relevant tool available at www.hiv.lanl.gov/content/immunology. The latter tool calculates a two-sided exact Fisher’s p-value and corrects the results for false discovery rate using Storey’s q-value calculation.

Wilcoxon rank-sum test

The Wilcoxon rank-sum test or MannWhitney U test is a non-parametric test for assessing whether two independent observation samples have similar distributions. The Wilcoxon rank-sum test is similar to the t-test when no underlying distribution is assumed for the data. Where applicable, independent p-values are combined using Fisher’s combined test. The statistical software R v2.9.2 was used for applying the test.

Explained Fraction

The Explained Fraction (EF) suggested in [190] can be used to estimate how much of heterogeneity in disease outcome is explained by known genetic or environmental factors, and hence how much is yet to be explained by unknown or non-included factors. In [190], the EF was used to estimate the amount of variation in progression to AIDS that can explained by 13 different genetic factors. The EF can provide insight into how much a disease status can be explained by a factor but also how much the absence of the factor explains the absence of disease [190]. This statistic is based on mutual information. Assuming only cases were both causal factors and
disease outcomes are described by categorical variables, we construct a contingency table of frequencies, denoted \([a_{ij}]\), with rows \(i\) indicating causal factors and columns \(j\) representing disease outcome. For multiple causal factors, its combination of factors is denoted by a row. The row marginals, \(a_i\), correspond to the frequency of the factor and the column marginals, \(a_j\) to the frequency of the disease. Now, the EF can be calculated using the values in the contingency table, the marginal values and Equations 2.3 and 2.4. In our analysis, we calculate both the combined EF that takes into account the combinations of causal factors and the additive EF which is the summation of the EFs of the individual factors. For independent factors, the combined EF is expected to be larger than the additive EF due to the convexity of the information measure [191].

The EF takes values between 0 and 100%. To estimate the 95% confidence intervals (CI) for the EF we used bootstrapping of the available data.

\[
EF = \frac{\sum_{i,j} a_{ij} \log\left(\frac{a_{ij}}{a_i a_j}\right)}{-I_{\text{max}}(1, 2)}
\]  

(2.3)

where the nominator is the mutual information \(I(1, 2)\) and the denominator, \(I_{\text{max}}(1, 2)\) is calculated by Equation 2.4 for two disease states and can be extended to include \(n\) disease states. \(I_{\text{max}}(1, 2)\) is the maximal value of \(I(1, 2)\) achieved when knowledge of the causal factors completely predicts an individual’s disease status.

\[
I_{\text{max}}(1, 2) = \sum_i a_{i1} \log\left(\frac{a_{i1}}{a_i a_1}\right) + \sum_i a_{i2} \log\left(\frac{a_{i2}}{a_i a_2}\right)
\]  

(2.4)

Multiple testing correction

In many cases we explore a lot of different hypotheses and multiple testing correction is required. However, a multiple testing correction approach might be too strict because a lot of the tests are not independent. In these cases, we only consider a result valid when it can be replicated in an independent cohort. Where applicable we also perform Monte Carlo simulations to estimate the False Discovery Rate (FDR) for our findings.
2.3.3 Epitope prediction

Software

Epitope prediction is a very useful tool in immunoinformatics. Prediction of T cell class I epitopes is now highly accurate (algorithms can achieve prediction accuracy (i.e. correct identification of true positives) of up to 94%) and has a wide allelic coverage [192]. This is particularly useful since many known HLA class I molecules do not have their binding specificity characterised. Some examples of epitope prediction software include Epipred [193], NetMHCPan [194], Stabilized Matrix Method (SMM) [195], Kernel-based Inter-allele peptide binding prediction SyStem (KISS) [12] and Adaptive Double Threading (ADT) [196]. These algorithms have been used in several studies in order to identify potential epitopes associated with differential outcome of viral infections and hence guide vaccine design [184, 197]. Many epitope prediction algorithms (including the ones used in this study) estimate the binding affinity of the pMHC complex (IC50). An affinity threshold of approximately 500nM (preferably 50nM or less) has been shown to determine the capacity of a peptide epitope to elicit a CD8+ T cell response [4]. It is worth noting that overall, prediction is based on the presence of specific motifs which can lead to a skewed sampling of peptides towards high-affinity binders and therefore potentially miss intermediate or weak binders.

We use the following epitope predictors:

1. **Metaserver v1.0** [198] is a combination of two web-based prediction methods, NetCTL v1.2 [14] and NetMHC v3.0 [199, 200]. NetCTL is an integrated method that predicts TAP transport, proteasomal cleavage and HLA binding. It has allelic predictors for 12 different class I alleles that are chosen to be representative of each of 12 supertypes. NetMHC v3.0 predicts pMHC binding, using artificial neural networks to predict binding affinities for 43 HLA molecules.
Metaserver combines the two methods (TAP transport and proteasomal cleavage from NetCTL and pMHC binding from NetMHC) and removes a normalising assumption (which held that all alleles bind the same number of peptides) to produce an estimate that shows improved accuracy in epitope prediction and predicts epitopes for 43 HLA molecules. Binding predictions for HLA-C molecules are not available in Metaserver due to scarcity of HLA-C binding data necessary for training the neural networks).

2. **NetMHCPan v2.2 and v2.4** [194] is a pan-predictor which uses an artificial neural network trained on the hitherto largest set of quantitative MHC binding data available, covering HLA-A, HLA-B, HLA-C as well as chimpanzee, rhesus macaque, gorilla, and mouse MHC class I molecules. Because it does not use a different artificial neural network for each allele NetMHCPan can be used (with caution) to predict peptide binding strength for any HLA allele. The model is built such that data from alleles that are similar to the allele under study are also taken into account.

We predict epitopes for both HCV and HTLV-1 strains. The specific protein sequences used for the epitope prediction can be found in Appendix A.

**Evaluation**

In order to investigate the accuracy of the epitope predictors we compared the predicted affinities with the experimentally quantified affinity for a set of 4 HLA molecules (A-02:01, A-24:02, B-07:02, B-35:01) for the HTLV-1 proteome (Table [21]). This ‘evaluation’ set of experimental values has not been used in the training of the predictors, i.e. overfitting is avoided, and therefore constitutes an independent measure of benchmarking the performance of the predictors. Although, many different predictors were investigated, Metaserver, NetMHCPan and SMM were overall outperforming the other predictors based on the evaluation dataset used. Here, we use Metaserver and
NetMHCPan for our analysis.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>A-02:01 (50)</th>
<th>A-24:02 (49)</th>
<th>B-07:02 (44)</th>
<th>B-35:01 (49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metaserver</td>
<td>0.76 (1.53 x 10^{-10})</td>
<td>0.68 (6.97 x 10^{-8})</td>
<td>0.67 (7.67 x 10^{-7})</td>
<td>0.66 (2.60 x 10^{-7})</td>
</tr>
<tr>
<td>NetMHCPan</td>
<td>-0.76 (1.23 x 10^{-10})</td>
<td>-0.72 (3.96 x 10^{-10})</td>
<td>-0.78 (3.94 x 10^{-10})</td>
<td>-0.77 (6.84 x 10^{-10})</td>
</tr>
<tr>
<td>SMM</td>
<td>-0.80 (2.02 x 10^{-12})</td>
<td>-0.62 (1.59 x 10^{-6})</td>
<td>-0.70 (1.40 x 10^{-7})</td>
<td>-0.78 (5.44 x 10^{-11})</td>
</tr>
<tr>
<td>Epipred-Web</td>
<td>0.48 (4.27 x 10^{-4})</td>
<td>0.68 (7.65 x 10^{-8})</td>
<td>0.65 (1.95 x 10^{-6})</td>
<td>0.47 (6.44 x 10^{-4})</td>
</tr>
<tr>
<td>Epipred-Local</td>
<td>0.56 (2.20 x 10^{-5})</td>
<td>0.70 (2.76 x 10^{-8})</td>
<td>0.60 (1.66 x 10^{-7})</td>
<td>0.32 (2.35 x 10^{-7})</td>
</tr>
<tr>
<td>Epipred-New</td>
<td>0.54 (5.37 x 10^{-5})</td>
<td>0.68 (6.69 x 10^{-8})</td>
<td>0.67 (6.94 x 10^{-7})</td>
<td>0.40 (4.18 x 10^{-7})</td>
</tr>
<tr>
<td>KISS</td>
<td>0.67 (1.36 x 10^{-7})</td>
<td>0.65 (1.98 x 10^{-6})</td>
<td>0.52 (3.34 x 10^{-4})</td>
<td>0.54 (8.30 x 10^{-3})</td>
</tr>
<tr>
<td>ADT</td>
<td>-0.68 (6.32 x 10^{-8})</td>
<td>-0.65 (3.66 x 10^{-7})</td>
<td>-0.51 (4.31 x 10^{-4})</td>
<td>-0.80 (6.29 x 10^{-12})</td>
</tr>
</tbody>
</table>

Table 2.1: Evaluation of epitope predictors. The measure used to evaluate the predictor’s is the non-parametric Spearman correlation. Depending on the binding score returned by the predictor (affinity is not always the default measure, other measures are used, such as binding energy) the correlation with experimental affinities might be positive (e.g. for affinity) or negative (e.g. for binding energy). The local version of the Epipred software was slightly different from the web version and was downloaded from the Epipred server. The new version of the software was provided after personal communication with the authors.

**Consistency**

The allelic coverage of Metaserver includes 43 HLA class I molecules (only HLA-A and B molecules are considered). In order to broaden the coverage and be able to study all the HLA alleles present in the HCV cohort we chose to use the combination of two epitope predictors: Metaserver and NetMHCPan. To verify that their predictions can be combined we explored the consistency of the predicted scores using 1) Spearman’s rank correlation and 2) a simple linear regression model that used the Metaserver predicted values as predictor and the NetMHCPan predicted values as response and vice versa. Since the Spearman correlation was very significant and the slope of the regression line was converging to the unit (Figure 2.1) we considered that the two
predictors can be combined.

**Rank measure**

Not all epitope prediction software provide the same measures of peptide binding strength. Some are only focused on binary classification (epitope or non-epitope) and some have a continuous measure that is usually given in units of binding affinity (IC50). The predictors used here return a continuous peptide binding score. In theory, the predicted binding score for each pMHC complex could allow the comparison of predicted binding affinities between alleles. However, between allele comparisons can be problematic. Firstly, within-allele comparisons (i.e. predictions for different peptides to the same allele) are thought to be more comparable than predictions between alleles. Secondly, whether or not a normalisation procedure should be applied for between-allele comparisons is still being debated in the community [198]. To avoid the potential problem of between-allele comparisons we used the rank measure technique introduced by [184]. For the rank measure, the strength of an alleles preference for a particular protein is quantified by ranking the strength of binding of the top binding peptide from the protein of interest amongst the strength of binding of peptides from the entire proteome to that allele. Specifically, we split the relevant proteome (HCV or HTLV-1) into overlapping nonamers offset by a single amino acid and predicted a binding affinity score for each nonamer. For each allele we rank all nonamers from the proteome from the weakest to strongest predicted binding scores. This produces a list of rank values for each protein to that particular allele that quantifies the binding relationship between that allele and the protein.
Figure 2.1: Comparison between the breadth calculated based on Metaserver and NetMHCPan respectively for the same alleles. Since Metaserver has a smaller coverage we were able to compare the breadth only for the 36 alleles covered by Metaserver. We examined the consistency of strong binders (<50nM), weak binders (>50nM and <500nM) and all binders (strong and weak) between the two predictors. The agreement is very good, as indicated by the Spearman correlation (r) and the slope of the regression line, and this allows us to combine the results of the two software with confidence in order to study the impact of allelic breadth on the outcome of the HCV infection. The line of equality (y = x) is also displayed to ease the comparison with the regression line.
2.4 HLA class I and HCV infection

In this part of the thesis, we tested a number of hypothesis to investigate the impact of HLA class I molecules on HCV infection outcome. The HLA class I related factors that we examine, have been significant in determining the outcome of other viral infections such as HIV and HTLV-1 but have not been explored extensively in the context of HCV infection. Interestingly, we find a weak overall impact of HLA class I-related factors on HCV outcome, suggesting a less efficacious role for HLA class I in HCV infection.

2.4.1 Heterozygotes’ advantage

First, we test the hypothesis that the patients who are heterozygotes in terms of their HLA class I alleles can present a broader spectrum of viral peptides to the immune system and obtain an advantage in clearing HCV infection. To investigate this hypothesis we examine the following possibilities:

1. Homozygotes at HLA-A locus have disadvantage over heterozygotes at HLA-A locus.

2. Homozygotes at HLA-B locus have disadvantage over heterozygotes at HLA-B locus.

3. Homozygotes at HLA-C locus have disadvantage over heterozygotes at HLA-C locus.

4. Homozygotes at exactly one locus have disadvantage over heterozygotes at all loci.

5. Homozygotes at exactly two loci have disadvantage over heterozygotes at all loci.
6. Homozygotes at **all three** loci have disadvantage over heterozygotes at all loci.

7. Homozygotes at **any one or more** loci have disadvantage over heterozygotes at all loci.

The analysis is performed on 4-digits (HLA-YXX:XX, Y=A,B,C), 2-digits (HLA-YXX) allele resolution and supertypes, i.e. the broader HLA grouping suggested in [201] (for HLA-C alleles there are no defined supertypes). HLA class I supertypes are a plausible classification scheme for alleles with similar peptide binding specificities [181].

It is necessary to perform all three resolutions because in the literature results are given for different resolutions and we need to obtain comparable results. Additionally, because of data aggregation for coarser resolution such as 2-digit and supertypes our analysis obtains more statistical power. Finally, for this part of the analysis we only consider individuals with full known HLA background (N=782).

The results for the 4- and 2- digits resolution are presented in Table 2.2 and for supertypes in Table 2.3 for each cohort, respectively. Although the direction of the Odds Ratio indicates a heterozygote advantage in almost all the hypothesis tested, the only statistically significant indication for a heterozygote advantage is observed for HLA-B alleles in the UK cohort (see Table 2.3). Interestingly, in [150] a dominant involvement of HLA-B in influencing HIV disease outcome is also reported. However, although the OR is in the same direction, the same result is not statistically significant in the larger USA cohort. As a consequence, the heterozygote advantage at the supertype level for the HLA-B locus is only a trend in the pooled cohort (see Table 2.3). The main explanation that we can provide for the lack of replication, other than the case of a false positive result, is the significantly different underlying allelic distribution between the two cohorts. To examine this hypothesis we compared the allelic distributions of the two cohorts for each locus (A, B and C) using a paired Wilcoxon sum-rank test. For a more detailed comparison we calculated the allele frequencies for heterozygotes and homozygotes in each case. We found that the distribution of the alleles between
the two cohorts when taking only the homozygous individuals into account are not significantly different, i.e. the alleles which are frequent among homozygotes have similar prevalences in the two cohorts. However, the distribution of the alleles at B locus for the individuals that are heterozygotes at the specific locus is significantly different between the two cohorts (Table 2.4). This implies that the HLA-B alleles that are driving the heterozygote advantage in the UK Cohort might not be as prevalent in the USA cohort.

Additionally, when we predict the breadth (i.e. number of binding peptides) of the HLA class I molecules for all individuals in the pooled cohort and categorise the individuals based on their zygosity we can readily observe that the HLA breadth distribution of heterozygotes is not significantly different from the one of homozygotes at any locus (Figure 2.2); nevertheless, consistently with a stronger heterozygote advantage in the UK cohort there was a trend for higher allelic breadth in heterozygotes in the UK cohort (p=0.09) but not in the USA cohort (p=0.8). When we pool the two cohorts we find that at the supertype level, there is a statistically significant advantage of heterozygotes (OR=0.7, p=0.03) over individuals that are homozygotes at one or two loci (Table 2.2(c)). In general, the overall lack of a strong heterozygote advantage in HCV is supported by findings in others studies [99,202,203]. To examine the validity of our approach we also applied this analysis in the context of HTLV-1 infection where there is a heterozygote advantage associated with proviral load [127]. We hypothesised that this can be partially explained by the breadth of HLA binding and we indeed found that a higher breadth of HLA-A alleles is significantly associated with lower log-proviral load in ACs and a higher breadth of HLA-B alleles is significantly associated with a lower proviral load in HAM/TSPs; overall, in the HTLV-1 cohort heterozygotes have a higher allelic breadth compared to homozygotes at one or two loci (Figure 2.3). This indicates that breadth can be linked to heterozygote advantage but it is not necessarily the only driving factor. We further examine the role of HLA class I allelic breadth in HCV infection in section 2.4.4.
Table 2.2: Impact of zygosity on HCV infection outcome for UK, USA and pooled cohort for 4- and 2- digits allelic resolution. OR<1 implies a heterozygote advantage. The results in the brackets are for 2-digits resolution.

<table>
<thead>
<tr>
<th>Homozygosity</th>
<th>Odds Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) UK cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-A alleles</td>
<td>0.86(0.86)</td>
<td>0.78(0.77)</td>
</tr>
<tr>
<td>HLA-B alleles</td>
<td>0.23(0.31)</td>
<td>0.09(0.09)</td>
</tr>
<tr>
<td>HLA-C alleles</td>
<td>0.54(0.58)</td>
<td>0.40(0.35)</td>
</tr>
<tr>
<td>One allele</td>
<td>1.69(1.23)</td>
<td>0.32(0.63)</td>
</tr>
<tr>
<td>Two alleles</td>
<td>0.23(0.23)</td>
<td>0.19(0.19)</td>
</tr>
<tr>
<td>Three alleles</td>
<td>0.34(0.34)</td>
<td>0.42(0.42)</td>
</tr>
<tr>
<td>One or more alleles</td>
<td>0.89(0.81)</td>
<td>0.80(0.61)</td>
</tr>
<tr>
<td>(b) USA cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-A alleles</td>
<td>0.55(0.83)</td>
<td>0.07(0.51)</td>
</tr>
<tr>
<td>HLA-B alleles</td>
<td>1.15(0.93)</td>
<td>0.75(0.85)</td>
</tr>
<tr>
<td>HLA-C alleles</td>
<td>0.87(0.82)</td>
<td>0.69(0.46)</td>
</tr>
<tr>
<td>One allele</td>
<td>0.70(0.90)</td>
<td>0.20(0.64)</td>
</tr>
<tr>
<td>Two alleles</td>
<td>0.88(0.87)</td>
<td>0.80(0.74)</td>
</tr>
<tr>
<td>Three alleles</td>
<td>0.80(0.47)</td>
<td>0.86(0.51)</td>
</tr>
<tr>
<td>One or more alleles</td>
<td>0.72(0.86)</td>
<td>0.18(0.46)</td>
</tr>
<tr>
<td>(c) Pooled cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-A alleles</td>
<td>0.68(0.90)</td>
<td>0.18(0.69)</td>
</tr>
<tr>
<td>HLA-B alleles</td>
<td>0.72(0.68)</td>
<td>0.38(0.23)</td>
</tr>
<tr>
<td>HLA-C alleles</td>
<td>0.78(0.75)</td>
<td>0.41(0.25)</td>
</tr>
<tr>
<td>One allele</td>
<td>0.86(0.97)</td>
<td>0.53(0.86)</td>
</tr>
<tr>
<td>Two alleles</td>
<td>0.64(0.68)</td>
<td>0.32(0.34)</td>
</tr>
<tr>
<td>Three alleles</td>
<td>0.47(0.38)</td>
<td>0.43(0.28)</td>
</tr>
<tr>
<td>One or more alleles</td>
<td>0.76(0.84)</td>
<td>0.21(0.36)</td>
</tr>
</tbody>
</table>
Table 2.3: Impact of supertype zygosity on HCV infection outcome for UK, USA and pooled cohort for HLA-A and B supertypes. OR<1 implies a heterozygote advantage. There are no defined supertypes for HLA-C molecules.

<table>
<thead>
<tr>
<th>Zygosity</th>
<th>Locus</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homozygotes</td>
<td>HLA-A</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>HLA-B</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>HLA-C</td>
<td>0.92</td>
</tr>
<tr>
<td>Heterozygotes</td>
<td>HLA-A</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>HLA-B</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>HLA-C</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Table 2.4: Comparison of the allelic distribution for each HLA class I locus between the UK and USA cohorts stratified for zygosity. The results are based on paired Wilcoxon sum-rank test.
Figure 2.2: Breadth of HLA class I molecules based on strong predicted binders versus zygosity, in the pooled CV cohort. The x-axis indicates the homozygosity status of the individuals included in each boxplot. A, B, C=Homozygotes only at HLA-A, B, C locus, respectively, One, Two, Three=Homozygotes only at one, two or three loci, Any=Homozygotes at any number of loci, None=Heterozygotes at all loci.
Figure 2.3: Breadth of HLA class I molecules based on strong predicted binders versus zygosiy, in the HTLV-1 cohort. The x-axis indicates the homozygosity status of the individuals included in each boxplot. A, B, C=Homozygotes only at HLA-A, B, C locus, respectively, One, Two, Three=Homozygotes only at one, two or three loci, Any=Homozygotes at any number of loci, None=Heterozygotes at all loci.
2.4.2 Rare allele advantage

The rare allele advantage hypothesis is based on the underlying assumption that the patients who are expressing rare HLA alleles can have an advantage because the virus adapts to the most commonly encountered HLA-restricted immune responses but not to the rare ones. The frequency of the supertypes is calculated for 1) the USA Cohort, based on the data from the Allele Frequency Net Database [204] while for 2) the UK Cohort, based on the present data. The supertype grouping defined in [201] is used.

We find a statistically significant rare allele advantage for the HLA-B locus at the supertype level (see Table 2.5) for the UK but not the USA cohort. However, because the direction of the OR is the same, when combining the cohorts the rare allele advantage is also present. The rare allele advantage hypothesis cannot be considered independently from the heterozygote advantage hypothesis. An individual with rare HLA alleles is more likely to be heterozygote for these alleles. Therefore, it might not surprising that the weak heterozygote advantage that we have already found is followed by the rare allele advantage. This is further supported by the fact that the rare allele advantage for the HLA-B supertypes is no longer significant if we correct for the heterozygote advantage. In summary, there is a weak indication of a rare allele advantage for HLA-B alleles at the supertype level but this cannot be decoupled from the heterozygote advantage studied above. However, the heterozygote advantage constitutes a broader hypothesis and it does not follow immediately from the rare allele advantage. Of note, we found no evidence suggesting a role for HLA-A alleles in HCV infection.

2.4.3 Individual alleles

We also investigated whether individual HLA class I alleles have an important impact on the outcome of HCV infection. Many HLA class I associations with HCV clearance and persistence have been found in different cohorts (Appendix Table A.1). However,
<table>
<thead>
<tr>
<th>Alleles</th>
<th>Odds Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A</td>
<td>2.26</td>
<td>0.84</td>
</tr>
<tr>
<td>HLA-B</td>
<td>0.004</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 2.5: Impact of HLA-A and B supertype frequency on HCV infection outcome for pooled cohort. OR<1 implies a rare allele advantage.

many of the associations have not been replicated and some others give contradictory results in different studies. Here, we use the UK and USA cohorts as well as known associations reported in the literature in order to obtain a consistent and accurate set of HLA class I alleles which can explain part of the heterogeneity of the HCV infection outcome. In the multiple logistic regression model used, all identified confounding variables are included as covariates (Appendix Table A.2). We perform the analysis for both 4-digits (data not shown) and 2-digits allelic resolution. The 2-digit resolution analysis allows for more statistical power. Our results for all the cohorts are shown in Figure 2.4. In summary, there are three HLA molecules (see Table 2.6) which have a statistically significant effect on HCV outcome: 1) B*57 and C*01 are associated with viral clearance (protective effect) and 2) C*04 is associated with viral persistence (detrimental effect). Since this is a retrospective study and the UK and USA cohorts have been studied before, these associations have been reported in the literature [24].

Another important factor that might confound these results is linkage disequilibrium (LD). Therefore, we investigated which HLA alleles are in LD with B*57, C*01 and C*04. These are:
1. **B*57**: A*01, C*06, C*18

2. **C*01**: B*27, B*56

3. **C*04**: A*36, B*07, B*35, B*53

Only C*06 which is in LD with B*57 should be further explored since C*06 is statistically significant associated with viral clearance in the pooled cohort while the other linked alleles do not have an effect on outcome. Two observations can rule out the possibility that C*06 is driving the protective effect of B*57: 1) When both molecules are included as covariates in the multiple regression model only the B*57 effect is still significant and 2) the B*57 effect is still significant among C*06-negative individuals.

### Table 2.6: The HLA alleles that are associated with the outcome of the HCV infection and are replicated in independent cohorts. P=Protective, D=Detrimental, OR=Odds Ratio, n_c=Number of allele carriers. The bold numbers in the parentheses refer to the 4-digit alleles considered in each case.

<table>
<thead>
<tr>
<th>Highly likely-Protective</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B*57</strong> (01,02,03)</td>
<td>P in Pooled Cohort (OR=1.99, p=0.005, n_c=86)</td>
</tr>
<tr>
<td></td>
<td>P in Ghana in [25]</td>
</tr>
<tr>
<td></td>
<td>P in USA in [98]</td>
</tr>
<tr>
<td><strong>C*01</strong> (02)</td>
<td>P in Pooled Cohort (OR=2.24, p=0.02, n_c=40)</td>
</tr>
<tr>
<td></td>
<td>P in Japan in [99]</td>
</tr>
<tr>
<td></td>
<td>P in Ireland in [26]</td>
</tr>
<tr>
<td></td>
<td>P in USA in [98]</td>
</tr>
<tr>
<td>Highly likely-Detrimental</td>
<td></td>
</tr>
<tr>
<td><strong>C*04</strong> (01)</td>
<td>D in Pooled Cohort (OR=0.65, p=0.03, n_c=187)</td>
</tr>
<tr>
<td></td>
<td>D in Ireland in [27]</td>
</tr>
</tbody>
</table>

Note: In two studies reported P [99, 205]
Figure 2.4: HLA class I associations with HCV infection outcome in UK, USA and pooled cohorts. Only alleles with 10 or more observations are shown. The OR is calculated based on a multiple logistic regression model which includes all potential confounding variables as covariates. The bars represent 95% confidence intervals (CI). The dashed vertical line indicates OR=1. Here, OR>1 indicates a protective effect (significant associations in green) while OR<1 indicates a detrimental effect (significant associations in orange). The asterisks (*) are dropped from the HLA names for a clear display.
2.4.4 HLA class I breadth

Using predicted binding peptides for each HLA class I molecule present in the UK and USA cohort we test the hypothesis that the HLA molecules which have a larger breadth i.e. can present a larger number of HCV peptides to the immune system (in particular to CD8+ T cells) and give an advantage to the host in clearing HCV infection. Based on the values of affinity measures used by the predictors we can divide the predicted epitopes in 1) strong (>50nM) and 2) weak binders (>50nM and <500nM). Hence, in the analysis we consider three categories of pMHC binding: a) strong, b) weak and c) strong and weak, i.e. all binding peptides. The total or per protein breadth (see definitions below) are considered for each HLA class I locus separately.

Total breadth For each HLA type (A, B and C) we count the distinct number of HCV epitopes (\(d_e\)) which we call total breadth of each HLA locus (A, B or C). The number of epitopes is defined based on the default threshold used by the epitope prediction software. For example, if an individual recognises \(x\) HCV epitopes based on his HLA-A1 molecule and \(y\) epitopes based on his HLA-A2 molecule and \(z\) of these epitopes bind both A1 and A2 then \(d_e = x + y - z\) distinct epitopes can bind in total the HLA-A molecules. We found that total breadth of HLA-A, B or C molecules did not have a statistically significant impact on the outcome of infection (Table 2.7). Interestingly, as already mentioned homozygosity was not associated with a significantly narrower predicted breadth (Figure 2.2).

Breadth per protein Next, we investigated whether the breadth of an HLA molecule for peptides of a specific HCV protein has an effect on the outcome of HCV infection. Although we found that the total breadth of HLA class I molecules does not have an impact on outcome this can be masking an effect which might be attributed to the breadth for peptides belonging to only one or few HCV proteins. We call this quantity breadth per protein. To calculate it, we count the distinct epitopes per HLA
Table 2.7: Overall breadth of HLA-A, B, C molecules in HCV infection. The variable expressing the total breadth of each HLA molecule is fitted simultaneously in the model for both UK and USA cohorts. All confounding variables are included as covariates. Here, OR>1 indicates a protective effect while OR<1 indicates a detrimental effect.
locus as in the case of overall breadth but we split the overall breadth in breadth per HCV protein. We did not find any statistically significant impact of the HLA breadth per HCV protein on the outcome of infection (Table 2.8). Hence, we can conclude that based on this approach, allelic breadth does not influence the outcome of HCV infection. This is consistent with the lack of a strong heterozygote advantage that we reported above (see section 2.4.1).

2.4.5 Protein specificity

Following the analysis performed in [184, 185] we hypothesised that HLA molecules which can present peptides from specific HCV proteins to CD8+ T cells, can be associated with viral clearance. Hence, we explored the protein specificities of the HLA molecules in the context of HCV infection. The binding strength measure which we used is rank (see Methods). First, we looked at the protein specificities of protective and detrimental HLA molecules (Table 2.6). The ranks of the strongest 5 binding peptides from each protein to the alleles B*57 and C*01 (10 rank values) were compared against the ranks of the strongest 5 binding peptides to the allele C*04 (5 rank values) (C*01:02, C*04:01 and B*57:01 are the most frequent alleles among the C*01, C*04 and B*57 allelic groups, respectively). A Wilcoxon-Mann-Whitney test was performed for each protein to test for differences between the two sets of rank values. We could not establish a statistically significant difference in preference of binding for HCV proteins between protective and detrimental HLA molecules (Figure 2.5). The only significant difference observed was for protein NS3 (protective alleles have a stronger binding preference for NS3 compared to the detrimental allele) but it was lost after FDR correction (q-value > 0.05). The results were robust when the number of peptides was changed to 8 or 10. Interestingly, although no significant difference was detected between protective and detrimental molecules, we did observe that all three HLA molecules had an overall stronger binding preference for proteins NS3 and NS5B which are shown to be
### (a) UK cohort

<table>
<thead>
<tr>
<th>Protein</th>
<th>HLA-A</th>
<th></th>
<th>p-value</th>
<th>HLA-B</th>
<th></th>
<th>p-value</th>
<th>HLA-C</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F6</td>
<td>1.02</td>
<td>0.87</td>
<td>0.99</td>
<td>0.97</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>core</td>
<td>1.036</td>
<td>0.73</td>
<td>0.99</td>
<td>0.84</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>1.18</td>
<td>0.20</td>
<td>0.95</td>
<td>0.92</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>1.19</td>
<td>0.09</td>
<td>0.93</td>
<td>0.79</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>1.53</td>
<td>0.02</td>
<td>0.43</td>
<td>1</td>
<td>1.11</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td>1.05</td>
<td>0.39</td>
<td>1.56</td>
<td>0.34</td>
<td>1.54</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td>1.06</td>
<td>0.44</td>
<td>0.95</td>
<td>0.83</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4A</td>
<td>1.08</td>
<td>0.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4B</td>
<td>1.01</td>
<td>0.88</td>
<td>0.57</td>
<td>0.24</td>
<td>1.78</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5A</td>
<td>1.07</td>
<td>0.35</td>
<td>0.97</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5B</td>
<td>1.03</td>
<td>0.50</td>
<td>0.92</td>
<td>0.70</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Strong**

**Weak**

<table>
<thead>
<tr>
<th>Protein</th>
<th>HLA-A</th>
<th></th>
<th>p-value</th>
<th>HLA-B</th>
<th></th>
<th>p-value</th>
<th>HLA-C</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.99</td>
<td>0.78</td>
<td>0.81</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>core</td>
<td>0.99</td>
<td>0.82</td>
<td>0.91</td>
<td>0.18</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>1.02</td>
<td>0.77</td>
<td>1.03</td>
<td>0.63</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>0.98</td>
<td>0.63</td>
<td>0.97</td>
<td>0.37</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>0.97</td>
<td>0.71</td>
<td>0.93</td>
<td>0.42</td>
<td>0.75</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td>1.01</td>
<td>0.67</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td>1.02</td>
<td>0.56</td>
<td>0.99</td>
<td>0.57</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4A</td>
<td>1.10</td>
<td>0.57</td>
<td>1.46</td>
<td>0.32</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4B</td>
<td>1.04</td>
<td>0.29</td>
<td>0.98</td>
<td>0.73</td>
<td>0.99</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5A</td>
<td>1.02</td>
<td>0.63</td>
<td>0.97</td>
<td>0.41</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5B</td>
<td>0.98</td>
<td>0.61</td>
<td>0.99</td>
<td>0.59</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (b) USA cohort

<table>
<thead>
<tr>
<th>Protein</th>
<th>HLA-A</th>
<th></th>
<th>p-value</th>
<th>HLA-B</th>
<th></th>
<th>p-value</th>
<th>HLA-C</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.99</td>
<td>0.68</td>
<td>0.98</td>
<td>0.26</td>
<td>0.91</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>core</td>
<td>1</td>
<td>1.00</td>
<td>0.97</td>
<td>0.41</td>
<td>0.93</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>1.01</td>
<td>0.86</td>
<td>1</td>
<td>0.98</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>0.99</td>
<td>0.56</td>
<td>1.00</td>
<td>0.79</td>
<td>0.96</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>0.93</td>
<td>0.18</td>
<td>1.03</td>
<td>0.40</td>
<td>0.87</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td>1.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.54</td>
<td>1.09</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td>1.01</td>
<td>0.41</td>
<td>0.99</td>
<td>0.36</td>
<td>0.98</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4A</td>
<td>1.01</td>
<td>0.82</td>
<td>0.97</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS4B</td>
<td>1.00</td>
<td>0.89</td>
<td>0.99</td>
<td>0.49</td>
<td>0.88</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5A</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>0.37</td>
<td>0.92</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5B</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
<td>0.88</td>
<td>0.96</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.8: Breadth per protein of HLA-A, B, C molecules in HCV infection. The variable expressing the breadth per protein of each HLA molecule was fitted in the same model for UK and USA cohorts. When no predicted epitopes were identified, a model could not be fitted (dash). All confounding variables are included as covariates. Here, OR>1 indicates a protective effect while OR<1 indicates a detrimental effect.
more immunogenic compared to the other HCV proteins [115, 206, 207] (Figure 2.5) while there was no binding preference for NS4A.

Figure 2.5: Binding preference of HCV proteins for protective and detrimental HLA class I molecules. Protective molecules: B*57:01 and C*01:02 and detrimental molecules: C*04:01 (Table 2.6). To establish the binding preference of the HLA molecules we used the top 5 ranking peptides per molecule. The results are unchanged if other thresholds are used instead. The p-values were corrected for multiple comparisons using the Benjamin-Hochberg method (q-values).

Secondly, we explored whether the protein specificity was different between resolved and chronic individuals. We found similar results when we looked at two different measures: 1) overall specificity and 2) specificity per HLA locus.

For overall protein specificity, for each HLA molecule (A1, A2, B1, B2, C1, C2), we took only the top ranking epitope out of the whole HCV proteome and recorded
the protein to which it belonged. Then we counted how many out of the 6 HLA molecules of each individual preferred binding to each HCV protein. For example, if all the HLA molecules of an individual preferred binding to epitopes from the same HCV protein then this protein has value 6 while all the others have value 0. This gave us a specificity measurement to which we fitted a multiple logistic regression model for each HCV protein including all the identified confounding factors (see Methods). No statistical significant impact of protein specificity on the outcome of infection was identified (data not shown). We also looked at more than just the top ranking epitope and followed the same process in order to avoid being biased by the immunodominant proteins. The results were not altered. Additionally, for each HCV protein we set the top 40% of the HLA molecules which bind to these protein to be strong binders for it. Then we counted how many of the 6 alleles (A1, A2, B1, B2, C1, C2) that an individual has, binds to each HCV protein. We also performed our analysis for different thresholds for defining the strong binding alleles for each protein (10-40%). The results were the same as in the previous approach.

To examine protein specificity per HLA type, for each HLA molecule, we took only the top ranking epitope for each HCV protein. Then we calculated the mean rank \( m_r \) of each HLA molecule per protein. For example, if the top ranking epitope of an HCV protein for HLA-A1 has rank \( x \) and respectively for HLA-A2 has rank \( y \) then the binding preference of the individual’s HLA-A molecules for that HCV protein can be represented by \( m_r = \frac{x+y}{2} \). No statistical significant impact of protein specificity on the outcome of infection was observed (data not shown). For completeness, for each HLA molecule, we also calculated the mean of the top 3 ranks for each HCV protein (less biased towards the top rank) and calculated the mean rank of this quantity per HLA type but the results were not altered.

In summary, we investigated 1) whether protective and detrimental HLA class I alleles have different protein specificity in HCV infection and 2) if the protein specificity differentiates between resolved and chronic individuals but no statistical significant
differences in protein specificity were identified.

2.4.6 Epitope specificity

The last hypothesis that we tested is whether the HLA molecules which can present specific peptides (epitope specificity) to CD8+ T cells give an advantage to the host in clearing HCV infection. For each individual, we looked at the top 10 binding peptides for each HLA molecule (maximum 60 peptides if there are no overlapping ones). We considered this list to be a list of epitopes for each individual. Then, for each peptide in turn (N=3084) we attributed a score of 1 if an individual had this peptide in the list and 0 otherwise. This provided a binary variable for each peptide which we fitted in a logistic regression model to predict disease outcome. All the confounding variables were included as covariates. This approach, because of the very large number of peptides examined, requires a correction for multiple testing. Hence, applying the Bonferroni correction would require a statistical significance level of $p < 0.00005$. We found no such statistical significance for any of the peptides neither in the UK nor in the USA cohorts. The peptides with statistical significance $p < 0.05$ are given in Tables 2.9. Importantly, none of these peptides were included in the prediction of both the cohorts so no result could be replicated. The same results were reached when we created the list of epitopes for each individual considering only the strong binding or the strong and weak binding peptides. Therefore, because of lack of high statistical significance and lack of replication our findings suggest that epitope specificity cannot explain the heterogeneity in the outcome of HCV infection and the identified peptides with $p < 0.05$ are most likely spurious associations.
### (a) UK cohort

<table>
<thead>
<tr>
<th>Peptide</th>
<th>OR</th>
<th>p-value</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYRSSAPLL</td>
<td>2.66</td>
<td>0.04</td>
<td>F</td>
</tr>
<tr>
<td>MNWSPTAAL</td>
<td>5.06</td>
<td>0.007</td>
<td>E1</td>
</tr>
<tr>
<td>RMYVGGVEH</td>
<td>2.70</td>
<td>0.04</td>
<td>E2</td>
</tr>
<tr>
<td>NIVDVQYLY</td>
<td>0.10</td>
<td>0.03</td>
<td>E2</td>
</tr>
<tr>
<td>FYGMWPLL</td>
<td>2.66</td>
<td>0.04</td>
<td>P7</td>
</tr>
<tr>
<td>MWPLLLL</td>
<td>2.66</td>
<td>0.04</td>
<td>P7</td>
</tr>
<tr>
<td>LLALFQRAY</td>
<td>2.70</td>
<td>0.04</td>
<td>P7</td>
</tr>
<tr>
<td>MAIKLGAL</td>
<td>1.94</td>
<td>0.03</td>
<td>NS2</td>
</tr>
<tr>
<td>RQAEVITPA</td>
<td>2.70</td>
<td>0.04</td>
<td>NS4B</td>
</tr>
<tr>
<td>NMWSGTFPI</td>
<td>1.81</td>
<td>0.03</td>
<td>NS5A</td>
</tr>
<tr>
<td>LLAAGVGIY</td>
<td>2.70</td>
<td>0.04</td>
<td>NS5B</td>
</tr>
</tbody>
</table>

### (b) USA cohort

<table>
<thead>
<tr>
<th>Peptide</th>
<th>OR</th>
<th>p-value</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSGAPTYSW</td>
<td>1.74</td>
<td>0.02</td>
<td>E2</td>
</tr>
<tr>
<td>LPALSTGLI</td>
<td>0.73</td>
<td>0.006</td>
<td>E2</td>
</tr>
<tr>
<td>FSILDPTFTI</td>
<td>1.39</td>
<td>0.02</td>
<td>NS3</td>
</tr>
<tr>
<td>RAYMNTPG</td>
<td>2.67</td>
<td>0.01</td>
<td>NS3</td>
</tr>
<tr>
<td>SWLGNIMF</td>
<td>0.69</td>
<td>0.01</td>
<td>NS5B</td>
</tr>
</tbody>
</table>

**Table 2.9:** No predicted epitopes were consistently associated with the outcome of HCV infection. These predicted epitopes are associated with the outcome of HCV infection at \( p < 0.05 \) statistical significance but not corrected for multiple testing (\( p < 0.00005 \)) for both UK and USA cohorts. None of these peptides were observed in both cohorts and are most likely spurious associations. All confounding variables are included as covariates. Here, \( OR > 1 \) indicates a protective effect while \( OR < 1 \) indicates a detrimental effect.
2.5 Quantification of HLA class I impact on viral infection

Having examined the impact of six different HLA class I-related factors on the outcome of HCV viral infection (see section 2.3), we find that apart from a weak heterozygote advantage and a significant effect of three individual HLA molecules, the other HLA class I-mediated factors do not have a substantial influence on hepatitis C infection outcome. Therefore, in this part of the project we aim at quantifying the impact of HLA class I molecules on HCV viral clearance and persistence. Since HLA class I associations with infection outcome are reported in other viral infections, we estimate the impact of HLA class I alleles in three different viruses: 1) HCV, 2) HTLV-1 and 3) HIV in order to have a more complete assessment. The quantification is based on the Explained Fraction (EF) which estimates how much of the variation in disease outcome is explained by a given factor (see Methods). We calculate both the combined EF and the additive EF. Additionally, because the EF depends on the frequency of the disease outcome and the disease factor (Appendix Figure A.2) we also normalise it; we set the frequency of disease outcome to known population frequencies (HCV: Resolved=30% and Persistent=70% [110]; HTLV-1:ACs=99% and HAM/TSPs=1% [127]; HIV: Elite controllers=0.3%, Vireamic Controllers=10% and Chronic progressors=89.7% [208]) and the frequency of the disease factor to 50%. The normalisation of the EF allows us to compare between alleles. The HCV and HTLV-1 cohorts are described here (see Methods) while the HIV cohorts are reported in [155] and [156].

The proportion of variation in disease outcome explained overall by the known HLA class I associations is nearly 3% and 6.6% in HCV and HTLV-1 (Tables 2.10 and 2.11), respectively. Additionally, the amount of variation in HCV outcome explained by the combination of all the protective factors available in the pooled HCV cohort (B*57, C*01, the SNP rs12979860, the KIR2DL3/2DL3+HLA-C1C1 and the HBV
infection) is 9.5%. Over all the infections studied here, we find that individual HLA class I molecules can explain 1.5% (median) of the heterogeneity observed in the disease outcome and no more than 6.5% (using the EF normalised for population frequency). Interestingly, the EF varies substantially among the viral infections studied. In summary, the impact of the individual HLA class I alleles is greater in HIV infection, less pronounced in HTLV-1 infection and very limited in HCV infection (Table 2.12 and Figure 2.4).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Comb EF(%)</th>
<th>LCI(%)</th>
<th>UCI(%)</th>
<th>Add EF(%)</th>
<th>Pop Freq</th>
<th>Freq=0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*57</td>
<td>0.74</td>
<td>0.05</td>
<td>2.25</td>
<td>-</td>
<td>0.73</td>
<td>1.79</td>
</tr>
<tr>
<td>C*01</td>
<td>0.66</td>
<td>0.03</td>
<td>2.12</td>
<td>-</td>
<td>0.66</td>
<td>3.23</td>
</tr>
<tr>
<td>C*04</td>
<td>0.98</td>
<td>0.15</td>
<td>2.53</td>
<td>-</td>
<td>0.96</td>
<td>1.43</td>
</tr>
<tr>
<td>B<em>57C</em>01</td>
<td>1.57</td>
<td>0.49</td>
<td>3.7</td>
<td>1.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B<em>57C</em>04</td>
<td>1.93</td>
<td>0.8</td>
<td>4.27</td>
<td>1.72</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C<em>01C</em>04</td>
<td>1.85</td>
<td>0.68</td>
<td>4.02</td>
<td>1.64</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B<em>57C</em>01C*04</td>
<td>2.92</td>
<td>1.57</td>
<td>5.71</td>
<td>2.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Homoz 1 or 2</td>
<td>0.54</td>
<td>0.02</td>
<td>1.85</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs12979860</td>
<td>3.56</td>
<td>1.57</td>
<td>6.27</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HBV</td>
<td>1.68</td>
<td>0.42</td>
<td>3.64</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HIV</td>
<td>0</td>
<td>0</td>
<td>0.52</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2DL3/2DL3+C1C1</td>
<td>0.95</td>
<td>0.11</td>
<td>2.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Protective</td>
<td>9.52</td>
<td>7.64</td>
<td>14.8</td>
<td>7.59</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>All</td>
<td>13.57</td>
<td>12.83</td>
<td>21.35</td>
<td>9.11</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.10: The proportion of heterogeneity in outcome of HCV infection which is explained by HLA class I alleles and other known factors. In the protective factors we include: B*57, C*01, the SNP rs12979860, the KIR2DL3/2DL3+HLA-C1C1 and the HBV infection. The confidence intervals are obtained using the bootstrap method. Abbreviations, Homoz: Homozygosity, LCI: Lower 95% Confidence Interval, UCI: Upper 95% Confidence Interval, Comb EF: Combined EF, Add EF: Additive EF, Pop Freq: EF normalised for disease outcome at population frequency, Freq=0.5: EF normalised for 50% factor frequency.
<table>
<thead>
<tr>
<th>Factors</th>
<th>Comb EF(%)</th>
<th>LCI(%)</th>
<th>UCI(%)</th>
<th>Add EF(%)</th>
<th>Pop Freq</th>
<th>Freq=0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*02</td>
<td>3.13</td>
<td>0.98</td>
<td>6.57</td>
<td>-</td>
<td>1.61</td>
<td>3.27</td>
</tr>
<tr>
<td>B*54</td>
<td>2.55</td>
<td>0.54</td>
<td>5.26</td>
<td>-</td>
<td>1.51</td>
<td>4.02</td>
</tr>
<tr>
<td>C*08</td>
<td>1.61</td>
<td>0.23</td>
<td>4.26</td>
<td>-</td>
<td>0.78</td>
<td>3.09</td>
</tr>
<tr>
<td>A<em>02B</em>54</td>
<td>5.06</td>
<td>2.43</td>
<td>9.6</td>
<td>5.43</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A<em>02C</em>08</td>
<td>4.36</td>
<td>1.96</td>
<td>8.58</td>
<td>4.74</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B<em>54C</em>08</td>
<td>3.67</td>
<td>1.56</td>
<td>7.58</td>
<td>3.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A<em>02B</em>54C*08</td>
<td>6.6</td>
<td>4.08</td>
<td>12.28</td>
<td>7.04</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.11: The proportion of heterogeneity in outcome of HTLV-1 infection which is explained by HLA class I alleles. The confidence intervals are obtained using the bootstrap method. Abbreviations, LCI: Lower 95% Confidence Interval, UCI: Upper 95% Confidence Interval, Comb EF: Combined EF, Add EF: Additive EF, Pop Freq: EF normalised for disease outcome at population frequency, Freq=0.5: EF normalised for 50% factor frequency.

<table>
<thead>
<tr>
<th>Factors</th>
<th>EF(%)</th>
<th>LCI(%)</th>
<th>UCI(%)</th>
<th>Add EF(%)</th>
<th>Pop Freq</th>
<th>Freq=0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*57-Emu et al.</td>
<td>5.40</td>
<td>1.86</td>
<td>11.33</td>
<td>-</td>
<td>3.33</td>
<td>7.69</td>
</tr>
<tr>
<td>B*57-Pereyra et al.</td>
<td>4.56</td>
<td>1.68</td>
<td>9.75</td>
<td>-</td>
<td>6.45</td>
<td>5.69</td>
</tr>
<tr>
<td>B*27-Emu et al.</td>
<td>0.14</td>
<td>0.02</td>
<td>2.37</td>
<td>-</td>
<td>0.02</td>
<td>0.38</td>
</tr>
<tr>
<td>B*27-Pereyra et al.</td>
<td>1.69</td>
<td>0.31</td>
<td>5.14</td>
<td>-</td>
<td>4.97</td>
<td>3.96</td>
</tr>
<tr>
<td>B*35-Pereyra et al.</td>
<td>1.87</td>
<td>0.49</td>
<td>5.13</td>
<td>-</td>
<td>2.60</td>
<td>5.14</td>
</tr>
</tbody>
</table>

Table 2.12: The proportion of heterogeneity in outcome of HIV-1 infection which is explained by specific HLA class I alleles. The confidence intervals are obtained using the bootstrap method. The data for the HIV calculations are obtained from [155] and [156]. Abbreviations, LCI: Lower 95% Confidence Interval, UCI: Upper 95% Confidence Interval, Comb EF: Combined EF, Add EF: Additive EF, Pop Freq: EF normalised for disease outcome at population frequency, Freq=0.5: EF normalised for 50% factor frequency.
Figure 2.6: The proportion of variation in viral infection outcome explained by known HLA class I associations (the non-normalise EF is used here). The data for the HIV calculations are obtained from [155] and [156]. 95% confidence intervals were estimated by bootstrapping the data 5,000 times, trimming the 5% extremes, and then calculating the range in which 95% of the remaining data lay. Due to linkage between the HLA alleles, the EF is not additive; so for instance, in HTLV-1 infection, the three alleles HLA-A*02, C*08, and B*54 together only explain 6.6% of the outcome. The outcomes explained are HCV: spontaneous clearance v persistence; HTLV-1: asymptomatic carriage v HAM/TSP; HIV-1: elite control v viraemic control v progression. The asterisks (*) are dropped from the HLA names for a clear display.
2.6 Discussion

HLA class I molecules are responsible for presenting viral peptides to CD8+ T cells and also for modulating NK cell activation. They are therefore considered a very important factor in determining the response of the immune system to viral infection. The fact that HLA class I genes are highly polymorphic suggests that they are maintained through selective forces, such as infectious disease morbidity [209].

In this study, we investigated six HLA class I-related factors that can potentially shape a protective or detrimental CD8+ T cell response in the context of hepatitis C infection. Furthermore, we quantified the impact of individual HLA class I alleles on HCV, HTLV-1 and HIV infection.

In HCV infection for HLA class I supertypes, we found a weak heterozygote advantage for HLA-B alleles as well as a weak overall protective effect for heterozygote individuals over individuals who are homozygotes at one or two loci. A rare allele advantage for HLA-B supertypes was also observed. However, an individual with rare HLA alleles is more likely to be heterozygote for these alleles. Therefore, it is not surprising that the weak heterozygote advantage that we found is followed by the rare allele advantage. Nevertheless, the heterozygote advantage constitutes a broader hypothesis and it does not follow immediately from the rare allele advantage. The lack of a strong heterozygote advantage is consistent with the fact that the predicted HLA allelic breadth distribution of heterozygotes is not significantly different from that of the homozygotes and is also further supported by findings in others studies [99, 202, 203]. However, it remains interesting that in HCV infection, HLA-B, as in the case of HIV-1 [150], is the locus that presents the strongest protective effect suggesting a special role for HLA-B alleles in controlling viral infections.

We also re-identified three HLA class I alleles which are associated with disease outcome: 1) the protective molecules: $B^*57$ and $C^*01$ and 2) the detrimental molecule:
All associations were corrected here for known confounding factors such as linkage disequilibrium and mode of infection. These HLA class I associations have been reported before and confirmed in multiple independent cohorts [25, 27, 98, 99]. However, we should note here that the evidence for C*04 -a detrimental allele in the context of HCV infection- remain controversial. C*04 has also been found to convey protection from HCV infection [99, 205] and no C*04 effect was identified in a large HCV-infected cohort which was recently analysed in [98].

Interestingly, the B*57 protective effect has also been reported in HIV infection. In [210, 211] a plausible explanation of this effect is the broad cross-reactivity of B*57 against variants of a dominant Gag-epitope as well as its plasticity in peptide binding which was also shown for B*27 in [212]. In the latter study, the ability of both bulk CD8+ T cells as well as epitope-specific TCR clonotypes to inhibit viral replication in vitro and upregulate perforin and granzyme B were also linked to the protective effect. Our analysis cannot rule out or verify that B*57 acts in a similar manner in HCV infection. Furthermore, 1) the frequency of B*57 in the studied cohorts would not allow for enough statistical power so that B*57 could drive a significant difference in epitope specificity between resolved and persistent individuals and 2) if the effect is moderate, the lack of a strong T-cell related detrimental HLA class I molecule hinders the possibility of identifying a difference in protein or epitope specificity. However, the overall lack of breadth and protein or epitope specificity impact on disease outcome suggests that an alternative mechanism to cross-reactivity might explain the protective effect of B*57 in HCV infection; such us driving the selection of costly viral escape mutations [213].

Next, we used epitope prediction software to examine whether there are differences between resolved and persistent individuals in the 1) number of peptides that their HLA class I molecules are predicted to bind, 2) the HCV proteins that their HLA class I molecules prefer to bind or 3) the specific predicted epitopes that their HLA class I molecules prefer to recognise. We found that none of these factors were significant.
determinants of the outcome. In particular, for protein specificity we found no significant differences in binding preference, either between protective and detrimental HLA class I alleles, or between persistent and resolved individuals. All this evidence, taken together suggests that the impact of HLA class I related factors, as described here, on the outcome of HCV infection is rather limited. The lack of a significant difference of breadth, protein specificity and epitope specificity is perhaps not surprising. The $C^*01$-protective effect in HCV infection has been shown to be NK-mediated [83].

The only detrimental allele in our study is an HLA-C allele, $C^*04$ for which epitope prediction performs poorly (similarly for $C^*01$) compared to HLA-A and HLA-B and which can also be NK-mediated. If the effect is NK-mediated, this approach would not be expected to relate pMHC properties to disease outcome because of 1) the different way that HLA class I molecules tune NK cell responses compared to CD8+ T cell responses and 2) the less pronounced effect of peptides in NK cell inhibition or activation by HLA class I molecules.

Motivated by the rather small role of HLA class I-mediated factors in HCV outcome, we quantified the impact of HLA class I alleles in HCV infection and compared it with known associations reported for HTLV-1 and HIV infection. We calculated that the proportion of heterogeneity in the outcome of these viral infections that can be explained by individual HLA alleles is on average 1.5% (median), with an observed maximum of 6.5%. Interestingly, that proportion varies substantially among the viral infections studied. The impact of the individual HLA class I alleles, in terms of the normalised explained fraction, is found to be greater in HIV infection, less pronounced in HTLV-1 infection and very limited in HCV infection. This is in line but not necessarily linked with functional and sequence-based studies on CD8+ T cell responses which suggest that a strong and effective CD8+ T cell response is detected in HIV [214,215] and HTLV-1 [216] but not consistently in HCV infection [217]. Notably, HLA class I associations with disease outcome inform us about differences in the effect of the HLA molecules and a small explained fraction does not imply that CD8+ T
cell responses are dismal; it could be that all CD8+ T cell responses are rather similar regardless of their restriction so no significantly different ‘best’ molecules exist. On the other hand, we should not assume that the CD8+ T cell response is highly important just because HLA class I associations are highly statistically significant. Furthermore, the small explained fraction implies that HLA class I genotype is unlikely to be the major determinant of between-individual variation in clinical outcome. Nevertheless, the normalised comparison of the ‘most protective’ and ‘most detrimental’ molecules in each viral infection, as performed here, can provide a potential ranking of HLA class I-mediated impact between the different infections.
Chapter 3

Impact of KIRs on HLA class I-mediated immunity


3.1 Aim

Our aim was to investigate known HLA class I associations with disease outcome in different KIR genetic backgrounds for both HCV and HTLV-1 infection.

3.2 Introduction

Killer cell immunoglobulin-like receptors (KIRs) are a family of transmembrane proteins that are expressed on natural killer (NK) cells and subsets of T cells [72, 74]. They bind HLA class I molecules and have activatory and inhibitory isoforms [75]. KIRs contribute directly and indirectly to antiviral immunity. Directly, KIRs on NK cells sense the loss of HLA class I molecules from the cell surface and trigger NK-mediated cytolysis. Indirectly, NK cells regulate adaptive immunity via crosstalk with dendritic cells and by the production of chemokines and cytokines [81, 82, 218].

HLA class I molecules can be grouped into allotypes with similar KIR binding properties [219]. For example, KIR2DL2 binds group C1 HLA-C molecules which
have asparagine at residue 80, and, with a weaker affinity, group C2 molecules which have a lysine at position 80 [76].

Early research on KIRs investigated NK-mediated protection by studying disease associations with KIRs in the context of their HLA class I ligands [83,86]. There is now compelling evidence that KIRs also regulate adaptive immunity [81,82,218], but it is not known whether this has a significant impact on the response to infection in vivo. Differences between human KIRs and their mouse functional homologues (the Ly49 receptors) and the paucity of KIR allele-specific antibodies have hindered work on the role of KIRs in controlling adaptive immune responses. Here we used immunogenetics to investigate whether KIR genotype modulates HLA-mediated anti-viral protection in vivo. We focused on HLA class I alleles which have previously been associated with disease outcome and investigated whether these effects were altered by the KIR background. We studied 4 well-documented HLA class I allele-disease associations in two viral infections: human T lymphotropic virus type 1 (HTLV-1) and hepatitis C virus (HCV).

HTLV-1 is a persistent retrovirus that infects 10-20 million people worldwide. Most infected individuals remain lifelong asymptomatic carriers (ACs). However, approximately 10% of infected individuals develop associated diseases including HTLV-1-associated myelopathy/ tropical spastic paraparesis (HAM/TSP), an inflammatory disease of the central nervous system that results in progressive paralysis. It is poorly understood why some individuals remain asymptomatic whereas others develop disease, but one strong correlate of disease is the proviral load, which is significantly higher in HAM/TSP patients than in ACs [121]. It has previously been shown that HLA-A*02 and C*08 are each associated with both a reduced risk of HAM/TSP and a reduced proviral load in ACs and that HLA-B*54 is associated with an increased prevalence of HAM/TSP and an increased proviral load in HAM/TSP patients [23,127].

HCV is among the most widespread viral infections, with 170 million infected people worldwide. As in HTLV-1 infection, the outcome of HCV infection is hetero-
geneous: the virus persists in approximately 70% of infected individuals while the
remainders clear the infection spontaneously, usually within 6 months. Chronic HCV
infection can cause serious liver damage including cirrhosis and hepatocellular carci-
noma [91]. The origins of this heterogeneity are not completely understood but several
genetic determinants have been identified, including HLA-B*57 which is associated
with spontaneous clearance in several cohorts [24, 25, 98, 178].

The aim of this study was to test the hypothesis that KIR genotype determines
the efficiency of HLA class I-mediated anti-viral immunity. We tested this hypothesis
for 4 HLA class I associations: HLA-C*08, A*02 and B*54 in HTLV-1 infection and
B*57 in HCV infection. We show, using multiple independent measures, that for both
HCV and HTLV-1, possession of the KIR2DL2 gene enhanced HLA class I-restricted
immunity.

3.3 Methods

Many parts of the methodology applied in this Chapter are very similar to the one
presented in Methods of Chapter 2, so we only describe below the parts of the analysis
that have not already been presented.

3.3.1 Data

The HCV and HTLV-1 cohorts and the HLA genotyping are described in the Methods
of Chapter 2 and were collected by our collaborators.

KIR genotyping For the HCV cohort, the presence or absence of ten KIR genes
(KIR2DL1, KIR2DL2, KIR2DL3, KIR3DL1, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4,
KIR2DS5 and KIR3DS1) was determined by PCR using sequence specific primers
as described in [85]. Additionally, KIR2DL4, KIR3DL2, KIR3DL3, KIR2DL5 and
KIR2DP1 were typed in the USA samples. For the HTLV-1 cohort, the presence or
absence of ten KIR genes (*KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DL4*, *KIR3DL1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5* and *KIR3DS1*) was also determined by PCR using sequence specific primers.

**Viral load** HCV RNA was assessed by a branched DNA (bDNA) assay (Quantiplex HCV RNA 2.0 assay; Chiron Corporation) [ALIVE] or the HCV COBAS AMPLICOR system (COBAS AMPLICOR HCV; Roche Diagnostics) [MHCS]. The HTLV-1 provirus load in peripheral blood mononuclear cells (PBMC) was measured as described in [121]. Quantitative PCR was performed using an ABI 7700 sequence detector (Perkin-Elmer Applied Biosystems). All DNA standards and samples were amplified in triplicate. A standard curve was generated by using the β-actin gene from HTLV1 -negative PBMC and the Tax gene from TARL-2, a cell line containing a single copy of HTLV-1 proviral DNA.

3.3.2 **Statistical analysis**

**Overview**

Two independent cohorts were studied: HCV-infected individuals (N=782) and HTLV-1-infected individuals (N=402).

Ten KIR genes were typed; of these, 4 were present at an informative frequency (see definition below) in the HTLV-1 cohort and 9 in the HCV cohort. The modulation of HLA class I impact on infection outcome by the informative KIRs was analysed; the models included all other known determinants of outcome in the cohorts as covariates.

All HLA class I alleles which were significantly associated with outcome in our cohorts and that were independently verified (either in a large independent cohort or on an independent outcome) were studied. For HTLV-1 there are three such alleles: *HLA-A*<sup>*02*</sup> and *C*<sup>*08*</sup> which are associated with reduced proviral load in ACs and reduced risk of HAM/TSP and *HLA-B*<sup>*54*</sup> which is associated with increased proviral...
load in HAM/TSP patients and increased risk of HAM/TSP [23,27]. For HCV there are two such associations: *HLA-B*57 and *C*01 which have both been associated with increased odds of viral clearance in two independent cohorts (see Chapter 1 and [98]). The *C*01-protective effect in HCV infection has been shown to be NK-mediated [83,220] and therefore we did not consider it here.

We investigated all KIR genes which were present at an informative frequency. We define an informative frequency as sufficient to detect a ‘moderate’ effect size (Δ=0.5); this is equivalent to at least 32 people carrying the gene and 32 not carrying the gene [221]. Of the KIR genes typed, 4 were present at an informative frequency in the HTLV-I cohort (*KIR2DL2*, *KIR2DS2*, *KIR2DS3* and *KIR3DS1*) and 9 in the HCV cohort (*KIR2DL2*, *KIR2DL3*, *KIR3DL1*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5* and *KIR3DS1*).

The effect of individual HLA class I alleles in different KIR genetic background was investigated by stratifying the cohorts for the absence and presence of the informative KIRs and the effect of the already described HLA class I associations were re-evaluated in each stratum. The impact of the HLA alleles on two response variables was studied: 1) status (AC vs. HAM/TSP for HTLV-1 infection and resolved vs. chronic for HCV infection) and 2) viral burden (log10[proviral load] for HTLV-1 infection and log10[viral load] for HCV infection). Multiple logistic regression was used to study the variation in status and multiple linear regression was used for the variation in viral burden. All statistically significant and potentially confounding variables were included in the models (HTLV-1: age and gender; HCV: HBV status, mode of infection, SNP rs12979860 and cohort). We focused on the difference between KIR+ and KIR- individuals in the size of the protective/detrimental effects associated with the individual HLA class I molecules (i.e. odds ratios and differences in viral load) rather than p values as the latter comparison is confounded by differences in strata sizes. As opposed to Chapter 2, here **OR<1 implies a protective effect and OR>1 implies a detrimental effect** so that the comparison with the previous studies is immediate.
HCV viral load

The HCV cohorts were analysed separately as viral load was measured using different assays. Two cohorts (MHCS and ALIVE) had sufficient numbers of individuals with measurements of all variables for this analysis. For MHCS we used the median of multiple measurements of viral load; for ALIVE, a single measurement was available for each subject. We only analysed viral load in patients with chronic infection, so any observed impact on viral load is independent of the impact on viral clearance.

Linkage Disequilibrium

The LD for HLA class I alleles is calculated as described in Chapter 1. The LD between KIRs is calculated based on the Chi-squared test on a 2x2 contingency table.

False Discovery Rate

We quantified the false discovery rate for our analysis using Monte Carlo methods. For each of the HLA associations studied, we performed 10,000 random stratifications of the relevant (HCV or HTLV-1) cohort with the size of the two strata being the number of KIR2DL2+ and KIR2DL2- individuals in that cohort and asked how many times we would see odds ratios equal to or more extreme than we observed in the actual cohorts. The probability of making our observations by chance is less than $p=2 \times 10^{-11}$ (Appendix Table B.1).

3.3.3 Epitope prediction

We produced a list of rank values (see Methods of Chapter 2) for each protein to each allele that quantifies the binding relationship between that allele and the protein. We then invert the rank so the bigger 1/rank, the stronger the preference of the allele for the protein; the logarithm of this measure is plotted on the y axis of Figure 3.1 as ‘HBZ binding score’. Each individual therefore contributes up to 4 values (alleles for which no
predictive algorithms were available were excluded from the analysis). Binding scores were compared between ACs and HAM/TSP patients using the Wilcoxon rank sum test and reported both as separate p values for HLA-A and B molecules and combined (since we found no evidence to reject the null hypothesis that the HBZ binding score of an individual’s A and B molecules was independent, spearman correlation=0.05 p=0.46). The median difference in binding score is the median of the difference of average HBZ binding between ACs and HAM/TSP patients expressed as a percent of the AC binding score for HLA-A and -B molecules.

3.4 **KIR2DL2 in HTLV-1 infection**

3.4.1 Disease status

In the cohort from Southern Japan, HLA-C*08 was associated with a significantly reduced odds of developing HAM/TSP (OR=0.47, p=0.03, OR<1 indicates a protective effect while OR>1 indicates a detrimental effect) [23]. We investigated the impact of KIRs on this protective effect by stratifying the cohort by KIR genotype. Of the KIR studied, one particular KIR, KIR2DL2, had a noticeable interaction with C*08 (Table 3.1 and Figure 3.3). We found that the C*08 protective effect was weakened and no longer statistically significant in the subset of individuals who were KIR2DL2- (OR=0.67, p=0.4) but enhanced in KIR2DL2+ individuals (OR=0.16, p=0.02). There were more KIR2DL2- individuals (N=300) than KIR2DL2+ individuals (N=102) so the absence of significance in the KIR2DL2- individuals was not simply due to reduced cohort size. Similarly, HLA-B*54, which is associated with a significantly increased risk of HAM/TSP (OR= 3.11, p=0.0009), had a weakened impact on disease risk in the absence of KIR2DL2 (OR=1.70, p=0.2) but an enhanced impact in the presence of KIR2DL2 (OR=12.05, p=0.004). Again, the absence of a significant effect of B*54 in KIR2DL2- individuals was not attributable to a loss of power. In contrast, although
$HLA-A^*02$ was associated with a reduced risk of HAM/TSP, there was no significant additional impact of $KIR2DL2$ genotype which could not be attributed to power.

### 3.4.2 Proviral load

As an independent test of the observation that $KIR2DL2$ enhanced the effect of both protective and detrimental HLA class I alleles in HTLV-1 infection, we investigated the interaction between HLA class I alleles, $KIR2DL2$ and HTLV-1 proviral load (pvl). We investigated pvl in ACs and HAM/TSP patients separately, so any observed impact on pvl is independent of the impact on disease status. $C^*08$ has previously been associated with a low pvl in ACs (difference in log10 pvl between $C^*08^+$ and $C^*08^-$ $\Delta=-0.33$, $p=0.05$); again, this effect was weakened in $KIR2DL2^-$ individuals ($\Delta=-0.29$ $p=0.18$) but enhanced in $KIR2DL2^+$ individuals ($\Delta=-0.66$ $p=0.07$); Table 3.1. Similarly, $HLA-B^*54$, which is associated with a high pvl in HAM/TSP patients ($\Delta=+0.24$ $p=0.01$) showed a weakened effect in the absence of $KIR2DL2$ ($\Delta=+0.22$, $p=0.05$) but an enhanced effect in the presence of $KIR2DL2$ ($\Delta=+0.42$, $p=0.01$).

Two previous observations on HTLV-1 immunogenetics have, until now, remained unexplained. Firstly, although $C^*08$ has been associated with a low pvl in ACs it has no detectable impact on pvl in HAM/TSP patients; similarly, $B^*54$, which was associated with a high pvl in HAM/TSP patients, had no impact on pvl in ACs [23,127]. Why some HLA class I alleles apparently ‘cease working’ in some populations was unknown. We hypothesised that the lack of the expected $C^*08$ and $B^*54$ effects in HAM/TSP patients and ACs respectively was due to a low frequency of $KIR2DL2$ in these groups and that the decrease or increase in pvl due to $C^*08$ or $B^*54$ respectively would be manifest only in $KIR2DL2^+$ individuals. Consistent with this hypothesis we found that the frequency of $KIR2DL2$ carriage in the groups that did not show the expected effect of HLA genotype on pvl was approximately half that of the groups in which HLA-associated effects were observed (prevalence of $KIR2DL2^+$ amongst
## Table 3.1: KIR2DL2 in HTLV-1 infection

<table>
<thead>
<tr>
<th>HLA Allele</th>
<th>OR in whole cohort (p value)</th>
<th>Genotype</th>
<th>OR in stratified cohort (p value)</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*08</td>
<td>2.126 (p=0.032)</td>
<td>+</td>
<td>6.249 (p=0.02)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>1.504 (p=0.364)</td>
<td>44</td>
</tr>
<tr>
<td>B*54</td>
<td>0.322 (p=0.0009)</td>
<td>+</td>
<td>0.083 (p=0.004)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>0.588 (p=0.173)</td>
<td>64</td>
</tr>
</tbody>
</table>

**HTLV-1 disease status (HAM/TSP v AC)**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>KIR2DL2</th>
<th>HLA Allele Carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACs</td>
<td>+</td>
<td>C*08</td>
</tr>
<tr>
<td>HAM/TSP</td>
<td>-</td>
<td>B*54</td>
</tr>
</tbody>
</table>

**HTLV-1 proviral load**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Difference in logPVL in whole cohort (p value)</th>
<th>Genotype</th>
<th>Difference in logPVL in stratified cohort (p value)</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACs</td>
<td>-0.33 (p=0.047)</td>
<td>+</td>
<td>-0.66 (p=0.066)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-0.286 (p=0.181)</td>
<td>26</td>
</tr>
<tr>
<td>C*08</td>
<td>-0.173 (p=0.208)</td>
<td>+</td>
<td>-0.856 (p=0.005)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-0.087 (p=0.578)</td>
<td>18</td>
</tr>
<tr>
<td>HAM/TSP</td>
<td>0.057 (p=0.778)</td>
<td>+</td>
<td>-0.218 (p=0.709)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>0.095 (p=0.676)</td>
<td>21</td>
</tr>
<tr>
<td>ACs</td>
<td>0.24 (p=0.012)</td>
<td>+</td>
<td>0.417 (p=0.014)</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>0.224 (p=0.049)</td>
<td>43</td>
</tr>
<tr>
<td>B*54</td>
<td>0.24 (p=0.012)</td>
<td>+</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3.1: KIR2DL2 in HTLV-1 infection: KIR2DL2 enhances the protective effect of C*08 and the detrimental effect of B*54 on HAM/TSP risk and, independently, on proviral load. In the disease status models an odds ratio (OR)<1 indicates a protective effect (decreased risk of HAM/TSP), an OR>1 indicates a detrimental effect (increased risk of HAM/TSP). In the proviral load models the dependent variable was log10(proviral load), a difference in PVL<0 indicates a protective effect (decreased PVL with the HLA allele), a difference in PVL>0 indicates a detrimental effect (increased PVL with the HLA allele). All models also included the two variables which can act as confounding variables in this cohort: age and gender. The impact on proviral load was considered separately in ACs and HAM/TSP and so the observation of an impact on proviral load is independent of the observation of an impact on status. If variables which were not significant predictors (p<0.05) were removed by backwards stepwise exclusion the conclusions were unchanged.
individuals: 12.5% in ACs vs 29.5% in HAM/TSPs; prevalence of KIR2DL2+ amongst C*08+ individuals: 18.2% in HAM/TSPs vs 27.8% in ACs; Table 3.1). The small numbers of individuals in the stratified cohorts (HAM/TSP KIR2DL2+: C*08+ N=4, C*08- N=50. ACs KIR2DL2+: B*54+ N=3, B*54- N=45) precluded a reliable test for an impact of HLA on pvl in KIR2DL2+ individuals. However, in the larger of these groups there was a significant impact; i.e C*08 was associated with a significant reduction in pvl in KIR2DL2+ individuals (Δ=-0.86, p=0.005). This provides, for the first time, a plausible explanation for the reported observation [23] that the B*54 effect on pvl was not manifest in ACs and the C*08 effect on pvl was not manifest in HAM/TSP patients.

3.4.3 HLA class I specificity

It was recently reported that in HTLV-1 infection, HLA class I molecules that bind peptides from the virus protein HBZ are associated with a reduced risk of HAM/TSP and, independently, a reduced pvl [185]. In the same study it was shown, using IFN-γ ELISpot, chromium release and CD107 staining, that HBZ-specific CD8+ T cells were present and functional in fresh PBMC from infected individuals. We therefore investigated the interaction between KIR2DL2 and the protective effect of binding HBZ. We used epitope prediction software [198] to predict the strength of binding of HBZ peptides to an individual’s HLA-A and B molecules. It was already reported that ACs carry HLA-A and -B molecules that are predicted to bind HBZ significantly more strongly than those carried by HAM/TSP patients (median difference 12%, p=0.00005). We found that this effect was stronger in KIR2DL2+ individuals (median difference 25%, p=0.00006) than in KIR2DL2- individuals (median difference 7%, p=0.06); Figure 3.1. We reasoned that this difference in HBZ binding between ACs and HAM/TSP patients was due to HLA-A*02 and B*54, which differ in their HBZ peptide-binding affinities [185] and are associated with different outcomes in HTLV-1
infection. We therefore removed all individuals with \textit{A*02} or \textit{B*54} from the cohort and repeated the analysis. Surprisingly, we still found the same pattern: possession of HLA molecules that bind HBZ strongly was significantly associated with remaining asymptomatic (median difference 10\%, \textit{p}=0.038) and this effect was strengthened in \textit{KIR2DL2+} individuals (median difference 23\%, \textit{p}=0.022) but not in \textit{KIR2DL2-} individuals (median difference 3\%, \textit{p}=0.177); Figure 3.1. This demonstrates that the protective effect of binding HBZ peptides by multiple HLA class I molecules, both A and B, is enhanced by \textit{KIR2DL2}.

\section*{3.5 \textit{KIR2DL2} in HCV infection}

\subsection*{3.5.1 Spontaneous viral clearance}

As previously reported, \textit{HLA-B*57} was associated with significantly decreased odds of chronic infection (OR=0.571, \textit{p}=0.023). This protective effect was enhanced in the presence of \textit{KIR2DL2} (OR=0.40, \textit{p}=0.007) but weakened in the absence of \textit{KIR2DL2} (OR=0.83, \textit{p}=0.63) (Table 3.2). Furthermore, the impact of \textit{B*57} was strongest in \textit{KIR2DL2} homozygote individuals (OR=0.21, \textit{p}=0.024), weaker in \textit{KIR2DL2} heterozygote individuals (OR=0.48, \textit{p}=0.07) and absent in \textit{KIR2DL2}-negative individuals (OR=0.83, \textit{p}=0.63). \textit{KIR2DL2} enhanced the association between \textit{B*57} and spontaneous clearance independently and with similar strength in both African-Americans and Caucasians (Table 3.3).

\subsection*{3.5.2 Viral load in chronic infection}

Next we investigated the impact of \textit{B*57} and \textit{KIR2DL2} on HCV viral load. We only considered patients with chronic infection, so any observed impact on viral load is independent of the impact on viral clearance. This analysis was possible in two cohorts: MHCS and ALIVE. We found that \textit{B*57} was associated with reduced chronic
Figure 3.1: In the whole cohort (left column) individuals with HLA-A or B molecules which were predicted to bind peptides from HBZ strongly were significantly more likely to be asymptomatic (top row); this effect was stronger in the individuals with KIR2DL2 (middle row) than without (bottom row). When individuals with A*02 or B*54 were removed (right column) the pattern remained the same. This indicates that the protective effect of binding HBZ peptides by multiple, different HLA class I molecules, both A and B, is enhanced by KIR2DL2. We did not predict HLA-C binding as the relevant algorithms are not available in Metaserver (due to scarcity of HLA-C binding data necessary for training the neural networks).
<table>
<thead>
<tr>
<th>HCV status (spontaneous clearance v chronic infection)</th>
<th>OR in whole cohort</th>
<th>KIR2DL2 Genotype</th>
<th>OR in stratified cohort</th>
<th>HLA Allele Carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(p value)</td>
<td>+</td>
<td>2.482</td>
<td>49</td>
<td>408</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>1.202</td>
<td>35</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+/-</td>
<td>4.796</td>
<td>16</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/+</td>
<td>2.104</td>
<td>33</td>
<td>324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-/-</td>
<td>1.202</td>
<td>35</td>
<td>374</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HCV Viral load</th>
<th>Difference in logVL in whole cohort</th>
<th>KIR2DL2 Genotype</th>
<th>Difference in logVL in stratified cohort</th>
<th>HLA Allele Carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(p value)</td>
<td>MHCS</td>
<td>ALIVE</td>
<td>MHCS</td>
<td>ALIVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-3.106</td>
<td>-0.180</td>
<td>+</td>
<td>-4.485</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-1.64</td>
</tr>
</tbody>
</table>

Table 3.2: KIR2DL2 in HCV infection: KIR2DL2 enhances the protective effect of HLA-B*57 on HCV status (spontaneous clearance v chronic infection) and, independently, on HCV viral load in chronic patients. In the HCV status models an odds ratio (OR)<1 indicates a protective effect (decreased odds of chronic infection), an OR>1 indicates a detrimental effect (increased odds of chronic infection). In the viral load models the dependent variable was log10[viral load]. Cohorts were analysed separately as viral load was measured using different assays. Two cohorts (MHCS and ALIVE) had sufficient numbers of individuals with measurements of all variables for this analysis. A difference in VL<0 indicates a protective effect (decreased vl with the HLA allele), a difference in VL>0 indicates a detrimental effect (increased vl with the HLA allele). All models included all variables which can act as confounders (see Methods). The impact on viral burden was considered within chronic carriers and so the observation of an impact on viral burden is independent of the observation of an impact on status.
### Table 3.3

<table>
<thead>
<tr>
<th>Cohort</th>
<th>OR</th>
<th>LCI</th>
<th>UCI</th>
<th>p-value</th>
<th>Allele carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All</strong></td>
<td>0.569</td>
<td>0.351</td>
<td>0.923</td>
<td>0.023</td>
<td>84</td>
<td>782</td>
</tr>
<tr>
<td><strong>African-Americans</strong></td>
<td>0.482</td>
<td>0.222</td>
<td>1.046</td>
<td>0.065</td>
<td>34</td>
<td>240</td>
</tr>
<tr>
<td><strong>Caucasians</strong></td>
<td>0.556</td>
<td>0.283</td>
<td>1.092</td>
<td>0.087</td>
<td>43</td>
<td>483</td>
</tr>
<tr>
<td><strong>KIR2DL2+</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>0.403</td>
<td>0.208</td>
<td>0.781</td>
<td>0.007</td>
<td>49</td>
<td>408</td>
</tr>
<tr>
<td><strong>African-Americans</strong></td>
<td>0.354</td>
<td>0.130</td>
<td>0.960</td>
<td>0.041</td>
<td>25</td>
<td>125</td>
</tr>
<tr>
<td><strong>Caucasians</strong></td>
<td>0.363</td>
<td>0.134</td>
<td>0.985</td>
<td>0.047</td>
<td>21</td>
<td>252</td>
</tr>
<tr>
<td><strong>KIR2DL2-</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>0.832</td>
<td>0.392</td>
<td>1.764</td>
<td>0.632</td>
<td>35</td>
<td>374</td>
</tr>
<tr>
<td><strong>African-Americans</strong></td>
<td>0.649</td>
<td>0.150</td>
<td>2.801</td>
<td>0.563</td>
<td>9</td>
<td>115</td>
</tr>
<tr>
<td><strong>Caucasians</strong></td>
<td>0.755</td>
<td>0.286</td>
<td>3.759</td>
<td>0.57</td>
<td>22</td>
<td>231</td>
</tr>
</tbody>
</table>

Table 3.3: There was a trend for *B*:*57* to be associated with an increased odds of HCV clearance in both African-Americans and in Caucasians. On stratifying the cohorts by *KIR2DL2* genotype this trend became significant for *KIR2DL2+* individuals but was lost for *KIR2DL2-* individuals. The same pattern and a similar strength of effect were seen in both African-Americans and Caucasians.
HCV viral load, particularly in the MHCS cohort (MHCS: difference in log10 VL $\Delta=-3.1$, $p=0.0003$; ALIVE: $\Delta=-0.18$, $p=0.336$. Combined $p=0.0006$). Consistent with our observations in HTLV-1 infection, this reduction was enhanced in the presence of KIR2DL2 (MHCS: $\Delta=-4.5$ $p<0.0001$, ALIVE: $\Delta=-0.46$ $p=0.046$. Combined $p=0.00003$) but weakened in the absence of KIR2DL2 (MHCS $\Delta=-1.64$, $p=0.24$, ALIVE $\Delta=+0.32$, $p=0.35$. Combined $p=0.66$); Table 3.2 and Figure 3.2. In MHCS (but not ALIVE) we also observed a progressive effect with KIR2DL2 copy number (2 copies: $\Delta=-6.5$, $p=0.0005$. 1 copy: $\Delta=-4.1$, $p=0.001$. 0 copies $\Delta=-1.6$, $p=0.24$); however, in this case, the number of homozygous individuals is too small and does not allow to draw conclusions.

### 3.6 No KIR2DL2 main effect on outcome

These data show that KIR2DL2 enhances both protective and detrimental HLA class I associations (Figure 3.3) with disease outcome and viral burden. Therefore, KIR2DL2 would not be predicted to have a significant net impact across all HLA class I molecules. That is, possession of KIR2DL2 (alone or with its C1 ligand) without a particular protective or detrimental HLA allele, would not be expected to be significantly protective or detrimental. This prediction was verified (Table 3.4).

### 3.7 Linkage disequilibrium

There is strong linkage disequilibrium between the KIR genes and between the HLA class I alleles (see below). Here, we thoroughly analyse linkage between HLA class I alleles and amongst KIR genes in order to identify which molecules are driving the effect on infection outcome.
Figure 3.2: In HCV infection, $B^*57$ is associated with a reduced HCV viral load. This protective effect is enhanced in $KIR2DL2^+$ individuals and reduced or absent in $KIR2DL2^-$ individuals. The cohorts were analysed separately as viral load was measured using different assays. Two cohorts, MHCS and ALIVE, had enough individuals with measurements for all factors to perform this analysis. HIV-1 status was a significant determinant of viral load in the ALIVE cohort, including HIV status in the model as a covariate did not change any of our conclusions (ALL $p=0.336$; $KIR2DL2^+ p=0.046$; $KIR2DL2^- p=0.350$). The $p$ values were obtained based on the linear regression model. Summary data from these plots is provided in Table 3.2.
Figure 3.3: The effect of both protective (A) and detrimental (B) HLA class I alleles on both status (first column) and viral load (second column) is stronger in \textit{KIR2DL2}\textsuperscript{+} individuals than in \textit{KIR2DL2}\textsuperscript{-} individuals. This effect was seen in both HTLV-1 and HCV infection. We also saw a progressive effect of \textit{KIR2DL2} copy number (C). We could not test for a progressive effect in the HTLV-1 cohort because there are no \textit{KIR2DL2} homozygotes. In all cases, the impact on viral burden was considered within (not between) disease status categories and so the observation of an impact on viral burden is independent of the observation of an impact on status. Effect sizes, p values and cohort sizes are provided in Tables 3.1 and 3.2.
Table 3.4: \textit{KIR2DL2} does not have an effect on status or viral burden, with or without its C1 ligand.
3.7.1 Linkage between HLA class I alleles

HTLV-1: *C*08 rather than linked HLA alleles appears to be the primary allele driving protection which is enhanced by KIR2DL2

In the HTLV-I cohort, *C*08 was in linkage disequilibrium with *B*40 and *B*48; we therefore sought to establish which was the primary HLA class I gene driving the protective association which was enhanced by KIR2DL2.

*C*08 & *B*40 In a logistic regression model to predict disease status (HAM/TSP v AC) when both *C*08 and *B*40 were included as factors *C*08 retained a protective trend (OR=0.52 p=0.07) whereas *B*40 lost significance (OR=0.88 p=0.3); stratifying on KIR2DL2 we again found that in KIR2DL2+ individuals, *C*08 retained a protective trend (OR=0.22 p=0.07) and *B*40 lost significance (OR=0.49, p=0.20) and that, as expected, neither was significant in the absence of KIR2DL2. Similarly, in a linear regression model to predict log10(proviral load) in ACs when both *C*08 and *B*40 were included as factors *C*08 retained significance (difference in logVL=-0.37 p=0.03) whereas *B*40 lost significance (indeed went in the other direction, difference in logVL=+0.14 p=0.3); stratifying on KIR2DL2 we again found that in KIR2DL2+ individuals the *C*08 effect was strengthened (difference in logVL=-0.64 p=0.07) and *B*40 lost significance (difference in logVL=+0.3, p=0.20) and that, as expected, neither was significant in the absence of KIR2DL2. We therefore concluded that *C*08 rather than *B*40 was the HLA gene most likely to be associated with protection whose effect was enhanced by KIR2DL2.

*C*08 & *B*48 In a logistic regression model to predict disease status (HAM/TSP v AC) when both *C*08 and *B*48 were included as factors both factors lost significance. However, stratifying on KIR2DL2 we found that in KIR2DL2+ individuals *C*08 retained a protective trend (OR=0.21, p=0.07) and *B*48 lost significance (OR=0.33, p=0.34) and that, as expected, neither was significant in the absence of KIR2DL2.
In a linear regression model to predict log10(proviral load) in ACs when both C*08 and B*48 were included as factors C*08 retained a trend (difference in logVL=-0.36 p=0.08) whereas B*48 lost significance (indeed went in the other direction, difference in logVL=+0.08 p=0.78); stratifying on KIR2DL2 we again found that in KIR2DL2+ individuals the C*08 effect was strengthened (difference in logVL=-1.1 p=0.03) and B*48 lost significance (again went in the opposite direction, difference in logVL=+0.7, p=0.24) and that, as expected, neither was significant in the absence of KIR2DL2. Taken together, these data suggest that C*08 rather than B*48 was the gene most likely to be associated with protection whose effect was enhanced by KIR2DL2.

HTLV-1: B*54 rather than linked HLA alleles appears to be the primary allele driving susceptibility which is enhanced by KIR2DL2

In the HTLV-I cohort, B*54 was in linkage disequilibrium with C*01. We therefore investigated i) whether B*54 or C*01 was the primary gene associated with increased susceptibility to HAM/TSP and ii) whether B*54 or C*01-associated susceptibility was enhanced by KIR2DL2. In a logistic regression model when both B*54 and C*01 were included as factors B*54 retained significance (OR=3.84 p=0.0009) whereas C*01 lost significance (indeed went in the opposite direction OR=0.73 p=0.3); stratifying on KIR2DL2 we again found that in KIR2DL2+ individuals B*54 retained significance and C*01 lost significance and that, as expected, neither was significant in the absence of KIR2DL2. If all B*54+ individuals were removed from the cohort then C*01 was no longer detrimental (OR=0.88, p=0.4, C*01+=96). Unfortunately, due to the large number of C*01+ individuals in the cohort (N=184) it was not possible to reverse this analysis and investigate the impact of B*54 in the absence of C*01.
HCV: \( B^*57 \) rather than linked HLA genes appears to be the primary gene driving protection which is enhanced by \( KIR2DL2 \)

In the HCV cohort, \( B^*57 \) is in linkage disequilibrium with \( A^*01, C^*06 \) and \( C^*18 \). Hence, we investigated whether the observed protective effect of \( B^*57 \) can be attributed to the other linked alleles. We found that \( A^*01, C^*06 \) and \( C^*18 \) do not have a significant impact on disease status neither overall nor in the context of \( KIR2DL2 \). We therefore conclude that \( B^*57 \) is the HLA allele associated with HCV clearance and also the one enhanced by \( KIR2DL2 \).

### 3.7.2 Linkage between KIR genes

The KIR genes are in tight linkage disequilibrium (Table 3.5), making it hard to definitively ascertain which KIR enhances the HLA-associated effects (i.e. the association between \( C^*08 \) and asymptomatic status in HTLV-1 infection, between \( B^*54 \) and HAM/TSP in HTLV-1 infection and between \( B^*57 \) and spontaneous viral clearance in HCV infection). To try to determine which was the primary KIR driving the enhancement of the HLA-associated effects we constructed a logistic regression model to predict status (HAM/TSP v AC for HTLV-1 infection, spontaneous clearance v persistence for HCV infection) in which the HLA molecule with the presence or absence of each KIR, depending on the stratum in which the HLA had the more significant effect, was included along with the known confounding factors. Then, we remove the HLA:KIR factors by stepwise backwards exclusion (i.e. by the highest p-value one at a time, refit the model and repeat). In all 3 cases (\( C^*08, B^*54 \) and \( B^*57 \)) the only HLA:KIR compound that remains in the model is the HLA:KIR2DL2+. This analysis suggests that \( KIR2DL2 \) is most likely to be the primary gene driving the observed effect.

\( KIR2DL2 \) is in particularly tight LD with \( KIR2DS2 \), an activating receptor. It could be argued that an activatory receptor is more likely to modulate CD8+ T cells
Table 3.5: Linkage between the KIRs in the HTLV-1 (top) and HCV (bottom) cohorts.
Positive linkage disequilibrium (LD) is shown in gray and negative in black. The statistical significance (p-values) of the LD is displayed in the cells of the tables.
and so we specifically investigated whether the observed effect was more likely to be driven by \textit{KIR2DL2} or \textit{KIR2DS2}. In all cases logistic regression models to predict status where HLA allele:KIR2DL2 and HLA allele:KIR2DS2 (plus confounders) were included as simultaneous covariates then the covariate HLA:KIR2DL2 was more significant than HLA:KIR2DS2 in every case (i.e. for \textit{HLA-B*57} in HCV, \textit{B*54} in HTLV-1 and \textit{C*08} in HTLV). However, in a model where HLA:KIR2DS2 but not HLA:KIR2DL2 was a covariate then HLA:KIR2DS2 was significant for \textit{B*54} and \textit{B*57} (but not \textit{C*08}). For \textit{B*54} and \textit{B*57} we therefore also investigated whether the HLA allele was significant in the cohort that was \textit{KIR2DL2+/KIR2DS2-}. Although the cohorts were small to draw concrete conclusions, both alleles had a significant effect in \textit{KIR2DL2+/KIR2DS2-} (\textit{B*54}: OR=18.5, p=0.016, n=10, N=48; \textit{B*57}: OR=0.47, p=0.02, n=2, N=13). The size of the\textit{KIR2DL2-/KIR2DS2+} did not allow further investigation (\textit{B*54}: n=2, N=6; \textit{B*57}: n=1, N=8). The small cohort sizes makes it impossible to conclude that \textit{KIR2DS2} does not also have an effect but it is clear that even in the absence of \textit{KIR2DS2}, \textit{KIR2DL2} does have an effect. Together these three observations indicate that \textit{KIR2DL2} rather than \textit{KIR2DS2} is more likely to be the primary gene driving the enhancement of HLA class I-mediated antiviral immunity.

Additionally, HLA-KIR factors which may be particularly relevant because they are known receptor-ligand pairs (e.g. HLA-B*57 with KIR3DL1 or KIR3DS1) were examined in more detail (Appendix Table B.3 and B.4) but all evidence indicated that the effect is driven by \textit{KIR2DL2}.

In summary, the evidence suggests that \textit{KIR2DL2} is most likely to be the KIR which is enhancing immunity. The one KIR for which it is impossible to assess whether it has a stronger effect than \textit{KIR2DL2} is \textit{KIR2DL3} as \textit{KIR2DL2} and \textit{KIR2DL3} segregate as alleles of the same locus. Both are inhibitory but \textit{KIR2DL2} provides stronger inhibitory signals than \textit{KIR2DL3} \cite{37}. The 2DL2/L3 locus is present in one copy in the majority of haplotypes so the observation that \textit{KIR2DL2} is present in an individual (1 or 2 copies) implies that there are 0 or 1 copies of \textit{KIR2DL3}. So the statement
that the presence of $KIR2DL2$ enhances class I mediated immunity can be restated in the reciprocal as lack of $KIR2DL3$ homozygosity enhances class I mediated immunity. However, functionally it is difficult to understand how the lack of homozygosity for a weaker receptor should enhance immunity more effectively than homozygosity for a stronger receptor.

### 3.7.3 Conclusions on LD

Analysis of the linked genes/alleles indicates that the primary molecules driving the observed associations are most likely to be $KIR2DL2$ in combination with $HLA-B*54$, $C*08$ and $B*57$ rather than individual linked KIR, multiple stimulatory linked KIRs or linked HLA class I alleles. We cannot rule out an effect of linkage between $KIR2DL2$ and neighbouring loci outside the KIR genes. However, there is little evidence of significant linkage between KIRs and even the next closest gene cluster, the LILR \[222\]. Furthermore, we observed the same effect of $KIR2DL2$ in three different populations (Japanese, African-American and Caucasian) so a putative linked locus driving the effect would have to be linked to $KIR2DL2$ in all three populations.

### 3.8 Canonical KIR-HLA binding

We next explore whether direct binding of KIR2DL2 with any of the HLA class I molecules studied (namely C*08, B*54 and B*57) can explain the enhancement that we observe.

Although HLA-C*08, as a group C1 molecule, is expected to bind KIR2DL2, the most frequent subtype in our cohort (C*08:01, 88%) binds KIR2DL2 very weakly (comparable to background \[223\]), furthermore HLA-B*54 and HLA-B*57 are not expected to bind KIR2DL2 and the most frequent subtypes in our cohorts (B*54:01 and B*57:01) have been shown not to bind KIR2DL2 \[76,223\]. Finally, KIR2DL2 enhanced the pro-
tective effect of binding HBZ peptides by multiple HLA-A and -B molecules. With the exception of B*46:01 and B*73:01 (which were not responsible for the enhancement, data not shown) KIR2DL2 is not thought to bind HLA-A and B molecules and has been shown not to bind 29/29 HLA-A and 54/56 HLA-B allotypes tested in [76]. We therefore hypothesised that the effect of KIR2DL2 on HLA class I-mediated immunity we have observed is not attributable to KIR2DL2 directly binding the HLA molecule whose effect is enhanced. To test this hypothesis we first investigated whether the other group C1 alleles had the same effect as C*08 in HTLV-1 infection. Grouping all the C1 alleles we found no significant association between C1 and decreased risk of HAM/TSP either in the whole cohort or in KIR2DL2+ individuals. Similarly, there was no relation between pvl and C1 in either ACs or HAM/TSP patients. Analysis of the individual C1 alleles confirmed the hypothesis that the C*08 effect we observed was not exhibited by other group C1 alleles (Appendix Table B.2).

HLA-B*54, a group Bw6 HLA allele, is not known to bind any KIR molecule. We therefore tested whether the observed B*54 effect was attributable to C*01, which is in linkage disequilibrium with B*54 and which encodes molecules that bind KIR2DL2. This analysis suggested that B*54, not C*01, was the gene driving the observed detrimental effect on HTLV-1 outcome (see Appendix B). This result, and the observation that no other C1 allele shows ‘B*54-like’ behaviour, indicate that, as postulated, the interaction between B*54 and KIR2DL2 cannot be explained by direct KIR-HLA binding.

The most frequent B*57 allele in our cohort is B*57:01, which does not bind KIR2DL2 [76]. There are therefore two ways in which the observed interaction between KIR2DL2 and HLA-B*57 could be attributed to ‘classical’ KIR-HLA binding: either KIR2DL2 might bind a class I HLA molecule whose encoding gene is linked to HLA-B*57, or the effect might be due to KIR3DL1/S1, which does bind B*57. Analysis of both these possibilities indicated that they did not explain the KIR2DL2-B*57 effect (see Appendix B).
As shown so far, the enhancement of C*08, B*54 and B*57-restricted immunity by KIR2DL2 is not explained by direct binding between the respective HLA molecules and KIR2DL2. Instead, we suggest that KIR2DL2 binds its HLA-C ligands and indirectly modulates C*08, B*54 and B*57-restricted T cells. Consistent with this, we found some weak evidence that KIR2DL2 enhanced HLA class I effects more strongly when it’s stronger C1 ligands are present (see section 3.9).

3.9 The role of KIR2DL2 ligands

KIR2DL2 binds HLA group C1 molecules and, with weaker affinity, C2 molecules [76]. We hypothesised that KIR2DL2-dependent enhancement of HLA-mediated immunity would be greatest in individuals bearing one or two copies of the group C1 ligand. This required further stratification of the cohorts and only the size of the HCV cohort allowed for such calculations. We indeed found that in KIR2DL2+ individuals who had at least one C1 ligand, the effect of B*57 was protective (OR=0.34, p=0.008, B57+=31, N=332) and it was slightly weakened when KIR2DL2 was present without its C1 ligand (OR=0.4, p=0.38, B57+=18, N=76). However, the differences are very small and certainly not conclusive, possibly because KIR2DL2 also binds C2 molecules and so there are no people in which KIR2DL2 does not have a ligand.

3.10 The role of other KIRs

3.10.1 Other Inhibitory KIRs

It seems unlikely that KIR2DL2 behaves fundamentally differently to other inhibitory KIRs. The effect of KIR2DL2 may be most apparent because KIR2DL2 is present at informative frequencies and its C1 and C2 ligands are ubiquitous [76]. We addressed the role of other inhibitory KIRs in 3 ways. (i) We studied the effect of individual
inhibitory KIRs (section 3.7.2), (ii) we investigated whether the number of inhibitory KIR:ligands had a cumulative effect and (iii) we examined the role of the group A KIR haplotypes which are dominated by inhibitory KIRs but do not contain KIR2DL2 (section 3.10.3). We found little evidence that the other inhibitory KIRs enhanced HLA class I-mediated immunity but this may be due to small cohort sizes and masking by the dominant KIR2DL2 effect.

We quantified the magnitude of KIR inhibitory signals per individual by counting the number of inhibitory KIR that were present with their ligand (KIR2DL1:C2, KIR2DL2:C1, KIR2DL3:C1, KIR3DL1:Bw4I [78, 224, 225]). Then, we stratified the cohort for ‘low inhibitory signal’, count<1, and ‘high inhibitory signal’, count<2. The cut-off (count<1, count<2) was chosen so that there were similar numbers of individuals in the two strata. In HCV infection, we found that B*57 had an increased protective effect for individuals within the ‘high inhibitory signal’ category (OR=0.48, p=0.03, B57+=50, N=518) but the effect was not significant for ‘low inhibitory signal’ individuals (OR=0.65, p=0.3, B57+=34, N=262). In HTLV-1 infection, the protective effect of C*08 was pronounced among individuals with a ‘high inhibitory signal’ (OR=0.30, p=0.02, C*08+=29, N=242) but absent in the ‘low inhibitory signal’ group (OR=0.98, p=0.97, C*08+=29, N=160). Similarly, the detrimental impact of B*54 was increased in individuals that possessed more inhibitory KIRs with their ligands (OR=5.31, p=0.001, B54+=43, N=242) compared to individuals with a ‘low inhibitory signal’ score (OR=1.58, p=0.36, B54+=42, N=160). However, the effect of the ‘high inhibitory KIR signal’ could be attributable to KIR2DL2 as the ‘high inhibitory signal’ group is heavily enriched for individuals with KIR2DL2 (especially for the HTLV-1 cohort). Additionally, the observation (discussed in section 3.10.3) that the KIR haplotype AA does not have an enhancing effect suggests that even if other inhibitory signals may contribute, KIR2DL2 is necessary for a detectable effect.
3.10.2 Activatory KIRs

We found no evidence that activating KIR were enhancing HLA class I-restricted immunity. Haplotype B, the more activatory KIR haplotype, enhanced HLA class I associations but this was only true if the haplotype contained KIR2DL2 (section 3.10.3). We found no evidence that the cumulative presence of activating KIR enhanced HLA class I restricted immunity. And, as far as it was possible to separate KIR2DL2 and KIR2DS2, which are in tight linkage disequilibrium, the enhancement of HLA class I restricted immunity appeared to be attributable to KIR2DL2 rather than KIR2DS2 (section 3.7.2).

KIR2DL2 is usually present on haplotypes which contain multiple activatory KIRs (section 3.10.3). We therefore also explored the possibility that the cumulative presence of multiple stimulatory KIR (rather than any one individual KIR) is driving the KIR2DL2 effect. In HCV infection, for each individual, we counted the number of stimulatory receptors of the KIR B haplotype (KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS5, KIR3DS1). Then we simultaneously investigated 1) the effect of B*57 combined with 1 or 2 stimulatory receptors, 2) the effect of B*57 in the presence 3, 4 or 5 stimulatory receptors and 3) the effect of B*57 in the presence of KIR2DL2. Using backward elimination, we found that only the B*57-KIR2DL2 had a significant effect on the outcome of infection. Similar analysis for viral load was not possible because of limited cohort sizes. In HTLV-1 infection, we could apply the same approach only for C*08 and the status variable because of limiting numbers. The stimulatory KIRs available in the cohort were KIR2DS2, KIR2DS3 and KIR3DS1 so we considered individuals with only 1 KIR stimulatory receptor or with 2-3 stimulatory receptors. We found that, as for B*57 in HCV, using backward elimination only C*08-KIR2DL2 had a significant impact on outcome. These two results taken together suggest that KIR2DL2 rather than the cumulative presence of stimulatory receptors of KIR-B haplotype enhance the HLA class I-mediated immunity (further discussed in section 3.10.3).
No *KIR3DS1* effect in HTLV-1 infection

We examined more closely the role of the activatory receptor *KIR3DS1* in HTLV-1 infection. *KIR3DS1*, together with its putative ligand HLA-Bw480I, has been associated with lower viral load and slower progression to AIDS in HIV infection \[^{[85]}\]. A suggested mechanism driving this association is that the activation of *KIR3DS1*+ NK cells may not depend on an HIV-specific signal but occur as a direct or indirect consequence of retroviral infection \[^{[2]}\]. Under this explanation, it would be expected that the ability of *KIR3DS1* to suppress viral load in HIV might also be true in HTLV-1 infection, by lowering the proviral load and protecting from HAM/TSP development. We tested this hypothesis in the current HTLV-1 Japanese cohort but we found no significant association of *KIR3DS1*+Bw480I with either disease outcome or proviral load (Tables 3.6 and 3.7). This result was further confirmed in a Brazilian cohort in \[^{[2]}\].

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio</th>
<th>p-value</th>
<th>% in ACs</th>
<th>% in HAM/TSP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>KIR3DS1</em></td>
<td>1.19</td>
<td>0.5</td>
<td>37.8%</td>
<td>36.5%</td>
</tr>
<tr>
<td><em>Bw480I</em></td>
<td>0.79</td>
<td>0.4</td>
<td>35.6%</td>
<td>39.6%</td>
</tr>
<tr>
<td><em>3DS1+Bw480I</em></td>
<td>1.44</td>
<td>0.3</td>
<td>13.9%</td>
<td>13.5%</td>
</tr>
</tbody>
</table>

Table 3.6: Analysis of *KIR3DS1* and *HLA-Bw480I* on HTLV-1 disease outcome. We have 180 ACs and 222 HAM/TSP individuals in the cohort.

<table>
<thead>
<tr>
<th></th>
<th>ACs</th>
<th>HAM/TSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>p-value</td>
</tr>
<tr>
<td><em>KIR3DS1</em></td>
<td>-0.25</td>
<td>0.09</td>
</tr>
<tr>
<td><em>KIR3DS1+Bw480I</em></td>
<td>-0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 3.7: Analysis of *KIR3DS1* and *HLA-Bw480I* on HTLV-1 proviral load. We have 180 ACs and 222 HAM/TSP individuals in the cohort.
3.10.3 KIR haplotypes

Two broad groups of KIR haplotypes have been defined: A haplotypes and B haplotypes\(^{226}\). KIR-A haplotypes are dominated by inhibitory KIRs (having 2 activating and 5 inhibitory KIRs); KIR-B haplotypes which are less tightly defined, are considered more activatory (having up to 7 activating KIRs and on average 5 inhibitory KIRs). KIR-A haplotypes do not have \textit{KIR2DL2}, B haplotypes can but do not always include \textit{KIR2DL2}. We investigated whether different haplotypes were associated with a different enhancement of the HLA class I associations with infection outcome.

For each individual, the following rules for attributing a KIR haplotype were used. AA: only \textit{KIR2DL1}, \textit{3DL1}, \textit{2DL3} and \textit{2DS4} are present; AB: all the KIRs in haplotype A were present as well as one or more of \textit{KIR2DL2}, \textit{2DS2}, \textit{2DS3}, \textit{2DS5} or \textit{3DS1} and BB: not all the KIRs in haplotype A were present and at least one of the KIRs in haplotype B were present.

We found that the AA haplotypes did not have an effect either on outcome of infection (Table 3.9), nor on viral burden for any of the three HLA molecules studied (data not shown). The KIR-AB haplotypes enhanced the effect of both protective and detrimental molecules, for all three HLA molecules studied. The fact that AB but not AA enhances HLA associations suggests that it is the B haplotype which is responsible for the AB enhancement. Unfortunately, there are insufficient numbers of individuals with BB to draw any conclusions about BB haplotypes.

Alternatively, to investigate the enhancement conveyed by the B haplotype we split the B+ individuals into groups with and without \textit{KIR2DL2} (Table 3.9). This clearly showed that the haplotype B-enhancement was absent when \textit{KIR2DL2} was absent and present when \textit{KIR2DL2} was present. We conclude that a haplotype analysis offers little additional information, KIR-A and KIR-B simply being imperfect markers for absence or presence of \textit{KIR2DL2}.  

122
### Table 3.8: The role of KIR haplotypes.

We had less than 5 observations for BB haplotypes in the HTLV-1 cohort. AA: only KIR2DL1, 3DL1, 2DL3 and 2DS4 are present; AB: all the KIRs in haplotype A were present as well as one or more of KIR2DL2, 2DS2, 2DS3, 2DS5 or 3DS1 and BB: not all the KIRs in haplotype A were present and at least one of the KIRs in haplotype B were present.

<table>
<thead>
<tr>
<th>KIR Haplotype</th>
<th>AA (p=0.5, n=22, N=252)</th>
<th>AB (p=0.1, n=44, N=405)</th>
<th>BB (p=0.07, n=18, N=125)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV B*57 status</td>
<td>0.73</td>
<td>0.57</td>
<td>0.36</td>
</tr>
<tr>
<td>HTLV-1 C*08 status</td>
<td>0.70</td>
<td>0.38</td>
<td>-</td>
</tr>
<tr>
<td>HTLV-1 B*54 status</td>
<td>1.23</td>
<td>5.35</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 3.9: KIR-B haplotype enhancement is not seen in the absence of KIR2DL2.

<table>
<thead>
<tr>
<th>KIR Haplotype</th>
<th>B+(AB or BB) (p=0.02, n=62, N=530)</th>
<th>B+KIR2DL2+ (p=0.007, n=49, N=408)</th>
<th>B+KIR2DL2- (p=0.75, n=13, N=122)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV B*57 status</td>
<td>0.51</td>
<td>0.40</td>
<td>1.23</td>
</tr>
<tr>
<td>HTLV-1 C*08 status</td>
<td>0.35</td>
<td>0.16</td>
<td>0.69</td>
</tr>
<tr>
<td>HTLV-1 B*54 status</td>
<td>6.25</td>
<td>12.50</td>
<td>2.78</td>
</tr>
</tbody>
</table>

(p-value, allele carriers, cohort size)
3.11 Potential mechanism

The antiviral effector population which is enhanced by KIR2DL2 could be either NK (Figure 3.4(a)) cells or T cells (Figures 3.4(b) and 3.4(c)). Three observations indicate that the mechanism is more likely to be T cell-mediated. First, strong binding of the HBZ viral peptide by multiple HLA class I molecules was associated with asymptomatic status and this effect was enhanced by KIR2DL2. Although NK cells can exhibit some peptide dependence, such strong protein specificity is more consistent with T cells. Second, the KIR2DL2 enhancement could not be attributed to direct binding between KIR2DL2 and any of the 3 HLA class I molecules investigated. Third, there is no KIR2DL2 main effect (with or without ligand) on disease outcome or viral burden, an observation which would have implied a direct NK-mediated effect.

It has been demonstrated that CD8+ T cells which express inhibitory KIRs (Figure 3.4(b)) have elevated levels of Bcl-2 and are less susceptible to activation induced cell death (AICD). Of particular interest are reports [79,80,227,228] that inhibitory KIRs on CD8+ T cells promote the survival of a subset of memory phenotype CD8+ αβ T cells with enhanced cytolytic potential (Tm1 [229]) by reducing activation-induced cell death. Tm1 cells have been described in both HTLV-1 and HCV infections, where they constitute a minority of virus-specific CD8+ T cells but the majority of perforin-bright cells [230,231]. Consistent with our findings, these studies have shown that the HLA molecule that restricts the T cell whose survival is promoted was independent of the HLA-C molecules that ligated the KIR [79,229]. We suggest that KIR2DL2 may bind its HLA-C ligands so that when the CD8+ T cell is activated by engagement of its TCR by the cognate pMHC complex the CD8+ T cell is less likely to undergo AICD. However, we should also note that there are evidence in the literature suggesting that the expression of KIRs on CD8+ T cells might suppress their responsiveness [232]. Nevertheless, an extended -even if suppressed- response might result in an enhanced
Importantly, Ugolini et al. [79] proposed that inhibitory KIRs promote T cell survival by increasing the activation threshold of T cells. This may explain why the \textit{HLA-A*02} protective effect in HTLV-1 is not significantly enhanced by \textit{KIR2DL2}. A*02 molecules bind peptides significantly more strongly than other alleles (Appendix Figures B.1 and B.2) and the immunodominant HTLV-1 peptide Tax 11-19 is bound exceptionally strongly. Therefore, even if the T cell activation threshold were increased, the strength of signalling may remain above the threshold and consequently the \textit{A*02} protective effect cannot be enhanced.

Alternatively, KIR2DL2 in addition to other iKIRs expressed on NK cells can suppress their activation and as a result influence the CD8+ T cell mediated response (Figure 3.4(c)). Recent data, in MCMV infection, show that NK cells can negatively regulate the duration and effectiveness of virus-specific CD4+ and CD8+ T cell responses by limiting exposure of T cells to infected antigen-presenting cells [233]. In another murine study of LCMV infection, activated NK cells cytolytically eliminate activated CD4+ T cells that affect CD8 T-cell function and exhaustion [234]. Moreover, in a similar experiment, the absence of an inhibitory receptor on NK cells led to the lysis of activated but not naive CD8+ T cells in a perforin-dependent manner \textit{in vitro} and \textit{in vivo} [235]. Further evidence supporting NK tuning of T cell responses have been found in [236]. Hence, one can postulate that the expression of the strong inhibitory receptor, \textit{KIR2DL2}, on NK cells might limit NK cell activation allowing CD8+ T cells to have an enhanced response. Interestingly, NK cells have also been shown to kill activated T cells and this killing is reduced by inhibitory KIR [237,238].

Under both hypotheses, the HLA molecule whose protective/detrimental effects are enhanced is the HLA molecule that binds the TCR not the HLA that binds KIR2DL2. As a result, if the CD8+ T cell is restricted by a protective HLA class I molecule (e.g. \textit{B*57} in HCV or \textit{C*08} in HTLV-1) it will survive for longer than in a person who is \textit{KIR2DL2+} and its protective effects will be enhanced. Similarly, if the CD8+ T cell
is restricted by a detrimental HLA molecule (e.g. B\textsuperscript{*}54 in HTLV-1) it will also survive for longer and its detrimental effects will be enhanced.

Both these postulated mechanisms could explain 3 striking features of our observations: i) that KIR2DL2, a receptor typically associated with innate immunity, enhances HLA class I-associated immunity ii) that the effect cannot be explained by KIR2DL2 binding the enhanced HLA molecules directly and iii) that both protective and detrimental HLA effects are enhanced.
Figure 3.4: Potential mechanisms of KIR2DL2 impact on HLA class I-mediated immunity. The antiviral effector population which is enhanced by KIR2DL2 could be either NK cells (a) or T cells (b,c). KIR2DL2 can be expressed directly on NK cells or T cells or it might be expressed on NK cells but mediate the response of T cells.
3.12 Discussion

We show that KIR2DL2 enhanced several independent HLA class I-mediated effects in two unrelated viral infections. In HTLV-1 infection, KIR2DL2 enhanced the protective and detrimental effects of HLA-C*08 and B*54 respectively on disease status. KIR2DL2 also enhanced the association between C*08 and low proviral load in ACs and between B*54 and high proviral load in HAM/TSP patients. Additionally, KIR2DL2 enhanced the protective effect of HBZ binding by multiple HLA molecules. Strikingly, on stratifying by KIR2DL2, we observed, for the first time, a protective effect of C*08 on pvl in HAM/TSP patients and explained the lack of impact of B*54 on pvl in ACs. In HCV infection, KIR2DL2 enhanced the protective effect of B*57 on spontaneous clearance and the association between B*57 and low viral load in chronic carriers; a ‘dose effect’ with KIR2DL2 copy number was also observed. This progressive effect is consistent with reports of an association between KIR gene copy number and the frequency of cell-surface expression of the respective KIR molecule [239][240].

There are two mechanisms by which KIR2DL2 could act: it could enhance either NK-mediated or T cell-mediated immunity. That is, NK cell killing of virus-infected cells could be altered by KIR2DL2 expression or, alternatively, the virus-specific CD8+ T cell response could be modified by KIR2DL2 expression either directly on the CD8+ T cells or indirectly on the NK cells. However, our findings indicate that it is the T cell response that is more likely to be enhanced. First, strong binding of HBZ viral peptides via multiple different HLA-A and B molecules was associated with asymptomatic status [185] and this protective effect was enhanced by KIR2DL2. KIR2DL2 is not known to bind HLA-A or-B molecules (with the exception of B*4601 and B*7301) [76][78][223] so it is unlikely that the enhancement of the protective effect of binding HBZ by KIR2DL2 is due to direct binding between KIR2DL2 and HBZ peptide in the context of HLA-A and -B molecules. Furthermore, although NK cells exhibit peptide
dependence [21], it is hard to reconcile protein-specificity via multiple HLA molecules with an NK cell-mediated mechanism. Second, the KIR2DL2 enhancement could not be explained by binding between KIR2DL2 and any of the 3 HLA class I molecules investigated. Additionally, there is no KIR2DL2 main effect on disease outcome or viral burden for HCV or HTLV-1 infection. One further observation also suggests a T cell-mediated mechanism. Two protective genotypes in HCV infection that are postulated to operate via innate immune mechanisms [220] (namely a SNP upstream of IL28B and KIR2DL3-HLA-C1) had no impact on viral load in chronic infection [83,102]. The authors hypothesised that this was because innate barriers offer little protection once overcome. In contrast, the KIR2DL2/B*57 effect that we report here had a significant impact on viral load: again, this is perhaps more consistent with adaptive immunity.

Our results indicate that KIR2DL2 enhances HLA class I-restricted CD8+ T cell-mediated adaptive immunity. KIRs expressed on both NK cells and CD8+ T cells have been reported to shape adaptive immunity [81,82,218]. We postulate that, in the face of chronic antigen stimulation, protective T cells survive longer if they carry KIR2DL2 and therefore exert stronger protection. Likewise, T cells restricted by HLA alleles associated with increased disease susceptibility also survive for longer in the presence of KIR2DL2 and so are more detrimental. Hence, KIR2DL2 enhances both protective and detrimental HLA class I associations.

Alternatively, it is known that NK cells kill activated T cells and that this killing is reduced by inhibitory KIR [237,238]. Furthermore, there are recent evidence supporting NK tuning of T cell responses in murine studies of MCMV and LCMV infections [233,234,236]. So again, in this context, T cells restricted by protective and detrimental HLA class I molecules may survive longer in the presence of inhibitory KIR that suppress NK cell activity and thus their protective and detrimental effect would be enhanced.

Why does KIR2DL2 enhance T cell responses whereas the other inhibitory KIRs

129
apparently do not? The effect of KIR2DL2 may be most apparent because KIR2DL2 is present at informative frequencies and its C1 and C2 ligands are ubiquitous; i.e. unlike the other KIR every individual carries a KIR2DL2 ligand.

It will be important to determine whether inhibitory KIRs play a similar role in enhancing CD8+ and possibly CD4+ T cell-mediated immunity to other pathogens and in autoimmune disease. KIR-expressing virus-specific CD8+ T cells have been reported in other chronic infections including HIV-1, CMV and EBV [232, 242, 243]. Furthermore, in HIV-1 infection, high expression alleles of an inhibitory KIR, KIR3DL1, in the context of HLA-Bw4I have been associated with slow progression to AIDS [86]. In order to explain protection by an inhibitory KIR the authors proposed a model based on NK cell development. Our results suggest an alternative explanation, i.e. that KIR3DL1 enhances protective HLA-B-restricted responses to HIV-1 (B*57 is a Bw4I allele).

In contrast to previous studies of KIR genotype, which investigated the antiviral action of NK cells, we investigated the impact of KIRs on HLA class I-mediated antiviral immunity. We find a clear and consistent effect of KIR2DL2. The effect sizes are substantial: KIR2DL2 carriers with B*57 are 2 times more likely to clear HCV infection spontaneously; if they fail to clear the virus they have a viral load that is reduced by 4 logs. Until now, the advantages offered by inhibitory KIRs in virus infections have been unclear. Our data support an alternative role in which inhibitory KIRs enhance both beneficial and detrimental T cell-mediated immunity in persistent viral infection.
Chapter 4

Do KIRs affect CD8+ T cell selection pressure? - Study design

The study design presented here has been included in an MRC grant proposal submitted by Dr. Becca Asquith in 2012. The proposal received full funding and will be implemented by our group.

4.1 Aim

Our aim was to design a study that can help us understand whether KIRs can influence CD8+ T cell selection pressure as suggested by our findings in Chapter 3.

4.2 Introduction

In Chapter 3 we show that the KIR2DL2 enhancement was observed for multiple independent HLA class I molecules in both HCV and HTLV-1 infections (Figure 3.3). In HCV infection, KIR2DL2 enhanced the protective effect of B*57 on spontaneous clearance and also on viral burden in chronic carriers. Importantly, we also observed a progressive effect with increasing KIR2DL2 copy number. In HTLV-1 infection, KIR2DL2 enhanced the protective effect of HLA-C*08 on disease status and proviral load in ACs; and also enhanced the detrimental effect of B*54 and its association with a high proviral load in HAM/TSP patients. Furthermore, KIR2DL2 enhanced the protective effect of binding the HTLV-1 viral protein HBZ by multiple HLA-A and -B molecules.
Our results have a number of striking features. Firstly, \textit{KIR2DL2}-enhancement is not mediated by direct binding between KIR2DL2 and the enhanced HLA class I molecule. Although HLA-C*08, as a group C1 molecule, is expected to bind KIR2DL2, the most frequent subtype in our cohort (C*08:01, 88\%) binds KIR2DL2 very weakly (comparable to background \cite{223}) and HLA-B*54 and HLA-B*57 are not expected to bind KIR2DL2 \cite{70,223}; also, KIR2DL2 is not expected to bind the HLA-A and B molecules that bind the HTLV-1 viral protein HBZ although it enhances its protective effect. Secondly, \textit{KIR2DL2} does not have a direct impact on disease outcome or viral burden in any of the examined associations. Thirdly, both beneficial and detrimental HLA class I associations with infection outcome are enhanced.

Many associations between KIR-HLA interactions and infection outcome have been reported \cite{83,89,244}. In each case the KIR-HLA effect was linked to direct NK killing where NK lysis of virus-infected cells is modulated by inhibitory and activatory KIR-HLA signals (Figure 3.4(a)). Our observations are more consistent with a CD8+ T cell response that can be modulated directly or indirectly by inhibitory KIRs (Figures 3.4(b) and 3.4(c)). Inhibitory KIRs directly expressed on CD8+ T cells could increase their survival by upregulating pro-survival molecules like Bcl-2 and by reducing activation induced cell death (AICD) \cite{79,227,228}. Indirectly, it is known that NK cells help downmodulate the acute response by killing activated CD8+ T cells and that this killing is reduced by inhibitory KIRs expressed on NK cells \cite{234,237,245}. In both scenarios, CD8+ T cells restricted by ‘protective’ and ‘detrimental’ HLA class I molecules would be expected to survive longer in the presence of inhibitory KIRs and thus their effect would be enhanced. Hence, we postulate that \textit{KIR2DL2} consistently enhances the strength of the CD8+ T cell response in a beneficial or harmful manner depending on the HLA class I restriction.

As a natural next step, we want to investigate this postulated mechanism further. Specifically, we want test the hypothesis that the CD8+ T cell response is enhanced by \textit{KIR2DL2} by examining whether the CD8+ T cell response is stronger in \textit{KIR2DL2}+
individuals. We plan to use two independent methods: 1) viral sequencing to quantify selection pressure and 2) modelling to quantify dynamics. We define a ‘strong’ CD8+ T cell response as one that exerts a large anti-viral pressure (by lytic and/or non-lytic mechanisms). The selection pressure can be quantified by the presence of non-coding changes in viral epitope regions. Here, we present only the study design of the viral sequencing approach and its implications.

4.3 Description of study design

Our aim is to design a framework to quantify how HCV-specific T cell selection pressure is altered in the presence of KIR2DL2 and other inhibitory KIRs. Our study has two parts: 1) the amplification and massive parallel sequencing of targeted HCV genomic regions from a longitudinal cohort of acutely HCV-infected individuals and 2) the extensive bioinformatics analysis of the high-throughput data in order to quantify CD8+ T cell selection pressure.

Briefly, in the first part of the study (Figure 4.1) our collaborators (see Statement of Collaboration) will synthesise cDNA from viral RNA, followed by nested PCR amplification (PCR1 and PCR2 - 35 cycles each) using genotype-specific degenerate primers. Phusion high fidelity polymerase will be used for the amplification process. The PCR product which consists of 3 amplicons (Figure 4.2) will be cleaved and ligated to bridge primers using the Illumina Nextera transposon technology. This will be followed by sequencing on the Illumina HiSeq platform.

For the bioinformatics analysis, we will follow a pipeline of data processing (Figure 4.3) in order to perform thorough quality control of the viral sequences, reliable sequence alignment and quantification of CD8+ T cell selection pressure. We will quantify the selection pressure based on both experimentally-defined and predicted epitopes which are most likely to be presented by each individuals HLA-A and B alleles and determine the ratio of non-synonymous to synonymous changes (dn/ds)
in these peptides. The dn/ds ratio will be compared for the same peptide between KIR2DL2+ and KIR2DL2- HLA class I matched individuals with CD4+ T cell count as a covariate (as a control for HIV coinfection). In each step of the analysis the appropriate software (Figure 4.3) or customised scripts will be used.

To this day, in a pilot study our collaborators have successfully sequenced viral genomes from 11 HCV-infected individuals which we used in order to construct and partially implement a preliminary version of this pipeline. Specifically, we investigate the error sources involved in the study, the quality of the bases and the reads, the sequence alignment and the available coverage. The full dataset will be analysed by my successor.

![Experimental setup followed for obtaining the NGS HCV data](image)

Figure 4.1: Experimental setup followed for obtaining the NGS HCV data
Figure 4.2: The HCV regions covered by the three amplicons sequenced in this study.

Note: This figure is adapted by a schematic provided by Dr Heather Niederer.

Figure 4.3: Bioinformatics pipeline for post-processing the HCV NGS data
4.4 Data

4.4.1 Longitudinal HCV cohort

We have 34 patients co-infected with HCV and HIV monitored prior to treatment. HCV and HIV viral load measurements and total CD4+ and CD8+ T cell counts are available in the cohort. Most of the patients are infected with genotype 1a and few with genotype 1b. For each individual, we have 2-6 timepoints with HCV viral load >10,000 RNA copies/ml over a period of 6-12 months (Figure 4.4). In total, we have 161 samples which are both KIR2DL2 and HLA typed. We are currently in the process of recruiting 20 more patients. Based on this information we estimate we will have sufficient power (80%) to detect a small to moderate effect size \([221](d=0.28, a=0.05\) two-tailed).

4.4.2 Predicted viral peptides

We used the NetMHCPan epitope prediction algorithm and the H77 HCV genotype 1a sequence (Appendix A) in order to identify the most epitope dense regions of the HCV genome. The HCV proteins NS3, NS5A and NS5B had the most strong binding epitopes. For these regions, we obtained the top 5 strong binding peptides for each HLA-A and -B molecule of each individual in the cohort. We obtained in total 45 predicted peptides (Figure 4.5) which will be used in order to quantify CD8+ T cell selection pressure. In the list of peptides we also include the ones that differ at 1 position relevant to the consensus. All the predicted peptides will be considered with caution and studied in parallel to experimentally-defined epitopes available in the literature.
Figure 4.4: Longitudinal HCV viral load data from 34 patients coinfected with HCV and HIV. There are 22 KIR2DL2+ individuals and 12 KIR2DL2- individuals in the cohort. Each color-symbol combination represents the viral load measurements in RNA copies/ml (log10 scale) for each individual.
Figure 4.5: Predicted HCV epitopes for viral proteins NS3, NS5A and NS5B. The schematics are obtained using the software Geneious.
4.5 **NGS data post-processing**

In order to construct and assess a preliminary version of the bioinformatics pipeline, we use 11 sequenced samples (a single time point from 11 HCV-infected individuals). We focus mainly on the quality of the bases, the depth of coverage and the overall reads, the alignment of the sequences and the challenges involved in identifying genuine sequence variation in NGS data.

4.5.1 **Error sources**

When generating NGS data there are many processes involved that can generate errors in the sequenced data. These errors can be randomly distributed across the data (e.g. base calling errors) or they can be propagated systematically in the data analysis pipeline (e.g. PCR errors). In Figure 4.6 we present the main sources of error involved in our project. These are: 1) reverse transcription, 2) PCR amplification, 3) sequencing and 4) contamination. Stringent steps are taken in order to minimise these errors during the data generation. For reverse transcription and PCR amplification, the highest fidelity enzymes are used (superscript 3 and phusion high fidelity respectively). For base calling the Illumina protocol is applied and an extensive quality control is performed (see below). Finally for limiting contamination, the data are processed in batches and a cleaning step takes place between pre- and post- PCR product handling; also a water solution is used as a control.

All the above errors occur in processes taking place before the bioinformatics analysis and can of course hinder SNP identification. Further errors in the context of ‘variant calling’ can also be introduced downstream of the bioinformatics pipeline and specifically at the sequence alignment step; we address this issue separately later on in this chapter.
Figure 4.6: Errors arising during the NGS data generation.
4.5.2 Quality control

Firstly, we perform a quality control on the pilot data. We use the standalone software FastQC \(^1\) in order to examine the confidence that we should have in the base calling process. In all the 11 pilot datasets we observe that the quality of the base calling drops significantly as we approach the end of the reads (Figure 4.7(a)). Hence, we decide to apply two ‘cleaning’ procedure in order to improve the quality of the data: a) trim a part of the end of each read and b) set a requirement on the percentage of bases in each read that have a Phred quality score $ \geq Q20$ (i.e. $\leq 1$ error in 100 bps). For this, we use the standalone software FastX \(^2\). Here, we show the results for one sample but similar results are obtained for all 11 samples.

There is a trade-off between quality and quantity, since we want to achieve the best possible sequence quality whilst maintaining the maximum possible amount of data. When we trim the last 50 bps from each read (Figure 4.7(b)), the quality improves remarkably but we loose $\approx 33\%$ of the bps. When we set the percentage of bases with quality $ \geq Q20$ to i) 80%, ii) 90% and iii) 100% per read (Figure 4.8(a)-4.8(c)) we loose approximately 15%, 27% and 80% of the total bases respectively. Since the bulk of the ‘bad’ quality bps is found at the end of the reads, we decide to combine the two ‘cleaning’ filters by trimming the last 10 bps of each read and setting the percentage of bases with Phred quality score $ \geq Q20$ for each read to 80% (Figure 4.9). We estimate that the combination of the two base-trimming processes results in the loss of approximately 25% of the total sequenced bases which still allows for a high overall coverage.

Next, we looked at the nucleotide frequencies per bp across all reads. We find that overall the HCV sequence reads have a higher GC content ($\approx 60\%$) compared to AT content ($\approx 40\%$) as also reported in the literature \(^{246}\). We observe no other

\(^{1}\)http://www.bioinformatics.babraham.ac.uk/projects/fastqc
\(^{2}\)http://molecularevolution.org/software/genomics/fastx
Figure 4.7: Base calling quality control (Filter a). The unfiltered data (a) and the data after trimming the last 50 bps from each read (b). The trimming improves the quality remarkably but results in the loss of approximately 33% of the sequenced bps.

Biases for any specific nucleotide at any read position apart from the first 10 bps (Figure 4.10). This is an extended signature for sites of transposase-catalyzed adaptor insertion that, as we also verified, weakly resembles the insertion preference of the native Tn5 transposase (AGNTYWRANCT, where N is any nucleotide, R is A or G, W is A or T, and Y is C or T) [247].
Figure 4.8: Base calling quality control (Filter b). We set the percentage of bases with quality ≥ Q20 to i) 80% (a), ii) 90% (b) and iii) 100% (c). We lose approximately 15%, 27% and 80% of the total sequenced bps, respectively.
Figure 4.9: Optimal base calling quality filter. As an optimal approach, we combine the two ‘cleaning’ filters by trimming the last 10 bps of each read and setting the percentage of bases with Phred quality score $\geq Q20$ for each read to 80%. This approach leads to the loss of approximately 25% of the total sequenced bps.
4.5.3 Sequence alignment and coverage

Having improved the quality of the sequence data, our next aim is align them to a reference sequence in order to obtain the coverage per bp, i.e. the number of bases aligning to each position of the HCV amplicon that we sequence. We obtain the reference sequence by creating a consensus sequence from 78 available HCV 1a genotype sequences available in GenBank. However, we should note that when all the timepoints per patient are sequenced, all the timepoints will be aligned to the first timepoint of each patient. Only the first timepoint will be aligned to this consensus sequence; hence, the analysis will be performed on a different consensus per patient so that the identification of variants involves less uncertainty.

We perform the alignment using the Mosaik software. One of the most important parameters for the alignment configuration is the percentage of bases per read that can be mismatched to the consensus. This parameter should be carefully chosen since

\(^{3}\text{http://bioinformatics.bc.edu/marthlab/Mosaik}\)
if it too strict true variants would not be identified but if it is too large, the quality of the alignment drops. In order to obtain an estimate of this parameter we align the 78 HCV genotype 1a sequences and calculate the percentage of the main variant (the second most common nucleotide) at each position. We find that at each position of the HCV genotype 1a genomes and for the regions that we are investigating (Figure 4.2), the second commonest nucleotide is present at an average frequency of approximately 11% (Figure 4.11). In other words, at each position approximately 8/78 sequences have the same nucleotide to each other but different to the consensus. We use this value for the mismatch parameter. The algorithm applies a high gap penalty which we set at the default value since we want to limit the presence of indels that can mask the presence of true variants.

Next, we examined the depth of coverage that we could obtain using this high-throughput approach. As shown in Figures 4.12, the depth of coverage that was achieved varied across the three amplicons. For some samples all the amplicons had a high coverage (> 20000, Figure 4.12(a)) while in some other cases only one or two out of the three amplicons provided a high coverage (Figure 4.12(b)). In the 11 samples analysed, the third amplicon that covers the NS5B region was consistently having the highest coverage (A3: 11/11; Figure 4.13) and this deteriorated for the other two amplicons (A1: 8/11 and A2: 6/11 were amplified successfully). This trend implies that we will have a better overall coverage of the NS5B region and hence, we might be more powered to detect variants occurring in this part of the HCV genome. The failure to amplify some of the amplicons can be potentially explained by two main factors: 1) the high sequence variation even within HCV genotype 1a that required the use of ‘degenerate’ primers and 2) the lack of a complete cDNA extension over the whole 7kb region which might imply that amplicons A1 and A2 have less viruses represented in the cDNA than A3. However, despite several strategies applied by our collaborators in order to improve the amplification of amplicons A1 and A2 such as the repeat of cDNA synthesis and the use of different primers, none provided a substantial
Figure 4.11: Sequence variation of 78 HCV genotype 1a strains.
improvement. Although, this will limit the amount of viral variants that we can infer, it will still allow for enough coverage to test our hypothesis since we also have multiple timepoints per patient.
Figure 4.12: Coverage of HCV amplicons. For some patient-timepoint samples all the amplicons were amplified (a) while for some others only some of them were amplified (b). For the alignment shown here the MOSAIK software is used; the consensus sequence is obtained based on the 78 HCV genotype 1a strains available on GenBank and the mismatch parameter is set to 10% per read.
Figure 4.13: Coverage per HCV amplicon. For the alignment the MOSAIK software is used; the consensus sequence is obtained based on the 78 HCV genotype 1a strains available on GenBank and the mismatch parameter is set to 10% per read.
4.6 Future work

The main aim of this project is to investigate how KIRs, and in particular KIR2DL2, can shape HLA class I-mediated immune control in HCV infection. We propose to address this question by studying the difference in selection pressure exerted on epitope regions of the HCV genome between KIR2DL2+ and KIR2DL2- individuals who are HLA class I matched. The available data include a cohort of 34 HCV/HIV coinfected individuals with multiple high-throughput sequenced timepoints.

In a pilot study that we performed on 11 sequenced timepoints (not from the same individual) we identified error sources, explored the quality of the base calling, applied cleaning filters to improve this quality and also examined how we can obtain an optimal sequence alignment. However, other possibilities can be explored in order to further improve the alignment such as the use of an ambiguous reference sequence (based on the IUPAC) code (Appendix Figures C.1, C.2 and C.3) and the local realignment of areas of particular interest such as the epitope regions.

In order to study selection pressure, the next step is to be able to reliably identify SNPs present in the amplicons and study whether they result in coding changes or not. To do this we need to use several ‘variant calling’ algorithms in order to identify true variants and set a reasonable threshold for defining a SNP as true positive. The high coverage obtained should help in reducing the uncertainty involved in ‘variant’ and allow for detection of low frequency SNPs. However, it is important to note that this is highly associated to the primers we used and the main genotype studied - 1a. Therefore, if a patient is infected with genotypes other than 1a that differ significantly from it, we will not not able to capture this variation. Once variants have been detected, the selection pressure in both predicted and experimentally defined HCV epitopes can be quantified using the ‘dn/ds’ ratio.

Additionally, using the available longitudinal cohort and the model described in
[248], we can examine the viral escape dynamics in \textit{KIR2DL2+} and \textit{KIR2DL2-} individuals. Of course, the next-generation sequencing of this longitudinal HCV/HIV coinfection cohort provides a unique set of data that can help in exploring hypotheses not only involving KIR- and HLA-mediated immunity but also other important immunological questions.
Chapter 5

Lytic and non-lytic CD8+ T cell responses

The analysis and findings presented in this chapter have been published in Elemans M., Seich al Basatena N.-K. et al., PLoS Comput. Biol. 7(9), 2011 [3].

5.1 Aim

Our aim was to investigate whether the experimental data presented in [42] best supports a lytic or non-lytic mechanism of CD8+ T cell control.

5.2 Introduction

There are many factors that suppress viral load in HIV/SIV infection: 1) CD8+ control of infected cells, 2) virus cytopathicity, 3) innate immunity, 4) target cell limitation, 5) antibody responses and others. Here, we focus on the CD8+ cellular arm which is initiated once the peptide-major histocompatibility complex molecules (pMHC) displayed on the surface of cells presenting antigen are recognised by the T cell receptor (TCR).

In SIV infection, one of the most direct and strong evidence of the importance of CD8+ T cells in controlling SIV infection in vivo is the observation that, on depleting CD8+ T cells and for both acute and chronic infection, SIV-1 viral load increases by 0.5 to 1 log [41][66][249].
There are two main CD8+ T cell mechanisms that can mediate viral growth in HIV/SIV infection [250]: (i) direct, lytic killing and (ii) indirect, viral suppression by secretion of soluble factors. For direct killing, the immunological synapse (IS) that forms at the contact site of the T cell and the infected target facilitates the transportation of lytic granules (which can contain the apoptotic protein called perforin and granzymes) to target cells [251]. For indirect killing, production of antiviral soluble factors such cytokines and chemokines (e.g. IFN-γ and RANTES) can potentially prevent infection of uninfected cells by free virus [56] or decrease production of viral particles by infected cells [58, 252, 253]. Interestingly, in [254] using a high-throughput single-cell assay, the authors found that the majority of antigen-stimulated CD8+ T cells that secrete IFN-γ did not exhibit cytotoxic responses, indicating that lytic and non-lytic effector functions might be independently regulated. In addition, in [255], using multiple lytic and non-lytic markers of CD8+ T cell activity it is shown that these responses can be functionally segregated in vivo. Hence, it can be suggested that CD8+ T cells might choose effector function depending on viral agent, site of infection and pathogenicity [256].

Recently, ground-breaking studies reported that following CD8+ cell depletion in SIV-infected macaques, viral load robustly increased, however the lifespan of SIV-infected cells was unaltered [42, 167]. The two groups used ART to block further rounds of infection to study the turnover of SIV-infected cells. They reported that when the lifespan of productively infected cells was measured there was no detectable difference between control macaques with an intact CD8+ T cell response and CD8+ T cell-depleted macaques. This unexpected result led to the suggestion that SIV might be controlled primarily via non-lytic mechanisms.

The aim of this project was to explore further, using an extensive set of ODE models that describe both lytic and non-lytic effector function, whether the experimental data presented in [42] best supports a lytic or non-lytic mechanism of CD8+ T cell control.
5.3 Methods

5.3.1 Experimental data

The experimental data were produced by collaborators for the study published in [42] and are briefly described below.

Ten rhesus macaques infected with SIV$_{mac239}$ were divided into two groups. Group A consisted of 5 CD8$^+$ T cell-depleted animals during early chronic phase. Group B included 5 CD8$^+$ T cell-depleted animals during the late chronic phase. Antiretroviral therapy (ART) was administered to all animals in both phases. To deplete CD8$^+$ T cells, the OKT8F was given for 3 consecutive days (Group A, days 5860 after infection; Group B, days 177179). ART (PMPA and FTC) was given for 28 consecutive days during both early and late chronic infection (starting at d63 and d168 for group A, and d63 and d182 for group B). Plasma viraemia was quantified by real-time reverse-transcriptase PCR. CD8$^+$ and CD4$^+$ T cell counts were measured using multicolor flow cytometric analysis. For almost all points for which viral load was recorded, CD4$^+$ and CD8$^+$ T cells measurements were available. These studies were approved by the Emory University and University of Pennsylvania Institutional Animal Care and Use Committees. All this experimental work was performed by our collaborators and further details can be found in [42].
5.3.2 Lytic models of infection

Classic model

The basic lytic model of HIV infection describes the dynamics of uninfected target cells \( T \), productively infected cells \( T^* \) and free virus \( V \).

\[
\dot{T} = \lambda - \beta TV - \delta_T T \quad (5.1a)
\]
\[
\dot{T^*} = \beta TV - \delta_I T^* - \kappa ET^* \quad (5.1b)
\]
\[
\dot{V} = pT^* - cV \quad (5.1c)
\]

Where \( \lambda \) is the influx of uninfected CD4+ cells \( \text{cells d}^{-1} \), \( \beta \) is the infection rate, \( \delta_T \) and \( \delta_I \) are the death rate of uninfected and infected CD4 cells respectively, \( \kappa \) is the CD8+ killing rate of infected CD4+ T cells, \( p \) is the production rate of free virions and \( c \) is the clearance rate of free virions (all measured per day). In all the models presented here, the fraction of CD8+ T cells, \( E \), is given by the empirical function calculated by a linear interpolation between time points.

We also considered a model with two populations of productively infected cells (e.g. CD4+ T cells and macrophages), \( T^* \) and \( M^* \) that follow different death rates, \( \delta_I \) and \( \delta_M \) respectively.

\[
\dot{T} = \lambda - \beta TV - \delta_T T \quad (5.2a)
\]
\[
\dot{T^*} = \beta TV - \delta_I T^* - \kappa ET^* \quad (5.2b)
\]
\[
\dot{M} = \lambda - \beta MV - \delta_T M \quad (5.2c)
\]
\[
\dot{M^*} = \beta MV - \delta_M M^* - \kappa EM^* \quad (5.2d)
\]
\[
\dot{V} = p_T T^* + p_M M^* - cV \quad (5.2e)
\]

Early killing model

In [257] an alternative lytic model is described where CD8+ T cells limit their killing to the phase prior to viral production. Two populations of uninfected cells are included
in the model; recently infected cells which are susceptible to CD8+ T cell killing, $I^*$ and productively infected cells that evade CD8+ T cell killing via MHC class I downregulation, $P^*$. Uninfected CD4+ T cells ($T$) are still described by [5.1a] but the rest of the dynamics are altered as follows:

\[ \dot{I}^* = \beta TV - \gamma I^* - \delta_I I^* - \kappa EI^* \quad (5.3a) \]
\[ \dot{P}^* = \gamma I^* - \delta_p P^* \quad (5.3b) \]
\[ \dot{V} = pP^* - cV \quad (5.3c) \]

Where $\gamma$ is the transition rate from $I^*$ to the $P^*$ population of infected cells and $\delta_p$ is the death rate of the $P^*$ population ($d^{-1}$).

**Late killing model**

Another known lytic model described in [258] suggests that infected cells produce only few virions early in infection and the majority of virions is produced at a later stage just before cell death. This cytopathic effect of the viral infection can be captured by two populations of infected cells; recently infected cells that do not produce virions and die at a negligible rate, $L^*$, and productively infected cells, $A^*$. The dynamics are given by:

\[ \dot{L}^* = \beta TV - \gamma L^* - \tau L^* - \delta_I L^* \quad (5.4a) \]
\[ \dot{A}^* = \tau L^* - \delta_A A^* - \kappa EA^* \quad (5.4b) \]
\[ \dot{V} = pA^* - cV \quad (5.4c) \]

Where $\tau$ is the transition rate from $L^*$ to the $A^*$ population of infected cells and $\delta_I$ and $\delta_A$ are the death rates of the $L^*$ and $A^*$ populations (both $d^{-1}$) respectively.
5.3.3 Non-lytic models of infection

Blocking infection model

The non-lytic models are similar to an extent with the lytic models regarding the dynamics of the uninfected and infected cells and the viral production but do not include direct CD8+ T cell killing. Instead, CD8+ T cells decrease the infection rate, $\beta$, or the virion production rate, $p$, by a fraction $u$:

$$\frac{1}{1 + \eta E(t)}$$  \hspace{1cm} (5.5)

The dynamics of a non-lytic model where viral infection is decreased can be described by:

$$\dot{T} = \lambda - \left(\frac{1}{1 + \eta E}\right) \beta TV - \delta_T T$$  \hspace{1cm} (5.6a)

$$\dot{T}^* = \left(\frac{1}{1 + \eta E}\right) \beta TV - \delta_I T^*$$  \hspace{1cm} (5.6b)

$$\dot{V} = pT^* - cV$$  \hspace{1cm} (5.6c)

We also fit a biphasic non-lytic infection model, similar to Eqs. \[\text{[5.2]}\]

$$\dot{T} = \lambda_T - \left(\frac{1}{1 + \eta_N E}\right) \beta TV - \delta_T T$$  \hspace{1cm} (5.7a)

$$\dot{T}^* = \left(\frac{1}{1 + \eta_N E}\right) \beta TV - \delta_I T^*$$  \hspace{1cm} (5.7b)

$$\dot{M} = \lambda_M - \left(\frac{1}{1 + \eta_N E}\right) \beta MV - \delta_T M$$  \hspace{1cm} (5.7c)

$$\dot{M}^* = \left(\frac{1}{1 + \eta_N E}\right) \beta MV - \delta_M M^*$$  \hspace{1cm} (5.7d)

$$\dot{V} = p_T T^* + p_M M^* - cV$$  \hspace{1cm} (5.7e)
Blocking production model

The dynamics of a non-lytic model where viral production is decreased can be described by:

\[
\dot{T} = \lambda - \beta TV - \delta_T T \\
\dot{T}^* = \beta TV - \delta_T T^* \tag{5.8a}
\]

\[
\dot{V} = \left(\frac{1}{1 + \eta E}\right) pT^* - cV \tag{5.8c}
\]

We also fit a biphasic non-lytic production model:

\[
\dot{T} = \lambda_T - \beta TV - \delta_T T \\
\dot{T}^* = \beta TV - \delta_T T^* \tag{5.9a}
\]

\[
\dot{M} = \lambda_M - \beta MV - \delta_T M \tag{5.9c}
\]

\[
\dot{M}^* = \beta MV - \delta_S M^* \tag{5.9d}
\]

\[
\dot{V} = \left(\frac{1}{1 + \eta E}\right) (p_T T^* + p_M M^*) - cV \tag{5.9e}
\]

5.3.4 Model fitting and selection

Model fitting The four different lytic and non-lytic models were fitted to data of virus load and total CD4+ T cell population, scaled to obtain equal mean. E(t) was obtained by linear interpolation of experimental data. ART-treatment was simulated by setting infection rate \(\beta = 0\). To fit the models to the data we used the pseudorandom algorithm in the modFit-function of the FME-package in the statistical software R.

Model selection The small sample (second-order) bias-adjusted Akaike Information Criterion (AICc) [259,260] was used to compare the fit of the models. AICc adjusts for differences in number of parameters [5.10]. The model with the lowest AICc is considered to describe the experimental data best and the comparison between the models is defined by \(\Delta_i\) (Equation 5.11). As a rule of thumb, we assumed
considerable support for models with an AICc within two of the lowest; if models differ by three to seven AICc units from the minimum AICc there is some support for the model with the higher AICc while a difference larger than 10 suggests that the model is very unlikely to describe the underlying data [260]. Note: This part of the study was performed by M. Elemans.

\[
AICc = N\ln\left(\frac{SSR}{N}\right) + 2K + \frac{2K(K + 1)}{N - K - 1}
\]  \hspace{1cm} (5.10)

where \(N\) is the data sample size, \(K\) is the number of parameters and SSR is the residual sum of squares of the fitted model.

\[
\Delta_i = AICc_i - \min(AICc)
\]  \hspace{1cm} (5.11)

where \(AICc_i\) is the AICc for model \(i\) and \(\min(AICc)\) is the minimum AICc value of all the models fitted to the data.

### 5.4 Model comparison

In summary, we fitted 4 lytic and 4 non-lytic models given in Table [5.1]

Although the quality of the fits is rather poor (Figure [5.1]), our results show clear and consistent support for the non-lytic model in which CD8+ T cells reduce infection; none of the lytic models receive any support for most data sets. The best-fitting non-lytic model (non-lytic model ii, in which non-lytic factors reduced new infection events and there are two populations of productively infected cells) was compared with each of the lytic models in turn. Furthermore the non-lytic model in which infection was reduced performed consistently better than the non-lytic model in which virion production was reduced though the differences in performance were not as large as for the comparison between lytic and non-lytic models (Table [5.2] and Figure [5.2]).

The reason why the lytic models fail can be seen from studying the equations. In the lytic models the rate of post-ART decline in viral load is determined by infected cell
### Table 5.1: Models fitted to experimental data.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Description</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lytic i (Li)</td>
<td>Classic model</td>
<td>5.1</td>
</tr>
<tr>
<td>Lytic ii (Lii)</td>
<td>Biphasic classic model</td>
<td>5.2</td>
</tr>
<tr>
<td>Lytic iii (Liii)</td>
<td>Early killing model</td>
<td>5.3</td>
</tr>
<tr>
<td>Lytic iv (Liv)</td>
<td>Late killing model</td>
<td>5.4</td>
</tr>
<tr>
<td>Non-Lytic i (NLi)</td>
<td>Blocking infection model</td>
<td>5.6</td>
</tr>
<tr>
<td>Non-Lytic ii (NLii)</td>
<td>Biphasic blocking infection model</td>
<td>5.7</td>
</tr>
<tr>
<td>Non-Lytic iii (NLiii)</td>
<td>Blocking production model</td>
<td>5.8</td>
</tr>
<tr>
<td>Non-Lytic iv (NLiv)</td>
<td>Biphasic blocking production model</td>
<td>5.9</td>
</tr>
</tbody>
</table>

The best fitting model, the biphasic non-lytic model where infection of new cells is blocked by non-lytic factors, was compared with each of the lytic models in turn as well as with the biphasic non-lytic model where viral production is blocked. The AICc consistently provided support for the non-lytic model of blocking infection. All the models fitted are summarised in Table 5.1. All the tests were performed using a paired two-tailed Mann-Whitney test.

### Table 5.2: The AICc mean difference and p-value

<table>
<thead>
<tr>
<th>Models</th>
<th>AICc mean difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLii vs Li</td>
<td>15</td>
<td>0.043</td>
</tr>
<tr>
<td>NLii vs Lii</td>
<td>26</td>
<td>0.028</td>
</tr>
<tr>
<td>NLii vs Liii</td>
<td>40</td>
<td>0.018</td>
</tr>
<tr>
<td>NLii vs Liv</td>
<td>39</td>
<td>0.018</td>
</tr>
<tr>
<td>NLii vs NLiii</td>
<td>10</td>
<td>0.018</td>
</tr>
</tbody>
</table>

161
death due to viral toxicity and CD8+ T cell killing. Under the CD8+ T cell depletion regime post-ART decline in viral load is solely determined by infected cell death due to viral toxicity. Thus, to predict the similar post-ART decline under the two treatment-regimes lytic models need to attribute a small role to CD8+ cells. This small role of CD8+ T cells is poorly compatible with the increase in viral load following CD8+ T cell depletion in the absence of ART. Therefore, lytic models cannot accurately fit both post-ART decline and the increase in viral load upon depletion whereas non-lytic models can. Consequently models with a non-lytic component are likely to consistently outperform similar models with a lytic component.

We conclude that although self-consistent hypotheses can be constructed in which CD8+ T cells exert their antiviral effects by lysis without a detectable impact on infected cell lifespan, these models are poorly predictive and a non-lytic model provides a better explanation of the viral load and CD4+ T cell dynamics.
Figure 5.1: Experimental viral load and percentage CD4+ T cells (filled dots on left column and right column respectively) and the best fitting non-lytic model (non-lytic model ii; solid lines) and the best-fitting lytic model (lytic model i; dashed lines). Number of parameters: 10 and 8 respectively. Model selection is based on AICc which controls for the number of parameters in the model. More fits can be found in the supplementary material of [3]. For animals in Group B, data for CD4+ and CD8+ T cells was only available from 45 days after infection, hence viral load data before this time point have not been fitted. **Note:** This figure was produced by M. Elemans.
Figure 5.2: The AICc of the model minus the AICs of the best fitting model for that animal is shown. A large difference represents a poor fit. Overall the model in which CD8+ T cells operate via a non-lytic mechanism which reduces infection (non-lytic model ii) offers the best fit in 6 out of 7 cases, a lytic model only offers the best fit in 1/7 cases. The 4 lytic models are i) the basic model ii) an extension of the basic model to include two populations of productively infected cells iii) a model following Klenerman et al. [258] in which SIV is cytopathic and iv) a model following Althaus et al. [257] in which CD8+ T cell killing is limited to the early non-productive stage of the viral lifecycle. The 4 non-lytic control models were: i) a model in which non-lytic factors reduced new infection events ii) an extension of model i to include two populations of productively infected cells iii) a model in which non-lytic factors reduce virion production iv) an extension of model iii to include two populations of productively infected cells. The values are given in Appendix Figure D.1. Note: This figure was produced by M. Elemans.
5.5 Discussion

SIV-infection of rhesus macaques is one of the most widely used animal models of HIV-1 infection. Therefore, the observation that the increase of viral load that followed CD8+ T cell depletion in SIV-infected macaques was not ‘accompanied’ by an increase of infected cell lifespan [42,167] raises important questions not only in SIV but also in HIV-1 infection. This observation is reminiscent of the findings of the Klenerman et al. study which found that CD8+ T cell-mediated lysis can reduce viral load but has small effects on the half-life of infected cells. [258].

A number of hypotheses can potentially explain this striking observation of the Klatt et al. and Wong et al. findings. Firstly, the CD8+ T cell control might be primarily non-lytic. Secondly, the CD8+ T cell control may be lytic but ART treatment impairs CD8+ T cell function. Thirdly, CD8+ T cell lytic killing might occur early and prior to viral production [257] or late and just before the cell would die anyway [258]. Fourthly, the CD8+ T cell control is lytic but the measurements of infected cell lifespan are accurate enough to discern a difference.

Our study finds evidence that a non-lytic mechanism, especially via blocking infection, explains best the SIV dynamics during acute and chronic infection across all animals studied. The models include a low number of parameters and require the fitting of both viral load and CD4+ T cell count data. As a consequence, the divergence between prediction and observation (particularly for CD4+ T cell count) was large. However, the model comparison presented here is based on the AIC and selects for the model that approaches better the underlying ‘true’ dynamics amongst all models fitted [259].

In [3], the rest of the above hypotheses are also explored. The authors find that ART treatment can impair CD8+ T cell function but the difference between control and CD8+ T cell depleted animals remains small. In addition, they find that both
the early and late CD8+ T cell killing models provide qualitative consistent results but only for relatively narrow model parameter ranges. Interestingly, the authors find that the limited accuracy of the data could offer a plausible explanation for the lack of difference.

Hence, taking all of the above observations together, we conclude that although the lack of an effect on the lifespan of infected cells found in [42, 167] cannot exclude a lytic CD8+ T cell effector function, the models of non-lytic control provided better overall predictions for the time course of infection and a potential reason why no difference in lifespan was detected between control and depleted animals. In [261], the authors use a biologically-directed modelling approach to study the Klatt et al. data. Their study involves a high number of biological parameters (which make the model hard to track) and suggests that aspects of the HIV lifecycle such as the eclipse phase and an exponential pattern of virion production can explain the lack of difference in infected cell lifespan between CD8+ T cell depleted and control SIV-infected macaques. However, in the models studied here we have considered such properties for the lytic effector function but the non-lytic response could still better explain the data although we could not exclude a lytic mechanism. It is after all likely that these two effector mechanisms act in parallel. Importantly, the findings in [3] shows that there is a clear need for more accurate experimental measurements before we can draw more robust conclusions regarding the mechanism of CD8+ T cell control in HIV/SIV infection.
Chapter 6

A Cellular Automaton model of CD8+ T cell responses

6.1 Aim

Our aim was to develop and implement a Cellular Automaton model that captures both temporal and spatial aspects of HIV/SIV dynamics and can be used to model both lytic and non-lytic CD8+ T cell responses.

6.2 Introduction

In order to study HIV/SIV dynamics including CD8+ T cell viral suppression and its effect on viral escape, we develop and implement a 3D Cellular Automaton (CA) model. A CA is an individual-based computer simulation of a system that evolves in time on a multidimensional (usually 2D or 3D) lattice of nodes. A CA model 1) simulates both spatial and temporal characteristics of the system, 2) can be used to study both the individual as well as the population behaviour and 3) can be easily redefined to study different aspects of the underlying biological processes. Agent-based models (ABMs) - which include CA models - have been used to study HIV infection in multiple studies [262, 264]. However, they can be computationally demanding and require a substantial knowledge of the underlying biological principles. A more detailed description of the successes and challenges of agent-based models is given in [265].

The CA model that we developed incorporates rules on cell motility, the reproduc-
tion and death of infected cells, the eclipse phase of viral production, CD8+ T cell killing and conjugate formation, as well as the lifespan and influx of uninfected cells. The timescale of the model can integrate fast and slow processes and can therefore be used to simulate the course of the infection during both acute and chronic phases. Using this approach, we can produce in silico experiments of viral escape dynamics under both lytic and non-lytic CD8+ T cell effector functions.

6.3 Model description

We use a three-dimensional Cellular Automaton (CA) to simulate the spread of HIV/SIV infection in a small portion of the spleen (roughly 0.05 – 0.5%). The CA (see Figure 6.1) is a lattice which consists of nodes and edges and is updated at every timestep (30 sec). Each node of the lattice represents a cell or part of the cell and has 26 neighbours. The grid is governed by toroidal boundary conditions where a cell leaving one side of the lattice reappears on the opposite side.

The cell populations included in the model are given in Figure 6.2. CD4+ T cells can be uninfected, infected with the wild-type strain or infected with a variant strain of the virus. The CD8+ T cell population that we explicitly model represents a population of all CD8+ T cell clones which are specific for a single epitope. The model also includes macrophages, the reticular network of the spleen and a generic splenocyte population.

Most cells -apart from macrophages (MΦs) that occupy four nodes- are represented by one node on the lattice; this is to reflect the difference between the diameter of T cell which is 7µm \[266\] and that of macrophages which is calculated to be 10 – 16µm and has been implemented as such in \[267\]. An additional set of nodes represents the Reticular Network (RN), a rigid cellular structure, found in the spleen. These nodes are immobile and act as spatial obstacles to the overall movement of cells. Only a small percentage of nodes is considered to be unoccupied since the spleen is
a dense organ ($\approx 1\%$). The nodes that are left after setting the known frequencies of the specific cell populations within biological ranges (see Appendix Tables E.1 and E.2) are considered to be unspecified splenocytes. The initialisation values of all the populations included in the simulations are given in Appendix Table E.1. The cellular automaton is implemented (by the author of this thesis) in C++.

Figure 6.1: A snapshot of the 3D cellular automaton model. The different coloured nodes represent different cell populations considered in the model (see Figure 6.2).

Figure 6.2: Cell populations considered in the 3D cellular automaton model.
6.4 Model assumptions

Modelling the dynamics of HIV/SIV infection demands the incorporation of many complex rules. However, focusing on the specific questions that we are aiming to address and in order to speed-up the simulations and ease the interpretation of our results, we allowed for multiple assumptions to be included in the model as long as these would not alter the outcome of the tested hypotheses. The following assumptions were made (the exact parameter values used are discussed in the relevant sections):

1. The same motility parameters apply for all cell types since we mostly focus on T cells for which the parameters have been experimentally defined in multiple studies [268–273].

2. No free virus is included in the model. Viral spread occurs via an increased probability of infection of neighbouring uninfected cells. This probability is set based on the HIV viral reproduction rate which has been estimated experimentally [274].

3. We do not make any distinction between activated and non-activated CD4+ T cells. However, there is evidence in the literature that activated CD4+ T cells can be the main targets of HIV-1 [275].

4. We only model a single HIV/SIV CD8+ T cell response specific for a single viral epitope and which is present at a magnitude reported for the chronic phase of infection. Because we focus on the dynamics of infection after set point viral load we do not consider proliferation or contraction of this CD8+ T population. Although epitope escape can lead to decreased antigen stimulation of the specific CD8+ T cell population and consequently to the decline of the response, this is expected to happen with a lag of a few months [276]. Additionally, the estimates of CD8+ T cell decline range from: $0.0002 - 0.015 d^{-1}$ [277–280] which
correspond to an average specific CD8+ T cell lifespan of 200 days. In the latter studies, the rate of specific CD8+ T cell decline is estimated following HAART when loss of antigenic stimulus is much greater than following escape. Hence, these estimates of the rate of specific CD8+ T cell decline might be an overestimate of the rate following viral escape on which we focus here. The above observations taken together strongly suggest that during the timescale that we simulate viral escape (<3 months) the assumption of an approximately constant epitope-specific CD8+ T cell population should be valid.

5. There is no fitness cost associated with the escape mutations.

6. We do not consider superinfection of infected cells. The infected CD4+ T cells are either infected with the wild-type or the variant strain. Findings regarding superinfection remain controversial in the literature [281, 282].

6.5 T cell motility

The use of two-photon microscopy has allowed major advancements in the \textit{in vivo} study of the motility patterns of cells. Quantities that characterise motility such as speed, motility coefficient and displacement have been calculated experimentally for different cell types and under different experimental settings in multiple studies [268–273]. In summary, the average speed of non-activated T cells is reported to be 10 – 15 \( \mu m/min \) with a maximum of 25 \( \mu m/min \) [268, 269], the motility coefficient is 50 – 100 \( \mu m^2/min \) and the displacement is calculated to be proportional to the square root of time \( (d \propto \sqrt{t}) \). Most of these measurements have been in the lymph nodes; for the spleen, they are suggested to be comparable but lower [283, 284]. Interestingly, T cell motility is argued to resemble a random; in [285], it is shown that this resemblance can be dictated by the lymph node environment rather than an intrinsic motility program.
To simulate the motility of cells with the CA model we implement the rules of cell motility used in [267] (see Appendix E.1 for rule description). The two main parameters which calibrate the motility of the cells are the speed and the motility coefficient. In our model simulations, the average T cell speed is approximately 9 \( \mu m/min \) with a maximum of 25 \( \mu m/min \) and the average T cell motility coefficient is 75 \( \mu m^2/min \) (Figure 6.3). Additionally, we indeed find that under this motility rules, the displacement is proportional to the square root of time (Figure 6.4) suggesting a ‘random walk’ behaviour. As expected, the mean CD8+ T cell speed is reduced (Figure 6.5) because of conjugate formation when infected cells are introduced in the model; this is also consistent with experimental observations [283, 286].

Figure 6.3: The speed of four individual simulated CD8+ T cells in the CA model is shown. The mean speed is approximately 9 \( \mu m/min \) and the mean motility coefficient is 75 \( \mu m^2/min \). These values agree with the experimental measurements reported in the literature [268, 273].
Figure 6.4: The mean displacement of CD8+ T cells with respect to $\sqrt{\text{time}}$ is shown. The straight line suggests that the CD8+ T cell movement resembles a random walk.
Figure 6.5: The mean CD8+ T cell speed decreases as the ‘probability of successful scan’, i.e. recognition followed by CD8+ T cell activation, of infected targets increases. Each boxplot represents the results obtained by the simulation of 50 epitope-specific CD8+ T cells. Abbreviations: No Rec=CD8+ T cells do not recognise any infected target that they meet, All Rec=CD8+ T cells recognise all infected targets that they meet; the rest represent intermediate cases.
6.6 CD4+ T cell influx

In order to obtain a sustained infection we also need to simulate an influx of CD4+ T cells. This influx can include both proliferation and inflow of cells to the spleen but since we are modelling a closed system we only focus on proliferation. To define the probability of additional CD4+ T cell appearing at the site of infection (here being the spleen) we use the estimates of CD4+ T cell proliferation rates reported in [287]. In the latter study deuterated glucose is used to label DNA of proliferating cells in 4 healthy and 7 treatment-naive HIV+ subjects. By applying a mathematical model the authors estimate that the mean CD4+ proliferation rate in HIV+ subjects is 0.025 $d^{-1}$ when their mean CD4+ T cell count is approximately 400 $\mu l^{-1}$, i.e. about 33% of a normal count. As an estimate of the influx probability of CD4+ T cells appearing in the grid we use a Hill function (Equation 6.1 and Figure 6.6) where $j$ is calculated such that the probability of entering the grid, $p_{influx}$ when the CD4+ T cell count is decreased by 67% provides a CD4+ T cell proliferation rate of 0.025 $d^{-1}$ and $n$ is chosen to provide a plausible change of $p_{influx}$ with respect to the percentage of CD4+ T cells, $U$, at any given timepoint. We note that $p_{influx} \to 0$ when CD4+ T cell count goes to normal, i.e. 100% of the initial population. This formulation simulates a homeostatic influx mechanism for the replenishment of CD4+ T cells which allows for a sustained viral infection in the CA model.

$$p_{influx} = \frac{U^n}{U^n + j^n} \quad (6.1)$$

The CD4+ T cells entering the grid with a probability, $p_{influx}$, can be uninfected or infected (either with the wild-type or the variant strain). Equations 6.2-6.4 describe the set of probabilities which define whether the CD4+ T cell that entered the grid will be uninfected, $p_u$ or infected (with the wild-type, $p_w$, or the variant, $p_v$ strain). These probabilities depend on the current population of uninfected, $N_u$, wild-type infected, $N_w$ and variant infected, $N_v$, cells present on the grid as well as their respective
Figure 6.6: The influx of CD4+ T cells is modelled based on a Hill function (Equation 6.1) of order $n = -5$. This type of influx can be described as homeostatic.
lifespans, $L_u, L_w$ and $L_v$. We consider $L_w = L_v$.

$$p_w = \frac{N_u L_w}{N_u L_u} p_u$$  \hspace{1cm} (6.2)

$$p_v = \frac{N_v L_v}{N_u L_u} p_u$$  \hspace{1cm} (6.3)

$$p_u = \frac{N_u L_u}{N_u L_u + N_w L_w + N_v L_v}$$  \hspace{1cm} (6.4)

This fraction of newly introduced infected cells can represent reservoirs of latently infected CD4 T cells which can be important in maintaining active infection [288,289]. Additionally, it can reflect the small proportion of circulating infected CD4+ T cells that enter the spleen at any given time [290].

### 6.7 CD4+ T cell reproduction and death

#### 6.7.1 Reproductive rate

In the context of within host viral infection, the basic reproductive ratio, $R_0$, is defined as the number of secondary infected cells produced by a primary infected cell during its lifetime; this definition assumes that uninfected target cells are not limiting. In this model we consider $R_0 = 6$ which agrees with estimates reported in the literature during acute HIV infection [274,291] and therefore before CD8+ T cell responses take effect. At this viral reproduction rate, we record a peak in infected cell population 15 – 20 days post infection (dpi).

#### 6.7.2 Death rate

Since we are mostly interested in the dynamics of CD4+ T cells (uninfected and infected) during chronic HIV+ infection and specifically in the comparison of lytic and non-lytic responses, we only allow for reproduction and death of CD4+ T cells in the model. The rest of the cell populations remain of constant size during the course of the infection.
For uninfected CD4+ T cells, we set the death rate based on the proliferation rate estimated in \[287\]. As shown in \[292\] the death rate measured in \[287\] is not that of all CD4+ cells but only of labelled cells, which are not a representative sample of the whole. However, the proliferation rate measured is the proliferation rate of the whole CD4+ population. Hence, because at equilibrium the proliferation equals the death rate and since we are mainly interested in the chronic phase of the infection, we set the death rate of uninfected CD4+ T cells to 0.025 $d^{-1}$ which equates to a lifespan of 40 days.

For infected CD4+ T cells, we consider two phases: 1) An eclipse phase which represents a viral delay of 1 day \[293, 294\] and 2) a productively infected phase. During the eclipse phase, infected cells can neither be recognised by CD8+ T cells nor infect uninfected CD4+ T cells and die with the same rate as uninfected cells. In the productively infected phase, they die at an exponential rate with a mean of 1 $d^{-1}$ (i.e. they have a mean lifespan of 1 day) as estimated in \[295\] (Figure 6.7). In the case where cytotoxic CD8+ T lymphocytes have to search for infected target cells to deliver their lethal hit, a constant rate of death and, therefore, exponentially distributed lifespans of the cells may serve as appropriate descriptions of the virus dynamics \[296]\.

In summary, infected CD4+ T cells die at a rate of 0.025 $d^{-1}$ during the eclipse phase and at a rate of 1 $d^{-1}$ at the viral productive phase. The latter death rate estimates refer to total death which includes both CD8+ T cell control and viral cytotopathicity.

Focusing on CD8+ T cell control, we assume that CD8+ T cells as part of the second line of defence do not immediately start to cope with the infection. We set the time to CD8+ T cell appearance to 10 days. The delay in CD8+ T cell responses is affected by multiple factors such as the presentation of epitopes to naive CD8+ T cells in local lymph, the time needed for clonal expansion and the migration to the site of infection. We set the containment of the infection attributable to CD8+ T cells to 30% \[3, 248, 297\]. The total death rate of productively infected CD4+ T cells (attributed to all factors) is 1 $d^{-1}$, hence, we assume an average death rate of 0.7 $d^{-1}$.
before CD8+ T cell activation (occurring after 10 days) and an average death rate of 0.95 $d^{-1}$ after activation, allowing a small proportion of the death rate ($\leq 5\%$) to be attributable to the single epitope-specific CD8+ T cell population that is explicitly included in the model. In general, as mentioned in [298], most current models for CD8+ T cell killing in HIV assume the additive property [248, 299, 300]. To allow for stochastic effects, the lifespan of infected cells, which is the reciprocal of death rate, is sampled for each infected cell from a normal distribution (mean±sd) of: 1.4±0.25 days before introducing CD8+ T cell in the model and 1.05±0.25 days after activation (not including the death rate attributable to the explicitly modelled single epitope-specific CD8+ T cell population).

Figure 6.7: The lifespan of wild-type infected cells. During the eclipse phase of viral infection which lasts 1 day, the infected cells die at the same rate as the uninfected cells while at the viral productive phase they die at a rate of 1 $d^{-1}$. 
6.8 CD8+ T cell effector function

6.8.1 Conjugate formation

We explicitly model the formation of conjugates between the single epitope-specific CD8+ T cell population included in the model and its target cells, i.e. the wild-type infected CD4+ T cells. Once a CD8+ T cell encounters a target cell (that ‘presents’ its cognate antigen) it must make the decision of ‘acting or moving on’; we call the time this takes the scanning time. The time that it takes a T cell to scan its target is estimated to be $7 \pm 2 \text{ min}$ [301] and this process is explicitly incorporated in the model. In general, the decision to act is related to the characteristics of the TCR-pMHC interaction and the level of pMHC expression on the surface of the target cell [302]. This process is simulated in the model via a probability which defines the decision of the CD8+ T cell upon encounter to which we refer as probability of successful scan and is discussed below (see section 6.8.2). If the CD8+ T cell recognises the cell as infected then the interaction time begins. The conjugates remain immobile during both scanning and interaction time. This is a good approximation of the underlying biological process; in [286, 301] the conjugates of T cells with B cells and granuloma cells respectively end up in complete immobility of the complex within minutes after recognition.

In addition, two more important aspects of the CD8+ T cell and target cell encounter are incorporated in the model: 1) multiple CD8+ T cells can kill the same target [303, 304] and 2) multiple targets can be killed by a single CD8+ T cell as observed in vitro [305]. The latter is further supported by the configuration of the contact between TCR and APCs; during recognition of a target cell the microtubule-organising center (MTOC) of the CD8+ T cell polarises towards the immunological (IS). The high motility of the MTOC allows for rapid switch of polarization between targets [300]. However, as found in [304] most of the CD8+ T cells in conjugate for-
mation did not have more than four target cells bound to them, an observation that we implement in the model.

In summary, at the end of the interaction time between the CD8+ T cell and its target(s) the CD8+ T cell can respond as follows: 1) during a lytic response, deliver the lethal hit and therefore eliminate the target or 2) in a non-lytic response, ‘secrete’ soluble factors affecting a specified area or 3) disassociate from the conjugate formation and move on to another target (Figure 6.8). The lytic and non-lytic models are simulated separately with disassociation always being the alternative option to responding after scanning.

6.8.2 Target recognition by CD8+ T cells

The effectiveness of a CD8+ T cell response is obviously dependent on their ability to successfully survey potentially infected cells and recognise them as targets. The recognition is dictated by a series of factors such as: 1) the level of antigenic stimulation \[307, 308\], 2) the structural rearrangements of TCR-binding \[309\], 3) the confinement time of TCR-pMHC interaction \[310\], 4) the formation of the peptide-MHC complex \[311\] and others. We simulate the stochasticity of this process by setting a probability of successful scan of an infected target once the initial scanning time is completed. If the target is successfully scanned, the probability of lysis or secretion of soluble factors is set to 1. The probability of successful scan is one of the parameters that allows us to vary the immune control exerted by the specific CD8+ T cell population included in our simulations.

6.8.3 Lytic and non-lytic CD8+ T cell viral suppression

The lytic response can be divided into at least three stages \[312\]: 1) CD8+ T cells survey potential target cells and recognise a subgroup of them, 2) they then form a conjugate (see section 6.8.1) with their target in order to deliver the lethal hit and
Figure 6.8: The CD+8 T cell first scans wild-type infected cells and then decides between the following: 1) react to the infected cell (lyse or secrete soluble factors) or 2) disassociate from the target cell and move on to the next target. The lytic and non-lytic models are simulated separately with disassociation always being the alternative option to responding after scanning. Abbreviations: T=uninfected cells, WT= wild-type infected cell, VAR=variant infected cells, CD8=single epitope-specific CD8+ T cells.
3) once the target lyses, they continue hunting for new targets. All these stages are explicitly included in the model for the epitope-specific CD8+ T cell population. Only one CD8+ T cell ‘lethal hit’ can be delivered per target cell as observed in [313]. The conjugate formation in the model is set to have a 30 min duration time, in agreement with measurements obtained via intravital multiphoton microscopy [301,305] and also taking into account the disintegration time of the complex [314]. Of note, a CD8+ T cell might need to ‘rearm’ before moving on to their next target [315, 316] or pause before identifying a new target [301] therefore increasing the duration of this process (we vary this parameter in order to investigate its effect).

The non-lytic response can be summarised in the following steps: 1) CD8+ T cells survey potential target cells and recognise a subgroup of them, 2) they bind on a conjugate with their target and 3) they start secreting soluble factors while being in a conjugate, suppressing therefore viral production or infection. We model the suppression of viral infection by not allowing the infection of uninfected CD4+ T cells that ‘interact’ with the soluble factors. In a similar way, we model the suppression of viral production by not allowing infected cells to infect other cells once the former ‘interact’ with the secreted factors. The duration of the conjugate is set to 10 min as reported in [301] when no lysis is observed. The in vivo area of diffusivity of soluble factors is poorly quantified, we therefore simulate the following different secretion patterns (see Figure 6.10): 1) localised effect, where only the wild-type infected cell in conjugate with the CD8+ T cell loses its ability to infect new targets for a given interval (called the duration of the effect), 2) a polarised (grid radius, r = 1) effect where all the cells (infected or uninfected depending on the type of viral suppression) present at the immediate area on the one ‘side’ (in grid terms) of the CD8+ T cell are affected and 3) a diffusive pattern where all the cells equidistantly found in the immediate or broader area (r = 1, 2) around the CD8+ T cell are affected by the soluble factor. The duration of the effect depends on the soluble factor under consideration. Here we focus on RANTES as a case study. The recycling time for the CCR5 receptors
after interaction with RANTES has a half-life of 6-9 hrs \[317\] so we set the total duration of the RANTES ‘protective effect’ on uninfected CD4+ T cells to 10 hrs; after that period the uninfected cells can again be infected either with the wild-type or the variant strain. We consider a full protection of the uninfected targets unless otherwise stated. The ‘radius’ of non-lytic effect which can clearly influence the viral escape is closely investigated using the different simulated patterns. A summary chart of all the non-lytic models simulated is given in Figure 6.9.

![Diagram](image.png)

Figure 6.9: Non-lytic models simulated with the 3D cellular automaton model.

For lytic killing, the target cells are lysed at the time that the conjugate is detached. For non-lytic responses, the secretion of soluble factors is immediately stopped after conjugate resolution. This is supported by experimental observations. In [318] they show that IFN-γ production from activated CD8+ T cells is terminated immediately after the contact of the T cell and its cognate antigen is disrupted. This can be a regulatory mechanism of preventing excessive cytokine production which can be destructive for the host. In the same study [318], in vitro experiments indicate that peptide-pulsed spleen cells cause the production of cytokines (IFN-γ and TNF-α) by LCMV-specific CD8+ T cells within 30 mins and in vivo, no cytokine production was observed in the absence of stimulation.
Figure 6.10: Secretion patterns of soluble factors by CD8+ T cells under a non-lytic response. The red nodes represent the activated CD8+ T cell and the green nodes depict the area that is affected by soluble factors secreted by this CD8+ T cell.
6.9 Quantifying CD8+ T cell killing and viral escape

We quantify CD8+ T cell killing and viral escape on the population level using two different models: 1) a model presented in [218] which assumes mass-action killing, i.e. killing is proportional to infected target cells and effector cells and 2) a model that we describe here and assumes that the rate of killing of the effector cells can eventually saturate in terms of target cells. Using the simulations of the CA model, we can also calculate the actual killing rate in the simulated datasets and use this information to investigate which of the models better predicts the CD8+ T cell killing rate.

6.9.1 Simulated killing rate

The CA model is implemented such that we record the exact number of wild-type infected cells killed by the epitope-specific CD8+ T cell population at each timestep (30 sec). Hence, using the simulated data we can readily calculate the absolute number of targets killed per day by the epitope-specific CD8+ T cell population, \( n_s \), as well as the ‘actual’ killing rate per day, \( k_s \), as follows:

\[
\begin{align*}
  n_s &= \# \text{ WT Infected cells killed (day)} & (6.5) \\
  k_s &= \frac{n_s}{\# \text{WT Infected cells (day)}} & (6.6)
\end{align*}
\]

6.9.2 Mass-action killing model

First, we use the model developed in [218] to calculate the escape rate of the variant infected cells which under the assumption of mass-action killing equals the epitope-specific CD8+ T cell killing rate (see below). Based on this model, CD4+ T cells productively infected with wild-type virus, \( y \), replicate at rate \( a \), die at rate \( b \) (including
the death attributable to specific-CD8+ T cells recognising epitopes other than the escape epitope) and are also killed by specific-CD8+ T cells recognising the escape epitope at rate \( k \). CD4+ cells productively infected with an escape variant, \( x \), replicate at rate \( a' \) and die (including death by specific CD8+ T cells recognising epitopes other than the escape epitope) at rate \( b \). Hence,

\[
\dot{y} = ay - by - ky \quad (6.7)
\]

\[
\dot{x} = a'x - bx \quad (6.8)
\]

If \( p(t) \) is the proportion of productively infected cells infected with the variant viral sequences, the following holds [248]:

\[
p(t) = \frac{x(t)}{x(t) + y(t)} = \frac{1}{ge^{-ct} + 1} \quad (6.9)
\]

where \( g = \frac{y(0)}{x(0)} \), \( t=0 \) is the time when the variant is introduced in the viral population and \( c = k - (a - a') \), is the escape rate of the variant viral strain. If we consider no fitness cost for the replication ability of the variant virus, \( a = a' \), then there is no difference in replication rate between the escape variant and the wild-type, \( a - a' = 0 \) and hence \( c = k \). It should be noted that \( k \) quantifies the overall efficiency of the epitope-specific CD8+ T cell population.

In the simulations, viral escape is quantified during the chronic phase of the infection (starting at 50 dpi). Equation (6.9) can be used to estimate the variant escape rate irrespective of the assumptions made in the model described by Equations 6.7-6.8 but it can only be used as an estimate of the killing rate if the mass-action killing assumption holds.

**6.9.3 Saturated killing model**

Then, we also formulate a model where the killing rate of CD8+ T cells saturates after a specific infected target cell threshold is exceeded. The same saturation killing term
is also discussed in [319]. This model can be described as follows:

\[ \dot{y} = ay - by - \frac{v_{\text{max}}y}{k_m + y} \]  
\[ (6.10) \]

\[ \dot{x} = a'x - bx \]  
\[ (6.11) \]

where the variables \( y \) and \( x \) and the parameters \( a, b \) and \( a' \) are the same quantities described in the previous model. Additionally, \( v_{\text{max}} \) is the maximum number of infected cells that can be killed per day when \( y \to \infty \) and \( k_m \) is the number or infected target cells that would give half the maximum number killed per day, \( v_{\text{max}} \).

The Equations 6.10-6.11 can not readily be rewritten in a similar format to Equations 6.9 but can still be fitted to 'escape data' as described by the ratio \( p(t) = \frac{x(t)}{x(t)+y(t)} \).

In addition, because we assume no fitness cost for the variant viral strain we set again \( a' = a \). Hence, we can easily reduce the above model in the following form:

\[ \dot{y} = fy - \frac{v_{\text{max}}y}{k_m + y} \]  
\[ (6.12) \]

\[ \dot{x} = fx \]  
\[ (6.13) \]

where \( f = a - b \). This model (reduced or not) includes two main different sub-models of killing depending on the availability of infected target cells. If \( y << k_m \) then the model describes mass-action killing whilst if \( y >> k_m \) then the number of infected cells killed per day saturates and converges to \( v_{\text{max}} \) (Figure 6.11).

Using the estimates of the wild-type population, \( y \), estimated when fitting the model to the data, we can obtain the average killing rate per day by calculating the average of the quantity \( \frac{v_{\text{max}}y}{k_m + y} \) over all the fitted values of \( y \). This provides an estimate of the average killing rate when assuming a saturated killing process.
Figure 6.11: Schematic of a model which includes saturated epitope-specific CD8+ T cell killing per day. This model includes two different sub-models of killing depending on the availability of infected cells (see Equation 6.12). If $y << k_m$ then we have mass-action killing but if $y >> k_m$ then the number of infected cells killed per day converges to a maximum value, $v_{max}$. 
6.10 Discussion

We have developed and implemented a complex agent-based model that captures both acute and chronic phase dynamics of HIV/SIV infection. The model includes detailed rules of T cell motility, viral production, death rate of infected and uninfected CD4+ T cells, conjugate formation of CD8+ T cells and infected targets and CD8+ T cell response that can be manifested lytically or non-lytically.

Importantly, the set-up of the model allows for the future study of numerous additional aspects of HIV/SIV infection such as fitness cost of variant strains, simulation of multiple epitope-specific CD8+ T cell populations, combination of lytic and non-lytic CD8+ T cell responses, early killing by CD8+ T cells [257], superinfection of infected cells and others.

We next use this model to study the CD8+ T cell killing process on a population level, investigate the efficiency of lytic and non-lytic CD8+ T cell responses and explore ‘viral escape’ dynamics under these different CD8+ T cell effector functions.
Chapter 7

CD8+ T cell effector function and viral escape

7.1 Aim

Our aim was to investigate whether non-lytic CD8+ T cell responses can drive the growth of escape variants, identify factors involved in their outgrowth and quantify the relationship between viral control and the outgrowth rate.

7.2 Introduction

It is widely accepted that CD8+ T cells suppress replication of human (HIV) and simian (SIV) immunodeficiency virus but this process is not yet fully understood. CD8+ T cells can control viral replication via lytic or non-lytic effector mechanisms. Lytic mechanisms require direct contact with the infected target cell and result in apoptosis. Non-lytic mechanisms are mediated by soluble factors and reduce viral production by infected cells or viral infection of uninfected cells. Strong evidence that CD8+ T cell-secreted chemokines can play an important role in inhibiting HIV infection come from in vitro experiments involving RANTES, MIP-1α and MIP-1β which bind CCR5 and act as competitive inhibitors of CCR5-mediated HIV/SIV entry. Further findings in the same direction have been reported by multiple studies and extended to other soluble factors reviewed in.

Recently, ground-breaking studies reported that following CD8+ T cell depletion
in SIV-infected macaques, viral load robustly increased, however the lifespan of SIV-infected cells remained unaltered [42,167]. These unexpected results led to the suggestion that SIV might be controlled via non-lytic mechanisms; a suggestion which was further studied and corroborated in [3] (see Chapter 5). In the latter study, ODE models of CD8+ T cells with lytic and non-lytic effector mechanisms were fitted to the experimental data obtained in [42]. The non-lytic models were found to be a better predictor of viral load measurements.

Both HIV and SIV are characterised by the evolution of viral epitope mutants that can escape CD8+ T cell responses. The virus mutates at the base substitution rate \((10^{-4} - 10^{-3} \text{ per bp})\) defined by the error-prone replication process. Cells can be infected with the variant virus strains that have not lost their replicative capacity. Additionally, variants of an epitope can arise from point mutations in or around the epitope presented by MHC class I molecules to CD8+ T cells. These mutations may lead to impaired recognition of the variant infected cells resulting in a higher immune pressure exerted on the wild-type virus. Eventually, the variant infected cells might outgrow the wild-type infected cells; a process termed as ‘viral escape’. The outcome of escaping a CD8+ T cell response is given by the trade-off between evading the CD8+ T cell response and the cost that this poses on viral fitness. Hence, evading a CD8+ T cell attack does not necessarily provide a growth advantage to the virus and this is indicated both by the emergence of compensatory mutations as well as by reversion to the wild-type in MHC-mismatched hosts. Nevertheless, there are many studies in the literature reporting the rapid or slow successful escape of specific epitopes from CD8+ T cell control during primary [323,326] and chronic infection [164,165,276,327].

Whilst it is clear that under a lytic mechanism an escape variant has a fitness advantage compared to the wild-type, it is less obvious whether this holds in the face of non-lytic control where both wild-type and variant infected cells would be similarly affected by soluble factors. Here, we use the Cellular Automaton model described in Chapter 6 to explore whether non-lytic CD8+ T cell responses can result in viral
escape and how the escape rate compares to the one attained via a lytic response. We also investigate the killing process under a lytic mechanism and the efficiency of non-lytic mechanisms in controlling infection.

7.3 Lytic CD8+ response and viral escape

7.3.1 The CA model captures HIV/SIV dynamics

Before using the CA model to simulate the different CD8+ T cell effector functions, we investigate the correspondence of the model dynamics to the observed HIV/SIV dynamics described in multiple studies in the literature.

The model has a typical viral expansion phase, with a viraemic peak reached after approximately two to three weeks (Figure 7.1). The infected CD4+ T cells in the first days post infection (dpi) grow at a rate close to $1.5 d^{-1}$ which is within the $1-2 d^{-1}$ range that has been reported in other studies [274,328,329]. Close to the peak there is a large depletion of CD4+ T cells and the uninfected target cells become limited. The immune control introduced at day 10 post infection together with the target limitation result in the slowing down of viral growth and the subsequent decline to a viral set-point within a few weeks. The proportion of infected CD4+ T cells at the steady-state ($> 30$ dpi) is of the order of $10^{-3} - 10^{-2}$ in compliance with experimentally observed values [290].

7.3.2 Probability of recognition by CD8+ T cells and killing rate

We model different levels of specific CD8+ T cell pressure in the absence of variant virus. We simulate CD8+ T cell pressure expressed in terms of different epitope-specific CD8+ T cell killing rates. We calculate the average simulated killing rate observed
Figure 7.1: Dynamics of uninfected CD4+ T cells and wild-type infected CD4+ T cells over a course of 150 days when no variant infected cell population is present. The bars represent the 95% central range values. The sizes of the populations (y-values) are given on a log10 scale.
over 25-50 dpi using Equation 6.6. Given that we consider a constant CD8+ T cell population in magnitude and we have calibrated the cell motility to biological ranges, three parameters in the model could influence the killing rate: 1) the probability of successful scan (recognition of infected target cells which leads to cytolysis or secretion of non-lytic factors), 2) the scanning time and 3) the duration of the killing process. Our aim is to simulate an epitope-specific CD8+ T cell population that can kill 1%-5% of the infected cells per day, in the range estimated in [248] for HIV specific CD8+ T cell clones recognising one epitope. We next explore how these three parameters can affect the killing rate.

First, we vary the probability of successful scan. Initially we set this probability to 1, i.e. each ‘interaction’ of a CD8+ T cell with a wild-type infected CD4+ T cell leads to a response (lytic in this case) but we find that this leads to an unrealistically high killing rate (> 20 fold clearance per day). Interestingly, we find that if a CD8+ T cell has a probability as small as 0.001 to successfully scan its target upon collision then the killing rate achieved is on average 1% per day (Figure 7.2) and for an epitope-specific CD8+ T cell population to kill daily 100% of the infected cells, a probability of successful scan of 0.025 suffices (Figure 7.2 inset). In other words, the epitope-specific CD8+ T cell population that constitutes 0.5% of the total splenocyte population can kill 1% of the infected CD4+ T cell population within one day. By increasing the probability of successful scan to 0.002 and 0.003 we observed a median killing rate of 3% and 5% respectively (Figure 7.2). A probability of successful scan of 0.001 implies that a CD8+ T cell will attempt 1000 scanning events on average before it successfully recognises and responds to the antigenic stimulus (500 and 330 infected targets for probabilities 0.002 and 0.003 respectively).

Scanning time is restricted to a few minutes, is not expected to vary significantly and is already modelled as a distribution (7 ± 2 min [301]) hence, we do not vary this parameter further.

It has been estimated in multiple studies that the CD8+ T cell killing process takes
approximately 30 min[25][30][30]. However, a CD8+ T cell might, for example, need to ‘recharge’ before it is able to kill again and this time can be additive to the killing time[31][32]. So, to explore the robustness of the killing rate against the duration of killing, we vary the parameter in the range of 15 to 120 min. No substantial effect of the killing duration on killing rate is observed although there is a trend for the killing rate to decrease slightly with the increasing duration of killing (Figure 7.3). The very limited effect is perhaps not surprising since even the increase of the killing time by 4-fold would still result in a killing process that evolves in a fast timescale.

The fact that the probability of successful scan has a large impact on the killing rate while the duration of killing does not, implies that the process of efficiently identifying rather than rapidly eliminating a target defines the killing rate of infected CD4+ T cells.

### 7.3.3 Higher CD8+ T cell killing leads to faster viral escape

Next, we investigated the dynamics of infection when a variant infected population (seeded at a magnitude which is approximately 30% of the wild-type population) is included in the model. The variant infected cells appear at a random age and occupy an available free space on the grid. The variant population is introduced at 50 dpi when the steady-state has been reached and the specific-CD8+ T cell population is not expected to expand and contract substantially (as it does during the acute phase). We model single epitope-specific CD8+ T cell lytic responses averaging to three different killing rates, $k_s$: (i) 1%, (ii) 3% and (iii) 5% per day. We run 25 simulations for each response (not all result in escape) by using different ‘initial random seeds’ (Figure 7.4) and observe that the variant population outgrows the wild-type population at a faster rate as the CD8+ T cell pressure increases (Figure 7.6). We fit Equation 6.9[248] to the simulated datasets in order to estimate the variant escape rate (Figure 7.5).
Figure 7.2: Probability of successful scan by epitope-specific CD8+ T cells and the corresponding observed killing rates. The dotted line represents the lower and upper bounds of the single epitope-specific CD8+ T cell killing rate that we aim at simulating (1 – 5%). The inset plot shows the killing rates recorded for higher probabilities of successful scan. The killing rate is calculated based on 25 simulations using Equation 6.6 and is averaged over 25-50 dpi.
Figure 7.3: Duration CD8+ T cell killing and the corresponding observed killing rates. The killing rate is calculated based on 10-25 simulations in each case using Equation 6.6 and is averaged over 25-50 dpi. The probability of successful scan is set to 0.001 for all the simulations shown here. Varying the killing duration showed no statistically significant differences in killing rate compared to a killing duration of 30 mins (p>0.05) apart from the case where the killing duration was increased to 120 mins (p=0.03); however, even in this case the mean difference in killing rate was only 0.2%.
Importantly, when we increase the grid size of the CA this trend as well as the killing and escape rate estimates remain similar ($p > 0.05$, Appendix Figure F.1).

The model used to estimate the escape rate assumes a mass-action specific-CD8+ T cell killing term which we investigate closer. Given the lack of replicative cost for the variant population, we expect that the killing rate and the variant escape rate would be approximately equal (see section 6.9.2). However, as it is readily seen in Figure 7.6 whilst the escape and killing rates are indeed significantly ($p=0.002$) positively correlated, they are not of equal magnitude. This interesting result can potentially be attributed to aspects of the CA model that are not included in the ODE model described by Equations 6.7-6.8. The main differences are the inclusion of an ‘eclipse’ phase, the CD4+ T cell influx of uninfected cells and potentially the mass-action killing assumption considered in the ODE description. When we introduce an eclipse phase or a homeostatic influx (instead of a steady influx) of uninfected cells to the ODE model the estimated escape rate is unchanged, indicating that neither of these two model aspects can explain the observed difference. Hence, the form of the killing term remains the most probable explanation of the observed discrepancy and we explore this further.
Figure 7.4: Dynamics of uninfected, wild-type and variant infected CD4+ T cell populations over a course of 150 days when the variant infected cell population is introduced at 50 dpi at an initial magnitude of 30% of the wild-type. We model three levels of epitope-specific CD8+ T cell killing: i) $k_s \approx 1\%$ (a), ii) $k_s \approx 3\%$ (b) and iii) $k_s \approx 5\%$ (c) of infected cells killed per day. We run 25 simulations in each case. Only the simulations that lead to escape ($> 10$ in each case) are shown. The bars represent the 95% central range values. The sizes of the populations (y-values) are given on a log10 scale.
Figure 7.5: We estimate the variant escape rate in the simulated datasets using Equation 6.9 and under three levels of epitope-specific CD8+ T cell killing rate: (i) $k_s \approx 1\%$, ii) $k_s \approx 3\%$ and iii) $k_s \approx 5\%$ of infected cells killed per day. The variant infected cell population is seeded at 50 dpi. We run 25 simulations in each case. Only the simulations that lead to escape (> 10 in each case) are shown.
Figure 7.6: Higher CD8+ T cell killing leads to faster viral escape. We model three levels of epitope-specific CD8+ T cell selection pressure: (i) $k_s \approx 1\%$, ii) $k_s \approx 3\%$ and iii) $k_s \approx 5\%$ of infected cells killed per day. We run 25 simulations in each case. Only the simulations that lead to escape (> 10 in each case are shown). The killing rate is estimated over 25-50 dpi using Equation 6.6. The escape rate is estimated by fitting Equation 6.9 to the simulated datasets.
7.3.4 CD8+ T cell killing rate is higher during ‘escape phase’

Next, we investigated more closely the CD8+ T cell response under a lytic mechanism. We studied two quantities: 1) the absolute number of wild-type infected cells killed by the epitope-specific CD8+ T cell population per day and 2) the killing rate, described by Equations 6.5 and 6.6 respectively. As shown in Figure 7.7, the killing rate per day is not constant and it increases after the variant-infected population starts outgrowing the wild-type population. We call the phase after the variant population appears the ‘escape phase’. Given that the number of effector cells in our model is constant, this increase can be explained by two main hypotheses: i) the number of wild-type infected cells killed per day by the epitope-specific CD8+ T cell population increases over the course of infection or 2) the number of wild-type cells killed is approximately constant but the overall wild-type population decreases and therefore the fraction of wild-type infected cells killed increases. By calculating the absolute number of wild-type cells killed per day by the epitope-specific CD8+ T cells, we find that the number of cells killed remains approximately constant over time (Figure 7.8). This confirms our second hypothesis; the variant population outgrows the wild-type during the ‘escape phase’ and because the wild-type population shrinks the killing rate increases but the absolute killing does not.

Interestingly, the fact that the number of wild-type infected cells killed per day by the specific CD8+ T cells are almost constant although the number of wild-type infected cells varies a lot during this period (Figure 7.4), implies that there is an upper threshold to the number that the CD8+ T cells can kill and therefore the killing process might be better explained by a saturated killing term. In addition, as it is readily shown in Figure 7.7 we obtain different killing rate estimates depending on whether we perform the calculation before or during the ‘escape phase’ since during the escape phase the variant strain gradually replaces the wild-type strain.
Figure 7.7: CD8+ T cell killing rate is higher during 'escape phase'. The three time-points labelled are: CTLs=The day at which the HIV-1 epitope-specific CD8+ T cell population is introduced in the model, Set-point=The approximate day that steady-state is attained, Variant=The day that the variant infected population is introduced in the model.
Figure 7.8: The absolute number of wild-type infected cells killed per day is approximately constant. The three timepoints labelled are: CTLs=The day at which the HIV-1 epitope-specific CD8+ T cell population is introduced in the model, Set-point=The approximate day that steady-state is attained, Variant=The day that the variant infected population is introduced in the model.
7.3.5 Saturation term better estimates killing rate

The form of the killing function is important for determining the dynamics of the immune response. Usually, in models of immune control the killing process is defined by a mass-action term \[ \text{mass-action term} \] (e.g. Equation 6.7) which assumes that the killing of infected cells is proportional to both infected target cells and effector cells. In the CA model the number of epitope-specific CD8+ T cells, i.e. the effector cells is assumed constant over time because we focus on the chronic phase of infection. However, we find that if we increase the number of effector cells then the killing rate increases and that the relation between the two quantities is consistent with a mass-action behaviour in terms of effector cells (Appendix Figure F.2); however, the data are noisy so we cannot draw further conclusions at this point. So, we decide to explore if mass-action also holds in terms of infected target cells. We have already shown (Figure 7.8) that the number of wild-type infected cells killed during the expansion and contraction of the infected target cells is approximately constant. This suggests that the mass-action killing term is not the most appropriate term to describe the killing process under a lytic mechanism. As it has been argued in other studies, limiting aspects such as finding, recognising and adhering to the target can create a saturation effect similar to Michaelis-Menten enzyme-substrate kinetics [139,319,330].

In order to thoroughly address whether a mass-action term or a saturated killing term better explains the underlying killing process, we use the simulated ‘escape data’ that capture the HIV/SIV dynamics and fit the two different models (Equations 6.9 -which under the mass-action assumption and lack of fitness cost is equivalent to 6.7, 6.8 and 6.12[6.13] respectively). Both models provide satisfactory fits (Figures 7.5 and 7.9) but when we use the AICc (Equation 5.10) to select the best fitting model, we find that the model which assumes saturated CD8+ T cell killing (in terms of infected target cells) explains the data significantly better \( p=0.001 \), paired Wilcoxon test; Figure 7.10. In addition, when we compare the average killing rate estimated
using the two models before and after the ‘escape phase’ we find that the saturated killing model predicts the killing rate more accurately during the ‘escape phase’ while the mass-action model underestimates it substantially (Figure 7.11). We also find that both models overestimate the killing rate before the ‘escape phase’ (Figure 7.11). This is expected since the killing rate increases during the ‘escape phase’ and both models are fitted over the period before and during the ‘escape phase’. However, the killing rate that relates to the escape rate is expected to be better described by the ‘escape phase’ estimate, making therefore the saturation killing model a better selection.

By exploring the values of the infected target cells needed in order to start transitioning from mass-action killing to saturated killing, $k_m$ (Figure 6.11), we find that the number of infected CD4+ T cells is already higher than this value when the specific-CD8+ T cell population is introduced in the simulation (10 dpi) and remains higher for most of the timecourse. This suggests that the killing of the specific CD8+ T cell population is better described by a saturated rather than a mass-action killing term during the timecourse of the simulations. However, in the case of the higher specific CD8+ T cell killing rate ($\approx 3 - 5\%$) we observe that the absolute number of infected target cells decreases towards the end of the simulated time course (Figure 7.8). This is expected because only at this killing rate level, the number of available wild-type infected targets drops below the $k_m$ value (then the wild-type infected cells are almost entirely replaced by the variant infected population within the 3 months simulation) and therefore at this particular case the killing rate is better described by a mass-action term.
Figure 7.9: We use the saturated killing model (Equation 6.11) in order to estimate the killing rate in the simulated datasets under the three levels of epitope-specific CD8+ T cell selection pressure (i) $k_s \approx 1\%$, ii) $k_s \approx 3\%$ and iii) $k_s \approx 5\%$. We run 25 simulations in each case. Only the simulations that lead to escape (> 10 in each case) are shown.
Figure 7.10: The AICc of the fitted models over different simulated datasets. We fitted both the mass-action and saturated killing models to 51 simulated ‘escape datasets’ and calculated the bias-adjusted AICc which penalises for the number of model parameters used. A lower AICc implies a better fit to the data. The saturated killing model has a significantly lower AICc to the mass-action model (p=0.001, paired Wilcoxon test). It is also readily observed that the light blue bars that show the AICc of the saturated killing model provide a better fit for most of the datasets (35 out of 51).
Figure 7.11: Correlation of simulated and estimated killing rate using a mass-action or a saturated killing term. The estimates before ‘escape phase’ are calculated over 25-50 dpi and the estimates during ‘escape phase’ are calculated over 50-150 dpi (if the wild-type infected cell population is eliminated before 150 dpi then the last positive value defines the end of the phase).
7.4 Non-lytic CD8+ response and viral escape

7.4.1 Non-lytic suppression leads to slower viral escape dynamics

Although it is readily understood how a lytic CD8+ T cell response can drive the expansion of a variant strain, it is not obvious how the dynamics of escape will be reshaped under a non-lytic control. Here, we use the CA model in order to explore the dynamics of a non-lytic CD8+ T cell response which blocks viral infection or viral production and how it relates to the rate of outgrowth of the variant infected CD4+ T cells.

The term ‘non-lytic control’ can imply the effect of many different soluble factors which act in turn or simultaneously. We model -as a case study- the secretion of RANTES for which the viral suppressive effect has been studied [317] (see Chapter 6). The simulation of the non-lytic effect that blocks infection can be summarised as follows: We consider that the ‘protective effect’ is conveyed to all uninfected CD4+ T cells that travel through the area of secretion (polarised or diffusive, see Chapter 6) formed around the activated epitope-specific CD8+ T cells. This effect lasts for 10 hrs (as measured for RANTES) and totally abolishes the possibility that these uninfected targets become infected. Once this period expires the formerly ‘protected’ cells become again susceptible to infection. In a similar way, a non-lytic effect that blocks production can be modelled by abrogating for a given time the infectiveness of infected cells that cross the area of secretion (see Chapter 6).

Using the CA model, we simulate a non-lytic single-epitope specific CD8+ T cell response mechanism and introduce a variant infected population as described for the lytic mechanism at 50 dpi. We set all the other parameters of the model to the exact same values as in the model of the lytic response. This allows us to study the two
different control mechanisms under exactly the same parameter space with only the
viral control being different. Hence, we can observe how the CD8+ T cell pressure in
each case shapes the viral escape dynamics. We find that in the case of a non-lytic
response that blocks infection or production (for 10hrs), the viral escape (estimated
using Equation 6.9) is slower, on average 1% per day and it does not increase with
increasing CD8+ T cell control or increasing radius of secretion (Figure 7.12).

In light of this finding, we first examine whether the number of uninfected CD4+ T
cells ‘protected’ from blocking infection are increasing as both the radius of secretion
and the CD8+ T cell pressure (measured by the probability of successful scan) are
increasing. We find that in both cases, as expected, there is a clear increase of protected
uninfected targets (Figure 7.13). Similarly, we find that under a non-lytic response
that blocks viral production from infected CD4+ T cells, the number of cells ‘blocked’
increases with both increasing radius of secretion and higher CD8+ T cell pressure
(Appendix Figure F.3).

Next, in order to understand whether this slower escape rate can be attributed to
the lack of a strong immune control or to intrinsic properties of a non-lytic response, we
quantified the immune control in both the epitope-specific lytic and non-lytic responses
(blocking infection or production) in terms of the number of new infections prevented
per day. This quantity can be calculated by simulating the infection dynamics of two
models: a) one with and b) one without the epitope-specific CD8+ T cell population,
and by then calculating the number of new infections produced per day in each case.
We do this for 40-50 dpi, just before the variant infected cell population is introduced in
the simulations. The difference of new infections per day between these two models is
the number of infections prevented by the epitope-specific CD8+ T cell population. We
find that the number of infections prevented under a non-lytic pressure is comparable
to a lytic killing pressure of 1% per day (Figure 7.14 and Appendix Figure F.4); this
is also in agreement with our observation that a non-lytic control leads to a lower
set-point of productively infected cells during the steady-state compared to a lytic
control of 1% per day (Figure 7.10 and Appendix Figure F.5). Although the immune control is comparable for the low lytic response (1% killing per day) and the non-lytic response, the lytic control results in a significantly faster escape rate (approx 3.5% vs 1.2%, p<0.05; Figure 7.12). Interestingly, the lytic immune control is higher compared to the non-lytic when we increase the probability of successful CD8+ T cell scanning but the non-lytic control remains approximately unchanged (Figure 7.12). In line with the above findings, we also observe that for the lytic control, the escape rate is positive for ≈80% of the overall simulations (with this number increasing with probability of successful scan), while for the non-lytic response that blocks infection we only have an average of 40% of the simulations returning a positive estimate and for the non-lytic response that blocks viral production although this percentage goes up to 60% it is still lower than for lytic control and in neither of the latter cases there is a trend with increasing probability of successful scan (Appendix Table F.1).

Therefore, the lack of immune control cannot explain the slower dynamics of escape that we observe when the CD8+ T cell viral suppression is non-lytic. This suggests that intrinsic properties of the non-lytic control are responsible for the slower variant growth over the wild-type. In the case of blocking infection, the fact that the uninfected targets are ‘protected’ from both wild-type and variant infection limits the advantage of the variant over the wild-type even when the effect of the soluble factor is not extended to a wide area around the activated specific CD8+ T cell (e.g. polarised pattern). Similarly, the fact that both wild-type and variant infected cells can be ‘silenced’ by soluble factors obviously restricts the advantage of the variant population.

To summarise, we observe slower and less frequent viral escape under a non-lytic compared to a lytic control mechanism which remains similarly slow for the different non-lytic secretion patterns (polarised and diffusive) and despite of an increase in the probability of successful scan by the effector cells; this is observed both when the viral production or the viral infection are abrogated.
Figure 7.12: Escape rate is slower when the CD8+ T cell selection pressure is exerted in a non-lytic manner. We simulated different secretion patterns of non-lytic factors and different levels of selection pressure in each case. We estimate the variant escape rate in the simulated datasets using Equation 6.9. The ‘protective effect’ lasts for 10 hrs and it totally safeguards the cells from infection. Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, NLp: Non-lytic model - blocking viral production from infected CD4+ T cells, TO=Target Only, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Figure 7.13: Number of uninfected CD4+ T cells ‘protected’ from infection under a non-lytic CD8+ T cell response manifested in a polarised or diffusive secretion pattern. We show the cumulative number at 50 dpi. Abbreviations: NLI: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Figure 7.14: New infections prevented under a non-lytic CD8+ T cell response that blocks infection. We show the number of infections prevented by the epitope-specific CD8+ T cell clones for 40-50 dpi, just before the variant infected cell population is introduced in the simulations. Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Figure 7.15: Set-point of productively infected cells. We show the percentage of productively infected wild-type cells for 40-50 dpi, just before the variant infected cell population is introduced in the simulations and after the steady-state has been attained. Here, we present the results for the lytic control and the non-lytic control that blocks viral infection (see Appendix F for non-lytic response that blocks viral production). Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
7.4.2 Cluster formation of infected cells

Using the CA model, we showed that under a comparable immune control, the variant infected cells outgrow the wild-type infected cells at a slower rate when the CD8+ T cell effector mechanism is non-lytic. However, the reason why the variant infected cells have an advantage over the wild-type cells -even if limited- is not immediately understood. The non-lytic response, either the one that blocks viral infection or the one that blocks viral production, affects both infected populations; so under the assumption of spatially well-mixed cell populations, neither would be expected to have an advantage. We further explored whether there are spatial patterns that can explain this advantage of the variant population under a non-lytic control.

Although on average there are approximately similar numbers of wild-type and variant infected cells around infected CD4+ T cells of any of the two types, the overall distribution is highly skewed toward the presence of cells of the same type (Figure 7.16). The percentage of wild-type infected cells around a wild-type type cell is predominantly $\geq 50\%$ while the percentage of variant infected cells is mainly $\leq 50\%$. The same is true for the infected cells surrounding the variant infected cells, indicating the presence of clusters of wild-type and variant infected cells. The existence of such clusters rejects the well-mixed assumption and suggests that spatial patterns can provide an advantage to the variant infected cells since only the wild-type infected cells -found usually close to other wild-type cells -can trigger the secretion of soluble factors by a CD8+ T cell that would suppress viral spread. If a wild-type infected cell activates the CD8+ T cell and this in turn secretes factors that block infection in the surrounding area, it would be more likely to affect wild-type infected cells and thus the variant infected cells have an advantage even though they are equally susceptible to non-lytic factors. Nevertheless, this process cannot have an extensive effect because of the fast timescale of T cell motility that forces the spatial effects to quickly mix. However, this is a dynamic process and the repetitive formation of such clusters during the spread of the
infection is likely to provide a net whilst limited advantage to the variant infected cell population that can explain outgrowth albeit at a slower speed.

Figure 7.16: The percentage of wild-type infected cells around a wild-type cells is significantly higher than the percentage of variant infected cells that surround it. The same is true for variant infected cells. We track simulated wild-type and variant infected CD4+ T cells (≈ 500 cells) at multiple timepoints. Abbreviations: WT/WT=Wild-type infected cells in the immediate area surrounding a wild-type infected cell, WT/VAR=Wild-type infected cells in the immediate area surrounding a variant infected cell, WT/ALL=Wild-type infected cells on the whole grid, VAR/VAR, VAR/WT and VAR/ALL are defined in the same way.

### 7.4.3 The efficiency of the non-lytic control mechanism

Although we find that CD8+ T cells that act non-lytically can lead to viral escape, we also observe that non-lytic control results in a lower number of ‘prevented’ cell
infections and this does not increase either with the radius of secretion or the increase of the successful encounters with infected cells (Figure 7.14). Therefore, we explore further the efficiency of the non-lytic control not in terms of viral escape but in terms of limiting the growth of the wild-type.

There are four main quantities in the CA model that could potentially strengthen the non-lytic response: 1) the radius of secretion, 2) the probability of successful scan, 3) the duration of the effect of the soluble factor and 4) the size of the CD8+ T cell population. We vary each parameter separately by setting all other parameters at a constant value. We already saw that the first two parameters did not have a substantial effect on boosting the response. Additionally, when we double the effect of the soluble factor from 10 to 20 hrs, we find no consistent substantial impact on controlling the spread of the wild-type either by blocking infection (Figure 7.17) or by blocking production (data not shown). Finally, we increase the size of the effector population from 0.5% to 1.5% of the splenocytes. Although, the number of cells that are ‘protected’ by the soluble factors increases (Figure 7.18), again we observe no consistent substantial effect on the strength (in terms of infections prevented) of the non-lytic control (Figure 7.19); we also see no effect on the set-point of the infected cell population. This result becomes even clearer when we juxtapose it to the lytic control where the increase of the size of the effector population results in a stronger immune pressure both in terms of infections prevented (Figure 7.19) and set-point infected cell population (data not shown). Our findings taken together suggest that the non-lytic control can be less efficient in limiting the infection compared to lytic control and might require a very large number of effector cells in order to provide strong viral suppression.
Figure 7.17: New infections prevented under a non-lytic CD8+ T cell response that blocks infection for varying duration of the effect of the soluble factor (SE). We show the number of infections prevented by the epitope-specific CD8+ T cell clones for 40-50 dpi, just before the variant infected cell population is introduced in the simulations. We run 10 simulations for each case. The probability of successful scan is set to 0.002 in all the simulations. Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1= Polarised secretion (r=1), D1= Diffusive secretion (r=1) and D2= Diffusive secretion (r=2).
Figure 7.18: Number of uninfected CD4+ T cells ‘protected’ from infection under a non-lytic CD8+ T cell response manifested in a polarised or diffusive secretion pattern for different sizes of the epitope-specific effector cell population, C, as a percentage of the total splenocyte population. We show the cumulative number at 50 dpi. We run 10 simulations for each case. The probability of successful scan is set to 0.002 in all the simulations. Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Figure 7.19: New infections prevented under a non-lytic CD8+ T cell response that blocks infection. We show the number of infections prevented by different sizes of the epitope-specific CD8+ T cell clones, C, for 40-50 dpi, just before the variant infected cell population is introduced in the simulations. We run 10 simulations for each case. The probability of successful scan is set to 0.002 in all the simulations. Abbreviations: NLi: Non-lytic model - blocking infection of uninfected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
7.5 Discussion

Emerging evidence suggests that CD8+ T cells can control HIV viral burden through multiple mechanisms \cite{37, 42, 58, 167, 331} which can be broadly classified as lytic and non-lytic. Although, the importance of non-lytic effects in HIV infection is shown by multiple studies (reviewed in \cite{50}) little is known about the efficiency of such a response compared to the lytic mechanism. Quantifying the CD8+ T cell lytic and non-lytic immune control would require a complex set of \textit{in vivo} experimental data that would capture spatial and temporal aspects of the system. To address these questions \textit{in silico} we construct a 3D cellular automaton model of HIV/SIV dynamics. Our aim is two-fold: 1) to investigate the efficiency of lytic and non-lytic responses and 2) to explore whether non-lytic control can drive viral escape. Whilst it is readily understood how direct killing of wild-type infected cells can result in variant outgrowth, this is not as obvious under a non-lytic immune suppression mechanism that blocks either viral infection or viral production and which can equally affect wild-type and variant infected cells.

In summary, the 3D cellular automaton model includes an extensive set of rules on cell motility, infection spread and cell death that can recapitulate experimentally observed dynamics of HIV/SIV infection and can be used for the explicit simulation of both lytic and non-lytic epitope-specific CD8+ T cell responses.

For a lytic response, the model allows us to study closely the CD8+ T cell killing term. We find that the killing process is better explained by a model that incorporates both mass-action killing and saturated killing. This implies that CD8+ T cell killing can be substantially restricted by factors other than the availability of infected cells. Specifically, we find that the killing rate is predominantly limited by the time that the CD8+ T cell needs in order to successfully identify its infected target and less restricted by the duration of killing \textit{per se} or the number of targets. Interestingly, we find that
biologically plausible CD8+ T cell killing rates can be attained for a small probability of successful identification of infected cells. Furthermore, it has been suggested by other studies that a mass-action term is only adequate for well-mixed populations at intermediate effector:target ratios \[332\]. In the simulations however, we observe the formation of clusters that do not align with a well-mixed hypothesis and could possibly result in the break-down of the mass-action behaviour in terms of infected targets; at this point though, this is only a potential explanation. Nevertheless, in terms of effector cells, our findings are consistent with a mass-action behaviour, i.e. we cannot exclude that the killing rate increases proportionally to the size of the CD8+ T cell population.

Additionally, we study the relationship between immune control and viral escape dynamics. We compare a lytic mechanism with a non-lytic mechanism that either blocks viral infection or inhibits viral production. Our findings show that a non-lytic control mechanism that blocks viral infection (such as in the case of RANTES) or restricts viral production (such as in the case of IFN-\(\gamma\)) results in significantly slower escape rate even when the lytic and non-lytic immune pressures are comparable in terms of both number of infections prevented and set-point dynamics. Importantly, we find that although the number of uninfected cells which are ‘protected’ by infection increases as the radius of secretion of soluble factor increases, the escape rate remains relatively constant and is significantly slower than the escape rate observed under a lytic mechanism. In line with the latter observation, in \[298\] the authors use a theoretical ODE model which also shows that if CD8+ T cells control virus replication non-lytically the rate of viral escape can be reduced although numerically the reduction can be dismal; however, the latter model is deterministic and assumes that non-lytic mechanisms can drive viral escape while the CA model allows us to formally prove this.

Our simulations further indicate that the advantage which the escape mutant has under a non-lytic control can be potentially attributed to the formation of clusters of infected cells. A wild-type cell that triggers the secretion of soluble factors is more likely
to be surrounded by other wild-type cells rather than variant cells, allowing therefore an outgrowth advantage to the variant strain. Interestingly, patches of infected cell clusters have been reported in *in situ* hybridization studies of SIV-infected cells [333].

Focusing on non-lytic control, we find that the efficiency of a non-lytic response in terms of infections prevented at steady-state is lower compared to the efficiency of a lytic response. We show that an increase of the duration of the non-lytic effect (from 10hrs to 20hrs) does not result in a stronger response, neither does an increase by three-fold of the effector population. This implies that although many uninfected cells are ‘protected’ or many infected cells are ‘blocked’, if the infected cells are not eliminated by the effector cells, there are still enough available targets to sustain the set-point viral burden. Hence, potentially very high numbers of effectors cells are needed for an efficient non-lytic suppression.

Our results are quantitatively and qualitatively very similar for non-lytic responses of blocking infection and blocking production. This is not surprising since in the model we focus on the chronic stage of infection where under the quasi-steady state assumption that virus production is proportional to the number of infected cells, the two types of suppression are expected to have identical dynamics. This can also be readily shown by the ODE models that describe the two different effector mechanisms (see Appendix F.5) and have been discussed in other studies [3].

Of course, since our model is theoretical it relies on parameter estimates which have been previously published. We vary many of these parameters in our simulations in order to check for the robustness of our results. In some cases, such as the motility behaviour of T cells and the lifespan of infected cells the estimates have been corrobo-rated by many studies. However, there are other parameters in the model that are not that well-defined such as the effect and radius of secretion of soluble factors. Here, we have varied the radius of the effect of the non-lytic response in order to examine how it impacts the immune control dynamics. Regarding the actual effect of the soluble factor, e.g. if it totally abrogates viral infection or production, little is known for most
of the cytokines and chemokines. Therefore, we have assumed a fully abrogating effect but as more data become available the model can be adjusted easily for the study of other factors and different effect sizes.

It is worth noting that the CA model can easily be used to investigate many aspects of the HIV/SIV dynamics. Specifically, it would be interesting to study the effect of the CD4+ T cell response on the course of the infection. In [334], the authors find evidence that the CD4+ T cell effector function can influence HIV/SIV dynamics during primary infection; although this would not alter the conclusions of this study since we mainly focus on chronic infection, it would be interesting to quantify -using the CA- the impact of a CD4+ T cell-mediated antiviral effect on the post-peak vireamic decline.

Our findings have implications for future studies of HIV progression and vaccine-induced responses. The appearance of escape mutants has been associated with loss of viral control and disease progression [335]. Our results indicate that non-lytic control can indeed slow down the appearance of escape mutants. In addition, in theory, a non cytolytic effector mechanism has three main advantages 1) can act on more than one target 2) can suppress without eliminating potentially vital cell populations and 3) can exhibit a ‘bystander’ protective effect [50]. Therefore, it can be argued that the tipping of balance towards non-lytic responses can be proven a strong weapon in the fight of HIV infection; however, an efficient non-lytic response mechanism that will not only lead to slower viral escape but also restrict the growth of the infected cell population might require a large number of effector cells. On the other hand, a lytic response can result in faster escape but also higher immune control suggesting that an appropriate combination of both responses can possibly offer the most efficient viral containment.
Chapter 8

General Discussion

As in every research project, some questions were answered and many more exciting ones arose. In this last part of the thesis, I would like to briefly summarise the findings presented and discuss their relevance and implications for further studies.

There are many persistent infections that have afflicted the human population over the years. Arguably, the best solution to persistent infections is the existence of prophylactic vaccines. However, for many viruses such as HCV, HTLV-1 and HIV this has not yet been possible. A key question that still remains largely unanswered is what differs between individuals that progress to persistent infection and those who remain asymptomatic or even successfully clear the virus (as in the case of HCV). The fact that persistent infection develops though complex interactions between the immune system and the pathogen makes it hard to decipher. Hence, we need to concentrate on a few key players at a time and try to understand their role in this interplay and how they fit in the answer to this question. Here, we focus mainly on CD8+ T cells because they have been repeatedly shown to play a major role in many viral infections providing therefore a valuable means of revealing potential drivers that lead to persistent disease.

Both the quality and the quantity of CD8+ T cell response to viral infection have been associated with viral control. However, when studying the role of CD8+ T cells in viral infection, one main problem occurs: causality. It is not possible, in the lack of elaborate longitudinal data, to determine whether the CD8+ T cell response is the cause or the consequence of variations in viral burden. One of the strongest evidence for the importance of CD8+ T cells in controlling viral infection is the association between HLA class I molecules and viral infection load and outcome. Hence, in order to
address the issue of causality, we studied the impact of HLA class I molecules, for which the direction of causality is unequivocal, on disease manifestation. In the context of HCV infection we tried to understand which HLA class I-related factors could explain why some individuals resolve the infection while others develop persistent disease. We found that although there were molecules that had a protective effect, namely B*57 and C*01, and there was a weak heterozygote advantage for HLA-B alleles, the overall predicted breadth, protein specificity and epitope specificity of the HLA class I molecules of individuals that resolved or persisted did not show any significant differences. This made us question the magnitude of the role of HLA class I molecules in determining the outcome of HCV infection. We quantified the effect using the explained fraction - a measure that estimates how much a given factor explains the heterogeneity of an outcome - and found that the HLA class I molecules associated with HCV infection outcome (for B*57, C*01 and C*04) could explain a total of 3% of the outcome whilst in HTLV-1 the fraction was close to 6% (for A*02, B*54 and C*08 together) and even larger for HIV infection where for B*57 alone the explained fraction was approximately 5%. This suggests that the most protective molecule in HIV or HTLV-1 infection might provide a larger benefit to the host compared to the most protective molecule in HCV infection or -and not mutually exclusive- the most detrimental molecule in HCV infection might not be as detrimental as in HIV or HTLV-1 infections. The fact that B*57 is a ‘protective’ molecule in both HIV and HCV infection can potentially form a basis for experimentally testing this finding by comparing the B*57-restricted CD8+ T cell response against HCV and HIV peptides in order to understand under a different benchmark what defines a strong ‘protective molecule’.

Next, we wanted to investigate whether there were other genetic factors that could potentially enhance this response. For this, we turned to Killer Cell Immunoglobulin-like receptors (KIRs) who are ligated by HLA class I molecules and have been associated mainly with NK cell responses. Interestingly, we found that a specific inhibitory
KIR, namely \textit{KIR2DL2}, could enhance both protective and detrimental HLA class I-mediated immunity. This was true for molecules involved in both HCV and HTLV-1 infections suggesting a more general underlying mechanism. Perhaps, the most surprising indication was that \textit{KIR2DL2} appears to enhance the downstream effect of CD8+ T cells rather than NK cells (although it might still be expressed on NK cells alone). This introduces a novel role for KIRs in adaptive immunity which should be explored further. We postulate that the expression of iKIRs directly on NK or CD8+ T cells can eventually increase CD8+ T cell survival. Ligation of iKIRs on CD8+ T cells can upregulate pro-survival molecules and reduce activation induced cell death (AICD) \cite{79,80,227,228}. Alternatively, ligation of iKIRs on NK cells can downregulate the killing of activated T cells \cite{237, 238}. In both cases, CD8+ T cells would survive longer in the presence of inhibitory KIRs and thus their protective and detrimental associations would be enhanced. In order to better understand the underlying mechanism we propose to study the effect of \textit{KIR2DL2} (and also other KIRs) on CD8+ T cell selection pressure. The latter can be quantified using sequence data and estimating the number of coding changes within epitope regions. If inhibitory KIRs are found to tune CD8+ T cell responses it would provide a new way of modulating viral control. Of particular interest are studies suggesting that strategies to block inhibitory receptors such as iKIRs may be of potential use in increasing the efficacy of immunotherapy and this has already been achieved in simple mice models of cancer \cite{336,337}. Additionally, it would be important to test whether HLA class I-mediated protection against other viral infections or viral-associated malignancies can be enhanced by KIRs.

Until now our study focused on the immunogenetics of persistent viral infections, in terms of HLA and KIR genes, with a main focus on their impact on CD8+ T cell responses. However, we also aimed at investigating the different manifestations of these responses (lytic versus non-lytic activities) and how they relate to viral dynamics. For this part of the study, we chose to concentrate on HIV/SIV infection since recent studies \cite{42,167} of SIV infection challenged the cytolytic role of CD8+ T cells and pointed
towards a non-lytic control that can lead to viral decline without reducing the lifespan of productively infected cells. These two effector functions differ in that the secretion of soluble factors (non-lytic response) can have both local and systemic consequences, whereas cell lysis is restricted to the elimination of the direct target. So, the first question that rises is whether a non-lytic response could explain the findings of these recent studies and what is the potential anti-viral efficiency of non-lytic mechanisms compared to lytic mechanisms? If we also take into account the observation that HIV and SIV escape the CD8+ T cell control frequently, some additional important questions are posed. Is the consistent observation of viral escape proof that HIV-1-specific CD8+ T cells lyse infected cells? If CD8+ T cells control SIV primarily by non-lytic mechanisms then why is viral escape observed?

On a first approach, we studied whether lytic and non-lytic responses can explain the data observed in [42]. We focused on non-lytic effector functions that block infection of uninfected cells or inhibit viral production by infected cells. We fitted a number of ODE models to the experimental data obtained from SIV-infected macaques and found some evidence that non-lytic control, particularly the one that blocks viral infection, can better explain the data and describe the underlying CD8+ T cell response. Nevertheless, as discussed in [3] more accurate measurements of immunological data are needed before firmer conclusions are drawn.

It should be emphasised that lytic and non-lytic responses are not mutually exclusive. The inherent ability of viruses to induce variable antigenic stimulation may be a determinant of the lytic versus non-lytic virus-specific CD8+ T cell responses [338,339]. Additionally, different markers expressed on the cell surface can also regulate antiviral CD8+ T cell effector activities [340]. However, in order to understand how these two different functions alter viral dynamics it would be ideal to study them separately. This cannot easily be tested in vivo, it can however be explored in silico. Hence, we constructed a detailed computational model, namely a 3D cellular automaton, so that we can simulate lytic and non-lytic CD8+ T cell responses and study how they modulate
virus-infected cell dynamics. Our main aim was to understand how non-lytic responses compare in efficiency to lytic responses and whether a non-lytic response can result in the outgrowth of the cell population that is infected with a variant strain; this is after all one of the strongest consequences of CD8+ T cells acting in a lytic manner. Our approach although theoretical is based on many biological parameters and processes that have been quantified experimentally. As in every theoretical model assumptions needed to be made in order to ease the analysis as well as the interpretation of the results. The 3D cellular automaton model of HIV/SIV dynamics produced quantitative and qualitative results that are very consistent with experimental observations. Importantly, the flexibility of this model allows for the future study of other important questions such as target cell competition, early CD8+ T cell killing, simulation of other non-lytic soluble factors and CD8+ T cell dysfunction.

We first studied the CD8+ T cell lytic response and found that for a CD8+ T cell to eliminate 1-5% of the infected targets per day, it needs to successfully recognise on average one infected target every 1000 down to 300 encounters respectively. It would be interesting to investigate whether this observation could be corroborated by the use of an ODE model that takes into account the encounter rate of CD8+ T cells with their infected targets. We also found that a model of lytic killing that is described by a saturating behaviour in terms of infected targets can better explain the killing process. This suggests that there are limiting factors involved in CD8+ T cell killing that would cause the mass-action behaviour in terms of the infected targets to collapse; however, we find that a higher number of effector cells results in a higher killing rate and this relationship can be consistent with a mass-action behaviour in terms of effector cells. Our findings further suggest that one of the factors that might restrict killing in terms of infected targets can be the need to identify the target before actually eliminating it while the duration of killing per se has a dismal impact on the killing rate. Although important two-photon microscopy studies of CD8+ T cell killing [301,305] allow the measurement of the time that CD8+ T cells spent in delivering the lethal hit to antigen-
pulsed cells and record the time they browse the pulsed or unpulsed targets, they do not capture information on the time that a CD8+ T cell takes before it successfully locates its target. Based on our findings, this type of knowledge would facilitate a better understanding of what may lead to the restriction of the CD8+ T cell killing rate. Interestingly, other studies have also indicated a saturation killing term as a better description of a lytic CD8+ T cell response and although there is still a debate on the subject it can be potentially explained by the fact that many studies are being performed on different viruses and under different experimental setups.

Next, we investigated the efficiency of the two different CD8+ T cell effector mechanisms and how they can shape the dynamics of viral escape. Clearly, production of soluble factors could lead to viral escape only if it is MHC-dependent; but even if this is the case, can viral escape be observed? We found that lytic CD8+ T cells could reproducibly drive rapid viral escape. Escape from non-lytic CD8+ T cells (both those that block infection and those that block production) was also repeatedly observed however it was less frequent and slower than in the lytic case. We also found that the formation of infected cell clusters of the same type (wild-type or variant) could explain the advantage of the variant strain under a non-lytic mechanism. Overall, we saw that the non-lytic control is less efficient than the lytic control in terms of infections prevented and set-point viral burden. Furthermore, the non-lytic effect did not strengthen with increasing number of effector cells (while the lytic effect did) or when the duration of the abrogating effect (blocking infection or production) was extended. These findings taken together suggest that viral escape is consistent with both lytic and non-lytic CD8+ T cell responses. However, a non-lytic response although it can lead to slower and less frequent viral escape compared to a lytic response, it might not be as efficient in controlling the viral burden and may require very large numbers of effector cells in order to exhibit high viral suppression. Our findings also provide a potential reconciliation between studies that report rapid and slow escape by showing that the underlying CD8+ T cell effector function can be a main driver of the observed
differences. It would be very important, especially for future T cell-based vaccine research, to explore such a possibility experimentally by recording longitudinally the epitope-specific CD8+ T cell lytic and non-lytic profile along with the escape rate of different viral epitopes.

As a concluding remark, I would like to note that a large amount of this work was the result of a productive collaboration between theoreticians and experimentalists and can be considered as a small indication of how these disciplines can work alongside each other. Hopefully, it also shows how this can accommodate the advancing of our knowledge of the immune system and biology as a whole. As J. Cohen insightfully said [341] ‘Mathematics Is Biology’s Next Microscope, Only Better; Biology Is Mathematics’ Next Physics, Only Better’.
Appendices
A Appendix A

A.1 The HCV genome and gene products

Appendix Figure A.1: The HCV genome and gene products. HCV is positive-sense single-stranded RNA virus. This viral genome contains a single open reading frame (ORF) encoding a polyprotein, processed into at least 10 proteins: three structural proteins and seven non-structural proteins [342]. The image is taken from [342].
A.2 The HCV-1a protein sequence (isolate H77)

>F
MSTNPKQRKPVPTVAHRNTSSRRVAVRSLVEFTCCRAGALDWVCARRGRLPSGRNLNVDSLSPRHVG
PRAGPGLSPGTLGSPMAMRVAGGGRDSCLPVALGAPQTPVGRAIWVSSITPLRAASPTSWGTYRSS
APLLEALPGPWRMASGFWKTA

>core
MSTNPKQRKTNRNRPQDVKFPGQQIVGVYLLPRGPRPLGVRATKTSERSQPRGRQQPIPKARR
PEGRTWAGPQPYWLYGNECGWAGWLLSPRGRSPWGTDPARKRSNLGKVIDTLCGFADLMGYIPLV
GAPLGGAAARAHGVRVLEDGVNYATGNLPGCSFISLLALLSCLTVPASA

>E1
YQVRNSSLGLYHTDCPNSSSYVEAADAILHTTPCVPCVREGNASRCWVAATPTVATRDGKLPPTTLRHH
IDLLVGSATLCSALYGDLGSCVFLVGQLFTFSPRRHWTQDCNCSIYPGHITGHRMAWDMNNWSPTAA
LVVAQLLRIPQAIMDIAHAGHVGLAIGAYFMSMNVKLVVLVLFLAGVDA

>E2
ETHVTGGSAGRTTALGVLLTPGAKQNIQLINTNGSWHIINSTLANCNESLNVTGWLAGFLFYHKFNSGCP
ERLASCRRLTDFAQGWGISYANSGLDERPYCWHYPPRPGVPASKVCGPVYCFTPSVVPVVGTDRSG
APRTYSWAGNTDFVLLNPPLGNWMGSTGFTKVCGAPPVIGGPNNTLCTPTCFRKHPPEAT
YSRCGSGPWITPRCMVDPYRPLYWHYPCTINTYIFKVRMYVGGVEHRLEAANWTRGERCDLEDLDRESLS
PLLLSTTQWQLPSCSFTTLPALSTGLIHHLHHQVINIVDQYLYVGSSIASWAIKWEYVLLFLLLADARVCS
CLWMLMLISQAE

>P7
ALENLVILNAASLAGTHGLVSFLVFCCAFOWLYKGRVFGAYAFYGMWPLLILLALPQRAYA

>NS2
LDTEVAASCWVLVGLMALTSLPSYKRYISWCMWWLQYFLLRVEAQLHVWVPLNVRGGRDAVLLMCV
VHPILVFDITKLLLAIIFGPLWILQRASLKLKVPVFVRQQLLRICALARKIAGGHYVMAIIKLGALTGYV
YNHLTPLRLDWAHLNRDLAVAVEPVVFSRMETKLTWGDATACDINGLPSARRGQIELLPADGMV
SKGWRLL
VKAAASKVKANLLSVEEACSLTPPHSASKFGYGA<KDELVRCHARAKAVAHINSVWKDLLEDSVTPIDTTIMA
KNEVFCVQPEKGRKPARLIFPDGLGVRCEKMALYDVVS<KLPLAVMGSSSYGFYSPQGRVEFLVQAWKS
KKTPMGFSYDTRCFDSTVDESIRTEAAIYQCDDLPQARVAIKSLTERLYVGGLTNSRGENCGYRRCR
ASGVLTTSGNTLCTYKARACRAAACLQDCTMVCGDDLVVICESAGQ<DAASLRRAFTEAMTRYSAPP
GDPQQPEYEDELITSCSSNVSHAHDGAKRKVVYYL<RPPLARAAWETARHPTVNSLGNIMFAPTLW
ARMILMTHFFSVLIARDQLEQLNCIEYGACYSIEPLDLPPIIQRLHGLSAFLASHPSGEINVAACLRL
KLGVPPLRAWHRARSVRARLLSGRAAGICLYKLFNWAVR<KLTLTPIAAAGRLDSGWFTAGYSGGDI
YHSVSHARPRWFCLLLLLAAVGVIYLLPNR

A.3 The HTLV-1 protein sequence

The HTLV-1 reference strain is from [343] with the exception of HBZ, which was more recently described in [344].

>Gag
MGQIFSR<ASPPIPPRPLAAHHLNW<FLQAAYRLEPGSSYDFHQL<KFLKIALETPARICPNYSLLAS
LLPKGYPGRVNEILHILIQTQAQPSPRPAPPPSSPTPHDDPDSPQIPPPYE<PTAPVLPVMHHPGAPP
NHRP<MDLQAIKQEVSAQAPGGQFMQTIRLAVQFDPTAK<QLQDDLQYLCSSLVASLHHQQLDSLIS
EAETRGITGYNPLAGPLRVQANIPQQQGLREYQQLWLAFAALPGASKDPSASILQGLEEPYHA<VER
LNI<ALNGLPEGTPKDPIRLSAYSNANKECQKLQARGHTNSPLGMRLACQTWTPDKTKLVVVQPKK
PPNPQCFCRC<AGHWSRDCTQP<PPPPGCPLQDPTHKWDCPRLRKPTEPEEEAD LLDDLADIPKH
PKNFIGGEV

>Env
MGKFLATLLLFQFCPLFGYDSPCCTLTITGVSSYHSDKPCNPAPVCGWTLLLALSADQALQPPCPNL
VSYSSYHATSYLFPHWT<PKPNRNGGGYYSADCP<SLKCPYLGQSWTCPYGAVSSYPKFQHDVN
FTQEVSLNINLHFSKCGFPLSSLLDAPGDPF<WIINEQSLQ<APLLPHSNI<LDHILESPWKS<
LTLVQLTLQSTNYCIV<DRASLT<WHLYSPNSVPSSTPPPLLPSLAP<HLTPFNFWHCFDPQ
IQAIVSSPCHSSLILPFPSPLSPVPTLGSRSRRAVPAVWVLVSA<AMAGV<AGITGSM<ALSGKSSLHEV
DKDIIQLTFQAIVKNHKNLL<IAQYQAAN<RRGLLLL<WEQGGLCKALQECQCRFFPNI<NSH<VPILQERPLL
NRVLTGWGLNWDGLGSLQWAREALQTGITLVA LLLVLILAGPCILRQLRHLPSRVRYPHYSLIKPESSL

>Pro
HPTPKKLHRGGGLTPPTLTQQVLPQDAPASILPVIPDLPARRPVICKAQVDQTSHPKTIEALLDGTADM
VLPIAFSSNTPLKNTSLVGAGGQTDQHFKTLTSPLVIRLPFRTPVILTSCLVDKNNWAIIGRDALQQ
CQGVLYLPEAKRPFVILPIQAPAVLGLEHLPRPPEISQFPLNQNASRCPNTWSGFRWQAISNTPPGQGI
TQYSQKLRPMEPDSSTTCGPLTL

>Po
GKKAACNLANTGASRPWARTPPKAIPQPPQFPKPERLQLQALQLHVRKALEAGHEIPYTGPNGNPVPVFKKA
NGTWRFIHDLRTALSTLTSSLSSSPGPDSSLPTTLAHQTLTDLRDAFFQIPLPKQFQPYFAFTVPQQC
NYGPGRYAWKVLQPFKNSPTLFEMQLAHILQPRQAFPQCTULTQYMDDILASPSHEDLLLLEATMA
SLISHGPLVSENKTTQPGTIGKQLQIISPNHITDARYPTVRIPSRWALPQLAQGELIQUIVSKGTPTLR
QPLHSYCALQRTDPRDQIYLPSQVLQLRQLASQNSRSLVQRLTTLP.SHGALMLTGGTGTTVVFQSK
EQWPLVLWHLAPLHPSTQCPWQQLASAVLLDKYTLQYGLCQTIHHNISTQFQIQTFQSDHPSVPLIL
LHSHRFKNLGATGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSGLSTSAAYILWDKQIL
SQRSFPLPPHKSAQAELLLHLSSRSSWCLNIFLDKSLHYLRLALTGFTQGRSSQAPFQALLP
RLLSRKVYLHHRVSLNPLDPISRLNALTDALLIPVQLQSPAELHFSFTHCQQTALTQGATTTEASNI
LAASCHACRGGNPQHMRPGHRRGGLPLLHWWQGITHFKYKNTLYRLHVVWDTFSGAIASATQKRTESSE
AISLLQAIAHLGKPSINTDNGPAYSQDIFNMCTSLAIRHTTHVYPNTSSGLVERSNGILKTLLYKY
FTDKPDLPMNALSIALWTINHLVNTCHKTRWLHHSPLQLQIPETRSLSNKNQTHWYFKLPGLNSRQ
WKGPQEALEAQGAAALIPVSASSAQWIPWRLLKRAACPVRPPGPADPKEDLQLQHG

>Rof
MPKTRRRPQRSQRRPKTPWQLLLPSLQGQLHAFQSLSSIAINPQLLLHFFFSMTLFLRLLSLPSPALTAL
LLFLPPPSDVSSLRLPPAPCLLLFLPFQLSGLLFLFLFLPLFFSLPLLSSLPSLPITMRFPARWRFLPW
RAPSQPPAAFLF

>P12
MLFRLLSPLSPALTALLLLFLPPPSDVSSLRLPPAPCLLLFLPFQLSGLLFLFLFLPLFFSLPLLSSLPSL
SLPITMRFPARWRFLPWAPRPSQPAAALF

>Tof
MALCCFAFSAPCLHLRSRRSCSSCFLLATSAFFSARLLRRAFSSSFLFKYSAVCFSSSFSSRFFRLFS
SARRCRSRCVSPRGAFSPGPGRSRRPLSSKDSKPSSTASSSSLFSNSSKDNPSNSTSSTSRSSGHID
TGKHNSPAPDTKLTLISPLMRWTESSFRIISLRVRLCTRRLVPHLGTMGFPPPSTSSRPTGHLSRAS
DHLGHPHRWTRYRLSSTVPYFSTPPLLPHENL
>P13
MLIISPLPRWTESSFRIISLRVRLCTRRLVPHLGTMGFPPPSTSSRPTGHLSRASDHLGHPHRWTRYRLS
STVPYFSTPPLLPHENL
>Rex
MPKTRARPRRSQRKRPPTWPSTSQQGLDRVFFSDTQSTCLETVYKATGAPSLGDEVVPAYIVTPYWPPVQS
IRSPGTPSMDALSAQLYSSLDPSPPAPREPLRPSRSLPRQSLIQPPTFHPSSRPCANTPPSEMĐTWN
PPLGSTSQPCLFTQTPDGPKTCTPSGEAPLSACTSTSFPFPSPGPSCPT
>P21
MDALSAQLYSSLDPSPPAPREPLRPSRSLPRQSLIQPPTFHPSSRPCANTPPSEMĐTWNPPPLGSTSQ
PCLFTQTPDGPKTCTPSGEAPLSACTSTSFPFPSPGPSCPT
>Tax
MAHFPGFQQSLFFGPVYVFGDCVQGDCPISSGLCSARLHRHALALTCEHPQITWDPIDGRVIGSAQF
LIPRLPSFPTQRTSSTLKLVPITTHPNPISFGMRLQAMRKYSFRRNGYMEPTLGQLHPTLSFDPGLRP
QNLTYLWGGSVMYLYLQSLPITWPLLHPVFCHPGQLGAFLTNVPKRIEELLYKISLTTGALIIPE
DCLPTTTLFQPARAPVTLTAWQNLFPFSTLTTPGLTWTFTDTPMISGPPKDGQPVLQQLQSFIFHK
FQTAYHPSFLLSHG-liQYSSFHLHLLFEEYTNIPISLLFNEKEADNDHEPQISPQQPLEPSEQHFREV
TEV
>HBZ
MAASGLFRCLPVCPSDDLLVEELVDGLSLEEELKDKEEEEKAVLDGLSLSEESRGLRRGPPGEKAPP
GETHHRDQRAAEKRKRKKEREKKEEKQIAEYLKRKEEKARRARRAEKKAADVARKQEERQERRKWR
QGAEEKQHSARKEMQMELGIDGYTRQLEGESLEAERMKLQQEKEDLMGEVNYWQGRLEAMWLQ
A.4 HLA class I associations with HCV infection outcome

In Table A.1 we have recorded HLA alleles that have been associated with the outcome of HCV infection. Other factors, such as differences in the ethnic and geographical background of the infected individuals have also been shown to influence the outcome of the infection [28]. Hence, whenever possible we have identified the country where the study took place and the prevalent HCV genotypes in the given cohort. In addition, we have taken into account the size of the cohort so that we can build a measure confidence in the power of each analysis. However, it needs to be emphasised that conclusive association studies regarding the influence of HLA on infectious disease require large samples, proper ethnic-background stratification, accurate clinical information, and use of models that consider other known genetic effects on the disease [16].
<table>
<thead>
<tr>
<th>Allele</th>
<th>Effect</th>
<th>Country(G)</th>
<th>Cohort size</th>
<th>Method</th>
<th>Mediation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*02</td>
<td>P</td>
<td>USA(NC)/-</td>
<td>105(-/49/56)</td>
<td>OR</td>
<td>CTLs</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>USA(C)/-</td>
<td>105(-/49/56)</td>
<td>OR</td>
<td>CTLs</td>
<td>[25]</td>
</tr>
<tr>
<td>A*03</td>
<td>P</td>
<td>Ireland/1b</td>
<td>243(-/95/148)</td>
<td>OR</td>
<td>CTLs</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>USA(NC)/-</td>
<td>105(-/49/56)</td>
<td>OR</td>
<td>CTLs</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Korea/1b,2a</td>
<td>343(206/-/137)</td>
<td>RR</td>
<td>-</td>
<td>[33]</td>
</tr>
<tr>
<td>A*34</td>
<td>LVL</td>
<td>Taiwan/1b</td>
<td>160</td>
<td>WRST</td>
<td>-</td>
<td>[202]</td>
</tr>
<tr>
<td>A*19</td>
<td>D</td>
<td>Ireland/-</td>
<td>268(122/-/146)</td>
<td>-</td>
<td>-</td>
<td>[34]</td>
</tr>
<tr>
<td>A*2301</td>
<td>D</td>
<td>USA/-</td>
<td>675(-/-/4444)</td>
<td>OR</td>
<td>CTLs</td>
<td>[24]</td>
</tr>
<tr>
<td>B*56</td>
<td>LVL</td>
<td>Taiwan/1b</td>
<td>160</td>
<td>WRST</td>
<td>-</td>
<td>[202]</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[29]</td>
</tr>
<tr>
<td>B*1401</td>
<td>HVL</td>
<td>Taiwan/1b</td>
<td>160</td>
<td>WRST</td>
<td>-</td>
<td>[202]</td>
</tr>
<tr>
<td>B*14</td>
<td>D</td>
<td>Italy/-</td>
<td>606(489/-/117)</td>
<td>RR</td>
<td>-</td>
<td>[24]</td>
</tr>
<tr>
<td>B*54</td>
<td>D</td>
<td>Japan/1b,2a,2b</td>
<td>104(916/-/130)</td>
<td>OR</td>
<td>CTLs</td>
<td>[335]</td>
</tr>
<tr>
<td>B*55</td>
<td>D</td>
<td>Japan/1b,2a,2b</td>
<td>25(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*51</td>
<td>P</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*52</td>
<td>P</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*53</td>
<td>D</td>
<td>Korea/1b,2a,2a</td>
<td>343(206/-/137)</td>
<td>RR</td>
<td>-</td>
<td>[335]</td>
</tr>
<tr>
<td>B*46</td>
<td>D</td>
<td>Korea/1b,2a,2a</td>
<td>343(206/-/137)</td>
<td>RR</td>
<td>-</td>
<td>[335]</td>
</tr>
<tr>
<td>A*1101</td>
<td>P</td>
<td>USA/-</td>
<td>675(-/-/4444)</td>
<td>OR</td>
<td>CTLs</td>
<td>[24]</td>
</tr>
<tr>
<td>B*57(01,03)</td>
<td>P</td>
<td>USA/-</td>
<td>675(-/-/4444)</td>
<td>OR</td>
<td>CTLs</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Ghana/2</td>
<td>104(-/35/37)</td>
<td>OR</td>
<td>CTLs</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>USA/1a,b</td>
<td>758(-/-/622)</td>
<td>OR</td>
<td>CTLs/NK</td>
<td>[33]</td>
</tr>
<tr>
<td>Cw*04</td>
<td>D</td>
<td>USA/-</td>
<td>675(-/-/4444)</td>
<td>OR</td>
<td>CTLs</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Ireland/1,3,5,2</td>
<td>225(-/86/130)</td>
<td>OR</td>
<td>CTLs/NK</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*07</td>
<td>P</td>
<td>Ireland/1b</td>
<td>243(-/95/148)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*27</td>
<td>P</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>Cw*01 (02)</td>
<td>P</td>
<td>Ireland/1b</td>
<td>243(-/95/148)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Japan/1b,2a,2b</td>
<td>285(172/-/113)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>USA/-</td>
<td>675(-/-/4444)</td>
<td>OR</td>
<td>CTLs</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>USA/1a,b</td>
<td>758(-/-/622)</td>
<td>OR</td>
<td>CTLs/NK</td>
<td>[33]</td>
</tr>
<tr>
<td>B*18</td>
<td>D</td>
<td>Ireland/1b</td>
<td>243(-/95/148)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td>B*08</td>
<td>D</td>
<td>Ireland/1b</td>
<td>243(-/95/148)</td>
<td>OR</td>
<td>CTLs</td>
<td>[202]</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Saudi/-</td>
<td>268(122/-/146)</td>
<td>-</td>
<td>-</td>
<td>[32]</td>
</tr>
<tr>
<td>Cw*05</td>
<td>P</td>
<td>USA(NC)/-</td>
<td>105(-/49/56)</td>
<td>OR</td>
<td>CTLs</td>
<td>[25]</td>
</tr>
</tbody>
</table>

Appendix Table A.1: HLA class I associations with HCV outcome reported in the literature. Protective (P) or Detrimental (D), Low (LVL) or High (HVL) Viral Load, Odds Ratio (OR), Relative Risk (RR), Wilcoxon Rank Sum Test (WRST), Caucasians (C) and Non-Caucasians (NC).
A.5 Confounding factors in HCV infection

<table>
<thead>
<tr>
<th>Variable</th>
<th>UK(161)</th>
<th>USA(621)</th>
<th>Pooled(782)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race(Caucasians)</td>
<td>-</td>
<td>0.70(0.24)</td>
<td>0.73(0.29)</td>
</tr>
<tr>
<td>Mode(IVDU)</td>
<td>0.11(0.009)</td>
<td>1.18(0.58)</td>
<td>0.13(0.02)</td>
</tr>
<tr>
<td>Mode(Transfusion)</td>
<td>0.07(0.009)</td>
<td>-</td>
<td>0.08(0.01)</td>
</tr>
<tr>
<td>HBV</td>
<td>-</td>
<td>3.85(1.7 x 10^-5)</td>
<td>3.79(1.9 x 10^-5)</td>
</tr>
<tr>
<td>HIV</td>
<td>-</td>
<td>0.76(0.15)</td>
<td>0.76(0.15)</td>
</tr>
<tr>
<td>rs12979860</td>
<td>1.63(0.18)</td>
<td>3.09(3 x 10^-9)</td>
<td>2.76(2.5 x 10^-9)</td>
</tr>
<tr>
<td>KIR2DL3/2DL3+HLA-C1C1</td>
<td>2.64(0.03)</td>
<td>1.58(0.07)</td>
<td>1.76(0.009)</td>
</tr>
</tbody>
</table>

Appendix Table A.2: Impact of potential confounding variables on the outcome of HCV infection.

A.6 The Explained Fraction depends on factor frequency

In order to show that the explained fraction depends on factor frequency, we constructed many 2x2 contingency tables keeping the frequency of the outcomes (column marginals) fixed. We then computed the Odds Ratio, the Explained fraction and the frequency of the causal factors and plotted the variation of Explained Fraction with respect to frequency for the same factor effect i.e. for the same OR. In Appendix Figure A.2 we readily observe that the higher the effect of a causal factor on disease outcome (OR ≠ 1) the higher the impact of the factor frequency on the explained fraction. We also observe, that the maximum EF is reached in all cases for factor frequency 0.5. This is expected because the EF depends on mutual information and we have the maximum mutual information when the frequencies of the causal factors are equal i.e. each factor has frequency 0.5.
Appendix Figure A.2: For the same Odds Ratio i.e. the same factor effect, the Explained Fraction depends on the frequency of the factor.
Appendix B

B.1 False Discovery Rate of cohort stratifications

<table>
<thead>
<tr>
<th>Result</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*57 status</td>
<td>6%</td>
</tr>
<tr>
<td>B*54 status</td>
<td>&lt;0.0001%</td>
</tr>
<tr>
<td>C*08 status</td>
<td>&lt;0.0001%</td>
</tr>
<tr>
<td>B<em>54+C</em>08 status</td>
<td>&lt;0.0001%</td>
</tr>
<tr>
<td>B*54 PVL HAM/TSP</td>
<td>15%</td>
</tr>
<tr>
<td>C*08 PVL HAM/TSP</td>
<td>4%</td>
</tr>
<tr>
<td>C*08 PVL ACs</td>
<td>15%</td>
</tr>
<tr>
<td>B*57 VP MHCS</td>
<td>5%</td>
</tr>
<tr>
<td>B*57 VL ALIVE</td>
<td>1%</td>
</tr>
<tr>
<td>HBZ binding</td>
<td>1%</td>
</tr>
</tbody>
</table>

Appendix Table B.1: False Discovery Rate of cohort stratifications. We examined how many times we would see odds ratios equal to or more extreme than we observed in the actual cohorts (FDR).

B.2 Canonical KIR-HLA binding not supported by linkage effects

HTLV-1: other group C1 HLA alleles do not exhibit the same behaviour as C*08

We found that, in HTLV-1 infection, HLA-C*08 was associated with a protective effect: reducing the risk of HAM/TSP and reducing proviral load in ACs. This effect was
enhanced in the presence of KIR2DL2 but reduced/absent in the absence of KIR2DL2. We hypothesised that the effect of KIR2DL2 on the C*08 protective effect is not mediated via a canonical, direct KIR-HLA interaction. To test this hypothesis we first investigated whether the other group C1 alleles exhibited the same behaviour as C*08 in HTLV-1 infection. Virtually all HTLV-1 infected individuals possess at least one C1 allele (430/432) therefore it was not possible to look at presence or absence of the C1 ligand instead we looked at the impact of C1 homozygosity. Grouping all the C1 alleles we find no significant association between C1 homozygosity and disease status either in the whole cohort or in the context of KIR2DL2. Similarly, C1/C1 was not associated with decreased proviral load in either ACs or HAM/TSP patients; nor did this become significant in the context of KIR2DL2.

In case the failure to find that the C1 grouping behaved the same as C*08 was because we were forced to looked at homozygosity for C1 rather than presence or absence as we had done for C*08 we additionally, investigated the individual C1 alleles (Table B.2). This confirmed the hypothesis that the protective C*08 effect and its enhancement by KIR2DL2 we had observed was not exhibited by other group C1 alleles.

**HTLV-1: the B*54 effect cannot be attributed to linkage with C*01**

B*54, a group Bw6 HLA allele, is not known to bind any KIR and so we tested whether the observed B*54 effect was attributable to C*01 which is in linkage disequilibrium with B*54 and does encode molecules which bind KIR2DL2. We found that B*54 rather than C*01 appeared to be the primary gene associated with HAM/TSP whose detrimental effect was enhanced by KIR2DL2 (see HLA linkage).
### (a) Disease status

<table>
<thead>
<tr>
<th>HLA-C1 Allele</th>
<th>OR (whole cohort)</th>
<th>KIR2DL2 Genotype</th>
<th>OR (stratified cohort)</th>
<th>p-value</th>
<th>Allele Carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*01</td>
<td>0.768</td>
<td>+</td>
<td>0.307</td>
<td>0.028</td>
<td>43</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.296)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*03</td>
<td>1.683</td>
<td>+</td>
<td>1.331</td>
<td>0.561</td>
<td>51</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.038)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*07</td>
<td>0.824</td>
<td>+</td>
<td>0.335</td>
<td>0.07</td>
<td>26</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.495)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*08</td>
<td>2.126</td>
<td>+</td>
<td>6.249</td>
<td>0.02</td>
<td>14</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.032)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*10</td>
<td>1.122</td>
<td>+</td>
<td>1.888</td>
<td>0.342</td>
<td>15</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.794)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*14</td>
<td>0.576</td>
<td>+</td>
<td>1.447</td>
<td>0.553</td>
<td>21</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>(p=0.073)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (b) Proviral load

<table>
<thead>
<tr>
<th>HLA-C1 Allele</th>
<th>Difference in VL (whole cohort)</th>
<th>KIR2DL2 Genotype</th>
<th>Difference in VL (stratified cohort)</th>
<th>p-value</th>
<th>Allele Carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*01</td>
<td>-0.164</td>
<td>+</td>
<td>-0.358</td>
<td>0.233</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.234)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*03</td>
<td>0.113</td>
<td>+</td>
<td>-0.047</td>
<td>0.866</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.407)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*07</td>
<td>0.222</td>
<td>+</td>
<td>0.408</td>
<td>0.23</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.187)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*08</td>
<td>-0.33</td>
<td>+</td>
<td>-0.662</td>
<td>0.053</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*12</td>
<td>0.07</td>
<td>+</td>
<td>-0.134</td>
<td>0.716</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.675)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*14</td>
<td>0.008</td>
<td>+</td>
<td>0.164</td>
<td>0.614</td>
<td>11</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(p=0.964)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### HAM/TSP

<table>
<thead>
<tr>
<th>HLA-C1 Allele</th>
<th>Change in VL (whole cohort)</th>
<th>KIR2DL2 Genotype</th>
<th>Change in VL (stratified cohort)</th>
<th>p-value</th>
<th>Allele Carriers</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*01</td>
<td>0.183</td>
<td>+</td>
<td>0.402</td>
<td>0.011</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.033)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*03</td>
<td>-0.169</td>
<td>+</td>
<td>-0.047</td>
<td>0.775</td>
<td>26</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.049)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*07</td>
<td>-0.057</td>
<td>+</td>
<td>-0.074</td>
<td>0.682</td>
<td>16</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.554)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*08</td>
<td>-0.173</td>
<td>+</td>
<td>-0.943</td>
<td>0.002</td>
<td>4</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.208)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*12</td>
<td>-0.065</td>
<td>+</td>
<td>-0.175</td>
<td>0.472</td>
<td>7</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.965)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C*14</td>
<td>0.085</td>
<td>+</td>
<td>-0.019</td>
<td>0.923</td>
<td>10</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(p=0.408)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix Table B.2: Impact of HLA-C1 alleles on HTLV-1 disease status and proviral load. No other HLA-C1 alleles have the same effect with C*08 in the context of KIR2DL2.
HCV: the $B^*57$-$KIR2DL2$ effect cannot be attributed to $B^*57$ linkage with HLA molecules that do bind KIR2DL2

$HLA-B^*57$ is in linkage with 3 HLA class I alleles: $A^*01$, $C^*18$ and $C^*06$. None of these HLA molecules is expected to bind KIR2DL2; furthermore $A^*01$, $C^*06$ and $C^*18$ do not have a significant impact on HCV status either overall or in the context of KIR2DL2. We therefore conclude that the $B^*57$-$KIR2DL2$ effect cannot be attributed to $B^*57$ linkage with HLA molecules that do bind KIR2DL2.

HCV: the $B^*57$-$KIR2DL2$ effect cannot be attributed to KIR2DL2 linkage with other KIRs that do bind $B^*57$

$HLA-B^*57$ does not bind KIR2DL2 however $B^*57$ does bind KIR3DL1 and possibly KIR3DS1 both of which are in weak linkage disequilibrium with KIR2DL2 ($p=0.04$ and $p=0.1$ respectively). We therefore examined whether the KIR2DL2 enhancement of the $B^*57$ protective effect could instead be explained by KIR3DL1 or 3DS1. Four observations argue against this:

1. If $B^*57$ binding of KIR3DL1 or 3DS1 was associated with enhanced immunity then other HLA-B ligands of KIR3DS1 or KIR3DS1 with similar binding to $B^*57$, namely Bw4.80I, would be expected to show a similar pattern to $B^*57$. This was not observed (Table B.3).

2. Examining the enhancement of $B^*57$ by KIR2DL2, KIR3DL1 and KIR3DS1 it can be seen that the strongest enhancement is by KIR2DL2 (B.4).

3. If we exclude KIR2DL2+ individuals then neither KIR3DL1 nor KIR3DS1 enhance $B^*57$ (in KIR2DL2-3DL1+ OR =0.83 $p=0.63$. KIR2DL2-3DS1+ OR =0.98 $p=0.97$). However there are only 11 KIR2DL2-3DS1+B*57+ individuals (35 KIR2DL2-3DL1+B*57+) so loss of significance in the latter case may be attributable to power. Conversely, if we exclude KIR3DS1+ individuals then
<table>
<thead>
<tr>
<th>KIR</th>
<th>HLA ligand</th>
<th>OR</th>
<th>p-value</th>
<th>N KIR+HLA+</th>
<th>Cohort size</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DL1</td>
<td>Bw4</td>
<td>0.94</td>
<td>0.7</td>
<td>485</td>
<td>782</td>
</tr>
<tr>
<td>3DL1</td>
<td>Bw4.80I</td>
<td>0.88</td>
<td>0.5</td>
<td>291</td>
<td>782</td>
</tr>
<tr>
<td>3DL1</td>
<td>Bw4/Bw4</td>
<td>1.19</td>
<td>0.5</td>
<td>131</td>
<td>782</td>
</tr>
<tr>
<td>3DL1</td>
<td>Bw4.80I/Bw4.80I</td>
<td>1.85</td>
<td>0.1</td>
<td>47</td>
<td>782</td>
</tr>
<tr>
<td>3DS1</td>
<td>Bw4</td>
<td>0.93</td>
<td>0.7</td>
<td>159</td>
<td>782</td>
</tr>
<tr>
<td>3DS1</td>
<td>Bw4.80I</td>
<td>0.8</td>
<td>0.4</td>
<td>91</td>
<td>782</td>
</tr>
<tr>
<td>3DS1</td>
<td>Bw4/Bw4</td>
<td>0.94</td>
<td>0.9</td>
<td>45</td>
<td>782</td>
</tr>
<tr>
<td>3DS1</td>
<td>Bw4.80I/Bw4.80I</td>
<td>0.61</td>
<td>0.4</td>
<td>12</td>
<td>782</td>
</tr>
</tbody>
</table>

Appendix Table B.3: Only the effect of B*57 and not of alleles with similar binding is enhanced by KIR2DL2.

<table>
<thead>
<tr>
<th>Cohort stratification</th>
<th>OR for B*57</th>
<th>p-value</th>
<th>N B*57+</th>
<th>N B*57-</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIR2DL2+</td>
<td>0.40</td>
<td>0.007</td>
<td>49</td>
<td>359</td>
</tr>
<tr>
<td>KIR2DL2-</td>
<td>0.83</td>
<td>0.631</td>
<td>35</td>
<td>339</td>
</tr>
<tr>
<td>KIR3DL1+</td>
<td>0.56</td>
<td>0.021</td>
<td>80</td>
<td>658</td>
</tr>
<tr>
<td>KIR3DL1-</td>
<td>2.5</td>
<td>0.6</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>KIR3DS1+</td>
<td>0.43</td>
<td>0.05</td>
<td>30</td>
<td>214</td>
</tr>
<tr>
<td>KIR3DS1-</td>
<td>0.67</td>
<td>0.2</td>
<td>54</td>
<td>484</td>
</tr>
</tbody>
</table>

Appendix Table B.4: KIRs which are known to bind B*57 do not enhance the B*57 protective effect as much as KIR2DL2 does.
there is still a trend for KIR2DL2 to enhance B*57 (OR=0.47 p=0.092); there are 30 KIR2DL2+KIR3DS1- individuals with B*57. There are only 4 individuals who are KIR2DL2+ but KIR3DL1- so excluding KIR3DL1 individuals is not possible.

4. In a model to predict HCV status in which KIR2DL2 with B*57 and KIR3DL1 with B*57 were both included as covariates (plus confounders) and then non significant HLA:KIR were removed by backwards stepwise exclusion then only KIR2DL2 with B*57 remained as a significant predictor. Similarly, in a model to predict HCV status in which KIR2DL2 with B*57 and KIR3DS1 with B*57 were both included as covariates (plus confounders) then, following backwards stepwise exclusion, only KIR2DL2 with B*57 remained as a significant predictor.

B.3 A*02 binds peptides strongly

In order to show that A*02 molecules bind peptides significantly more strongly than other HLA molecules we used two approaches. First, we extracted the list of positive MHC Binding Assays from the Immune Epitope Database (IEDB) [349] which contained in total 78202 peptides that bind HLA molecules. We then compared the experimental affinity measurements of all the A*02xx molecules with the corresponding values of other frequent HLA molecules (only HLA-A and B molecules were considered). The frequencies of the HLA molecules were calculated based on available data for UK and USA populations at the Allele Frequency Net Database [204]. Secondly, we used the epitope prediction software Metaserver [198] to identify potential epitopes for both the HCV and HTLV-1 proteomes for 36 and 21 HLA molecules respectively (we only obtained predicted epitopes for the HLA alleles present in the cohorts included in the study). Then we compared, as in the previous approach, the predicted affinity of the A*02xx:peptide complexes which were less than 500 nM IC50 with the predicted
values for the rest of the pMHC complexes included in the analysis. Both the above approaches suggest that A*02 binds peptides significantly more strongly compared to other HLA molecules. In the first approach (experimentally measured affinities), A*02 bound epitopes significantly more strongly than each of A*01, A*30, B*07, B*35, B*44 and B*45 (p<0.00001 in each case, Wilcoxon Rank sum, two-tailed see Appendix Figure B.1). In the second approach (theoretically predicted affinities), A*02 bound epitopes significantly more strongly than the other A alleles considered (HTLV-1: p=0.015, HCV: p<0.00001), B alleles (HTLV-1: p<0.00001, HCV p<0.00001) and A and B alleles combined (HTLV-1: p<0.00001, HCV p<0.00001), Appendix Figures 2(a) and 2(b).

Appendix Figure B.1: A*02 binds epitopes significantly more strongly than other HLA molecules which are frequent in UK and USA populations. The binding measure shown here is the affinity (nM IC50) values obtained from the IEDB database for 16958 peptides.
Appendix Figure B.2: A*02 binds peptides significantly more strongly than other HLA molecules based on predicted epitopes from the HTLV-1 and HCV proteomes. The binding measure shown here is the affinity (nM IC50) values obtained from Metaserver [198] for 2446 and 3611 peptides, respectively.
C Appendix C

C.1 Alignment to ambiguous reference sequence

In Appendix Figure C.1, we show a very small part of a multiple alignment where some bases of the reads have aligned to ambiguous bases of the reference sequence based on the IUPAC code. An 85% ambiguous consensus sequence is used. This means that we first perform a multiple alignment of the 78 HCV genotype 1a strains available on GenBank and then use both the most common nucleotide and the variants that exceed >15% frequency at each position in order to define the consensus of that position.

Appendix Figure C.1: Sequence alignment using an 85% ambiguous reference strain. The alignment mismatch parameter is set to 5% per read.

\[\text{http://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html}\]
Appendix Figure C.2: Coverage of HCV amplicons. For some patient-timepoint samples all the amplicons were amplified (a) while for some others only some of them were amplified (b). For the alignment shown here the MOSAIK software is used; an 85% ambiguous consensus sequence is provided (i.e. positions that have >15% different nucleotides are defined by the IUPAC code) based on the 78 HCV genotype 1a strains available on GenBank. The ambiguous reference sequence is constructed using the Consensus Maker v2.0.0 software\textsuperscript{1}. The mismatch parameter is set to 5% per read.
Appendix Figure C.3: Coverage per HCV amplicon. For the alignment the MOSAIK software is used; an 85% ambiguous consensus sequence is provided (i.e. positions that have >15% different nucleotides are defined by the IUPAC code) based on the 78 HCV genotype 1a strains available on GenBank. The ambiguous reference sequence is constructed using the Consensus Maker v2.0.0 software. The mismatch parameter is set to 5% per read.
D Appendix D

D.1 AICc differences of the fitted models

The Appendix Figure D.1 shows the AICc of the model minus the AICc of the best fitting model for each animal. As a rule of thumb a difference of < 2 suggests substantial evidence for both models, values between 3 and 7 indicate that the model with the worse fit has considerably less support, whereas a difference > 10 indicates that the model with the worse fit is very unlikely.

Appendix Figure D.1: The fit of each model for all the animals. A large difference represents a poor fit. The best fit model is shaded, comparable models are shown in bold. For explanation of the models see Table 5.1. The last three animals in the list were euthanised after early chronic infection, receiving either ART-treatment alone or the combination of CD8+ T cell depletion and ART. These animals were excluded from the analysis as, in each case, they had received only half the treatment (i.e. just ART or just ART/depletion) so there was insufficient data to constrain the fits. Including these 3 animals in the analysis did not change the result (p=0.028, two tailed paired Mann-Whitney as before). Note: This table was produced by M. Elemans.
E Appendix E

E.1 3D Cellular Automaton - Cell motility rules

The following rules which govern the motility of the cell populations on the grid are the same with the ones used in [267] and where kindly provided by F. Graw. The implementation in C++ was performed by N.-K. Seich al Basatena (thesis author).

We distinguish between two types of movement:

I Movement into free space (FS)

II Neighbour swapping between neighboured cells

The cells are marked as swapped/moved in order to avoid moving them twice during one timestep. In addition, a check is performed in order to establish whether a cell is able to move or not (e.g. it is bound to a conjugate).

Move into free space:

1. Find an FS in the lattice

2. Determine the neighbouring cells of the FS which are able to move into this node according to their moving direction

3. If there is more than one cell in (2), choose one of these cells at random, change the position of the FS and the chosen cell, and mark both cells as moved in order to avoid to move them twice during one step

4. If there is no cell found in (2), change the preferred moving direction of each neighbouring cell valid for the next timestep

5. Repeat step (1)-(4) until all FS were updated

Neighbour swapping:

The algorithm is formulated for a CTL\(^2\). Movement of the other cell types occurs

\(^2\)Here we use the terms CTL and CD8+ T cell, as described in Chapter 5, interchangeably.
similarly.

1. Find a CTL in the lattice

2. Determine the type of the neighbouring cell to which the moving direction of
   the CTL is pointing and proceed obeying the following rules if the cell belongs
   to:

   (a) splenocytes: swap the position with the CTL if the splenocyte was not
       already moved during this timestep, otherwise determine a new moving di-
       rection for the CTL valid for the next timestep

   (b) CTL: check if the moving direction of this second CTL points to the first
       CTL

       • if yes then swap both cells

       • if no then determine new moving directions for both cells valid for the
         next timestep

   (c) infected cells: analogous to (b)

   (d) uninfected cells: analogous to (b)

   (e) RN: determine new moving direction for the CTL valid for the next timestep

   (f) FS: swap cells

3. Repeat step (1)-(2) until all CTLs have been updated

The moving direction \( \gamma \in \{1, \ldots, 26\} \) points to one of the 26 neighbours of a cell. All
   cells are attributed with a position in 3D-space. The moving direction is updated as
   follows:

1. Translate the moving direction, \( \gamma \), to cartesian coordinates \((x, y, z)^T\) with \(x, y, z \in \{-1, 0, 1\}\). The position of the cell under consideration is \((0, 0, 0)^T\).
2. Choose one coordinate $\epsilon \in \{x, y, z\}$ at random and change its values as follows:

$$
\epsilon_{\text{new}} = \begin{cases} 
\epsilon = 1 & \Rightarrow \begin{cases} 
\epsilon - 1, & p = 0.8 \\
\epsilon - 2, & p = 0.2 
\end{cases} \\
\epsilon = 0 & \Rightarrow \begin{cases} 
\epsilon + 1, & p = 0.5 \\
\epsilon - 1, & p = 0.5 
\end{cases} \\
\epsilon = -1 & \Rightarrow \begin{cases} 
\epsilon + 1, & p = 0.8 \\
\epsilon + 2, & p = 0.2 
\end{cases}
\end{cases}
$$

3. Replace $\epsilon$ by $\epsilon_{\text{new}}$ and translate the cartesian coordinates to the moving direction $\gamma$.

Based on this algorithm the cells prefer small turning angles. In general, cell movement involves restructuring of the actin-filament network in the cytoskeleton so cells are expected to prefer small turning angles \cite{350}.

### E.2 Model quantities: look-up tables
<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice edge</td>
<td>50</td>
<td>n/a</td>
<td>125000 cells, 0.05 − 0.5% of the splenic white pulp [351]</td>
</tr>
<tr>
<td>Timestep</td>
<td>30 sec</td>
<td>n/a</td>
<td>Integrates micro and macro scale</td>
</tr>
<tr>
<td>Reticular network</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free space</td>
<td>1%</td>
<td></td>
<td>Spleen is a densely packed organ</td>
</tr>
<tr>
<td>Epitope-specific CD8+ T cells</td>
<td>0.5%</td>
<td>[168, 352]</td>
<td>Estimated based on ELISPOT assays</td>
</tr>
<tr>
<td>CD4+ T cells</td>
<td>15%</td>
<td></td>
<td>Mouse spleen</td>
</tr>
<tr>
<td>MΦs &amp; DCs</td>
<td>1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Motility Parameters**

| T cell speed | 10 − 15 \( \mu m/min \) | [268, 269, 271, 272] | In spleen, estimates are lower [283, 284] |
| T cell motility coefficient | 50 − 100 \( \mu m^2/min \) | [285, 353, 354] |

Appendix Table E.1: CA model initialisation values and motility parameters.
<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell properties</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_0$</td>
<td>6-8</td>
<td>[274][291]</td>
<td></td>
</tr>
<tr>
<td>Total death attributed to CD8+ T cells</td>
<td>25 – 30%</td>
<td>[3][218][297]</td>
<td></td>
</tr>
<tr>
<td>Lifespan of uninfected CD4+ T cells</td>
<td>13 – 76 d</td>
<td>[355]</td>
<td></td>
</tr>
<tr>
<td>Lifespan of infected CD4+ T cells</td>
<td>2 d</td>
<td>[295]</td>
<td>Includes 1 d viral eclipse phase</td>
</tr>
<tr>
<td>Eclipse phase of viral production</td>
<td>1 d</td>
<td>[293][294]</td>
<td></td>
</tr>
<tr>
<td>Duration of cytokine effect</td>
<td>Half-life: 6-9 hrs</td>
<td>[317]</td>
<td>Case study: RANTES</td>
</tr>
<tr>
<td>CD8+ T cell cytokine secretion duration</td>
<td>≈ 10 mins</td>
<td>[301]</td>
<td>Ag+ recognition with no lytic response</td>
</tr>
<tr>
<td>CD8+ T cell scanning time</td>
<td>$7 \pm 2 \text{ min}^{-1}$</td>
<td>[301]</td>
<td>Normal distribution</td>
</tr>
<tr>
<td>CD8+ T cell lytic killing duration</td>
<td>30 mins</td>
<td>[251][301][395]</td>
<td></td>
</tr>
</tbody>
</table>

Appendix Table E.2: CA model parameters.
F Appendix F

F.1 No significant effect of the grid size on the estimated escape rates

Appendix Figure F.1: The effect of the CA grid size on estimated escape rates under lytic CD8+ T cell suppression. When comparing the estimated escape rates (or killing rates - data not shown) no significant difference ($p > 0.05$) is observed between grid sizes for any of the CD8+ T cell suppression levels.
F.2 Killing rate increases with effector cells size

We vary the size of the epitope-specific CD8+ T cell population in the range of 0.5%-1.5% of the total splenocyte population (in line with biological ranges for PBMCs) in order to investigate its effect on killing rate. We find that the killing rate increases as the effector size increases (Appendix Figure F.2-left). We further explore whether this increase is proportional and find that for all the different probabilities of successful scan (pss) the data are consistent with a mass-action behaviour. In other words, the relative increase in effector cells entails a relative increase in effector cell killing rate which is not statistically significantly different (p > 0.05; Appendix Figure F.2-right). However, some of the data are noisy so we cannot draw firmer conclusions at this point.
Appendix Figure F.2: We vary the magnitude of the epitope-specific CD8+ T cell population and we observe that a larger number of effector cells leads to a higher CD8+ T cell killing rate (left). The relative increase in effector cells is not significantly different to the relative increase in killing rate, i.e. the slopes of the lines that correlate the relative increases of the two quantities (right) are not significantly different to the unit, $p > 0.05)$. Abbreviations: pss=Probability of Successful Scan by CD8+ T cells.
F.3  The percentage of simulations for each model that do not result in viral escape

We also explore the number of simulations that do not result in viral escape as a potential surrogate measure of CD8+ T cell pressure on the wild-type infected cells.

<table>
<thead>
<tr>
<th>pss</th>
<th>Lytic</th>
<th>NLI-P1</th>
<th>NLI-D1</th>
<th>NLI-D2</th>
<th>NLp-TO</th>
<th>NLp-P1</th>
<th>NLp-D1</th>
<th>NLp-D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>33%</td>
<td>57%</td>
<td>33%</td>
<td>76%</td>
<td>50%</td>
<td>43%</td>
<td>35%</td>
<td>57%</td>
</tr>
<tr>
<td>0.002</td>
<td>24%</td>
<td>71%</td>
<td>57%</td>
<td>62%</td>
<td>36%</td>
<td>43%</td>
<td>21%</td>
<td>36%</td>
</tr>
<tr>
<td>0.003</td>
<td>0%</td>
<td>52%</td>
<td>52%</td>
<td>62%</td>
<td>36%</td>
<td>36%</td>
<td>58%</td>
<td>36%</td>
</tr>
</tbody>
</table>

Appendix Table F.1: The percentage of simulations for each model that do not result in viral escape. Abbreviations: pss=Probability of Successful Scan by CD8+ T cells, NLI: Non-lytic model - blocking infection of uninfected CD4+ T cells, NLp: Non-lytic model - blocking viral production from infected CD4+ T cells, TO=Target Only, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
F.4 Quantifying the immune control of a non-lytic response that blocks viral production

Appendix Figure F.3: Number of infected CD4+ T cells ‘blocked’ from viral production under a non-lytic CD8+ T cell response manifested in a polarised or diffusive secretion pattern. Abbreviations: NLp: Non-lytic model - blocking viral production from infected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Appendix Figure F.4: New infections prevented under a non-lytic CD8+ T cell response that blocks infection. We show the number of infections prevented by the epitope-specific CD8+ T cell clones for 40-50 dpi, just before the variant infected cell population is introduced in the simulations. Abbreviations: NLp: Non-lytic model - blocking viral production from infected CD4+ T cells, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
Appendix Figure F.5: Set-point of productively infected cells. We show the percentage of productively infected wild-type cells for 40-50 dpi, just before the variant infected cell population is introduced in the simulations and after the steady-state has been attained. Here, we present the results for the lytic control and the non-lytic control that blocks viral production. Abbreviations: NLp: Non-lytic model - blocking viral production from infected CD4+ T cells, TO= Target Only, P1=Polarised secretion (r=1), D1=Diffusive secretion (r=1) and D2=Diffusive secretion (r=2).
F.5 Equivalence of non-lytic models in chronic infection

It can readily be shown that under a quasi-equilibrium between infected cells and free virus which holds during the chronic phase of infection, a non-lytic model where the CD8+ T cells reduce viral production (Equation 5.6) and a non-lytic model where CD8+ T cells reduce infection of new targets (Equation 5.8) produce the same dynamics for the population of productively infected cells ($T^*$).

The quasi-steady assumption for the non-lytic model where infection of new cells is hindered results in:

$$V = \frac{p}{c} T^*$$ \hspace{1cm} (1)

while the quasi-steady state assumption for the non-lytic model where viral production is blocked results in:

$$V = \left( \frac{1}{1 + \eta E} \right) \frac{p}{c} T^*$$ \hspace{1cm} (2)

Considering a constant population of uninfected target cells ($S$) and substituting Equations [1] and [2] in the ones governing the behaviour of productively infected cells ($T^*$), Equations [5.6] and [5.8] respectively, they change into:

$$\dot{T}^* = \left( \left( \frac{1}{1 + \eta E} \right) \frac{\beta p S}{c} - \delta_I \right) T^*$$ \hspace{1cm} (3)

where $\beta$ is the infection rate, $p$ is the production rate of free virions, $c$ is the clearance rate of free virions and $\delta_I$ is the death rate of productively infected cells (all $d^{-1}$). The fraction of CD8+ T cells, $E$, is given by the empirical function calculated by a linear interpolation between time points.


273


[76] Achim K Moesta, Paul J Norman, Makoto Yawata, et al. Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for


[132] S Niewiesk, S Daenke, C E Parker, et al. The transactivator gene of human T-cell leukemia virus type I is more variable within and between healthy carriers than patients


[185] Aidan MacNamara, Aileen Rowan, Silva Hilburn, et al. HLA Class I Binding of HBZ Determines Outcome in HTLV-1 Infection. 6(9), September 2010. PMID: 20886101 PMCID: 2944806.


[243] Lars T van der Veken, Maria Diez Campelo, Menno A W G van der Hoorn, et al. Functional analysis of killer Ig-like receptor-expressing cytomegalovirus-specific CD8+


300


