The Contribution of HLA Class I-Mediated $CD8^+$ T Cell Responses to HIV-1 Immune Control and Evolution

Philippa Clare Matthews

St John's College and

Department of Paediatrics,

University of Oxford

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
15 JUL 2010
THESIS



Trinity Term 2009

Thesis submitted to the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

To Andrew

without whom this work

would not have been possible

Abstract

CD8$^+$ T cells are a critical component of the adaptive immune response to HIV infection. The CD8$^+$ response is orchestrated by presentation of HIV peptides in association with HLA Class I molecules on the surface of infected cells. Disease outcome varies according to HLA phenotype, with some alleles being characteristically associated with sustained viraemic control, and others with more rapid progression to AIDS. In order to evade the CD8$^+$ T cell response, HIV selects escape mutations that reduce or abrogate HLA Class I presentation or disrupt recognition by T cell receptors. These mutations often occur in predictable combinations, termed HLA 'footprints', in response to selection pressure imposed by specific alleles.

In the setting of a recent failed T cell vaccine trial, these studies aim to improve characterisation of favourable CD8$^+$ T cell responses and to investigate the impact of HLA footprints on future viral evolution.

The analyses highlight the potential benefit of selection of mutations that impose a viral fitness cost. By partially crippling the virus, these polymorphisms can contribute to viraemic control. The evolutionary effects of an HLA footprint are shown to be significant both in mediating convergent evolution, and in driving patterns of viral evolution across global populations. Studies of HLA-B*35 in HIV demonstrate that Class I alleles can have differential impacts according to the subtype of infecting virus. In this instance, there is evidence of disease control mediated by CD8$^+$ T cell responses to HIV Gag. Finally, this work explores *in vitro* viral suppression assays as a tool for quantifying the CD8$^+$ T cell response in individual subjects, presenting optimisation of techniques for assessing bulk and individual HLA-Class I restricted responses to HIV.

Overall, these data contribute to the development of an HIV vaccine by enhancing our understanding of a successful CD8$^+$ T cell response, suggesting the need to consider immunogens that match the strain of infecting virus, and investigating *in vitro* methods for assessing vaccine outcomes.

Contents

Manuscripts and publications relevant to this thesis ix
Acknowledgements xii
Glossary of terms xvi

Chapter 1: Introduction

1.1 Historical context of HIV 1
1.2 HIV epidemiology 1
1.2.1 Classification of HIV 1
1.2.2 The scale of the epidemic 1
1.2.3 Routes and risks of HIV transmission 2
1.3 Structure and function of HIV 4
1.3.1 Structure of HIV virion and genome 4
1.3.2 Gag 5
1.3.3 Pol 8
1.3.4 Env 9
1.3.5 Accessory and regulatory proteins 10
1.4 HIV diversity and evolution 13
1.4.1 Relationship between HIV and SIV 13
1.4.2 The origin of clades 14
1.4.3 Factors driving HIV diversity 16
1.5 HIV life cycle 17
1.5.1 Transmission across mucus membranes 17
1.5.2 Fusion and entry to host cell 17
1.5.3 Cells targeted by HIV infection 18
1.5.4 Reverse transcription and integration of HIV genome 20
1.5.5 Production and transmission of progeny virions 20
1.5.6 Gut immunopathology 21
1.5.7 Viral latency and reservoirs of infection 22
1.6 Dynamics and clinical correlates of HIV infection 22
1.6.1 Relationship between $CD4^+$ T cell count, viral load and disease 22
1.6.2 Mechanisms of $CD4^+$ T cell decline 23
1.7 HLA Class I and $CD8^+$ T cell mediated immunity 24
1.7.1 Evidence for the role of $CD8^+$ T cells in control of viraemia 24
1.7.2 Evolution and genetics of HLA 25
1.7.3 HLA structure and peptide binding 26
1.7.4 Antigen presentation to T Cell Receptors (TCR) by HLA 29
1.7.5 $CD8^+$ T cell mediated immune escape 30
1.7.6 HLA and HIV transmission 33
1.7.7 Determinants of a successful $CD8^+$ T cell response 35
1.8 Other immune responses to HIV 40
1.8.1 B cells and antibody-mediated immunity 40
1.8.2 Natural Killer (NK) cells 41
1.8.3 $CD4^+$ T cells 42
1.8.4 Host restriction factors 43
1.9 HIV Vaccine strategies 44
1.9.1 Current options for HIV vaccines 44
1.9.2 Goals of a successful vaccine 45
1.9.3 $CD8^+$ T cell vaccines 47
1.9.4 Antibody vaccines 49
1.10 Aims of these studies 50

Chapter 2: Materials and Methods

2.1 Patient cohorts 52
2.1.1 Durban, South Africa 52
2.1.2 Gaborone, Botswana 53
2.1.3 Thames Valley, UK 53
2.1.4 Additional sources of sequence data 54
2.2 Processing of fresh whole blood 55
2.2.1 Harvesting PBMCs from whole blood 55
2.2.2 Freezing and thawing cells 55
2.2.3 DNA extraction 55
2.2.4 RNA Extraction 56
2.3 HLA Typing 56
2.4 Interferon-gamma (IFN-γ) ELISpot assays 57
2.4.1 General methods for ELISpot assay 57
2.4.2 Screening for HIV-specific responses by megamatrix 58
2.4.3 ELISpot confirmation of optimal epitopes 58
2.5 Amplification of HIV genes by Polymerase Chain Reaction (PCR) 59
2.5.1 Amplification from proviral DNA 59
2.5.2 Amplification from RNA 59
2.5.3 Visualisation of PCR product 59
2.5.4 Caveats of population sequencing 60
2.5.5 Quality control of PCR 64
2.6 Purification and sequencing of amplified HIV genes 64
2.6.1 Purification of PCR product 64
2.6.2 Sequencing of PCR product 64
2.7 Bioinformatics and phylogenetic analysis 65
2.7.1 Statistical tests 65
2.7.2 Identification of HLA Class 1 associations with HIV polymorphisms 66
2.7.3 Calculation of Shannon Entropy 66
2.7.4 Phylogenetic analysis 66
2.8 Enrichment of $CD4^+$ and $CD8^+$ T cells from PBMC 68
2.8.1 Positive / negative selection using Magnetic Cell Separation (MACS) 68
2.8.2 Activation of enriched $CD4^+$ and $CD8^+$ T cell populations 68
2.9 Fluorescence activated cell sorting (FACS) 69
2.10 Use of MHC Class I tetramers 70
2.10.1 Tetramerisation of MHC Class I monomer 70
2.10.2 Testing tetramer using compensation beads 70
2.10.3 Selection of $CD8^+$ T cell populations using tetramers 71
2.11 Generation of $CD8^+$ T cell lines 71
2.11.1 Derivation of T cell lines from PBMC 71
2.11.2 Testing T cell lines by intra-cellular staining (ICS) 72
2.12 Culture and estimation of titre of HIV-1 *in vitro* 72
2.12.1 Laboratory viruses 72
2.12.2 Generation of viral stocks 73
2.12.3 Estimation of viral titre 73
2.13 Viral suppression assays 75
2.13.1 *In vitro* infection of target cells 75
2.13.2 Seven day viral suppression assay 75
2.13.3 ELISA assay to measure p24 concentration 76
2.14 Supplementary methods tables 78

Chapter 3: Effect of reverting mutations in control of viraemia in HIV

3.1 Introduction 80
3.1.1 Relationship between $CD8^+$ T cell responses that target Gag epitopes and control of HIV viraemia 80
3.1.2 The effect of reverting mutations on control of viraemia 81
3.1.3 Hypothesis and aims 81

3.2 Methods 82
3.2.1 Study cohort and origin of sequence data 82
3.2.2 HLA typing 83
3.2.3 Methods for identification of sites of HLA-mediated selection 83
3.2.4 Optimisation of a lineage-corrected approach to identification of HLA-selected polymorphisms 85
3.2.5 Interpretation of statistical data from lineage-corrected algorithm 87
3.2.6 Evaluation of lineage-corrected methodology 90
3.2.7 Identification of HLA associations with 18-mer peptide recognition 93
3.2.8 Statistical analysis of data 94

3.3 Results 94
3.3.1 HLA Class I expression in 710 C-clade infected South African subjects 94
3.3.2 HLA-associated HIV amino acid polymorphisms selected by HLA-B and located in p24 Gag predominate 95
3.3.3 Known epitopes containing HLA-associated polymorphisms are predominantly restricted by HLA-B 101
3.3.4 An increased number of HLA-associated polymorphisms correlates with lower median viral load 102
3.3.5 Correlation between HLA-B polymorphisms in Gag and viral load persists with reduced number of Gag sequences 104
3.3.6 Number of polymorphisms that revert following transmission correlates with reducing viral load 105
3.3.7 An increased number of HLA-B associated polymorphisms in Gag correlates with higher median $CD4^+$ T cell count 107
3.3.8 Identification of 45 new HLA-associations with HIV peptide responses using IFN-γ ELISpot assays 107
3.3.9 Number of HLA-B restricted Gag epitopes targeted is correlated with reducing viral load 109
3.3.10 Sites of HLA selection are associated with increased variability 111

3.4 Discussion 112
3.4.1 Relationship between selection of reverting mutations and control of viraemia 112
3.4.2 Relationship of this work to other recent studies 115
3.4.3 Advantages of lineage-corrected statistical approach 116
3.4.4 Limitations of lineage-corrected statistical approach 116
3.4.5 Mechanisms for HLA-mediated control of viraemia 118
3.4.6 Relevance of these studies to HIV vaccine design 119

Chapter 4: The impact of HLA selection on HIV phylogeny

4.1 Introduction 121
4.1.1 Selection of HLA footprints and relationship to HIV clades 121
4.1.2 Immune control of HIV in association with HLA-B*57 and HLA-B*27 122
4.1.3 Hypothesis and aims 124

4.2 Methods 124
4.2.1 Impact of HLA selection on sites of amino acid variation between clades 124
4.2.2 Evaluating the impact of clade-specific amino acid polymorphisms on phylogeny 125
4.2.3 Relationship between Shannon entropy and sites of HLA selection 126
4.2.4 Investigating phylogenetic clustering among sequences bearing HLA footprints 127

4.3 Results 130
4.3.1 HLA footprint sites overlap with sites of inter-clade amino acid polymorphism 130
4.3.2 Shannon entropy scores are higher at sites of HLA selection 134
4.3.3 Synonymous and non-synonymous nucleotide substitutions contribute to clade specific phylogeny 135
4.3.4 HIV CRF subtype AE bears an HLA-B*57 footprint 137
4.3.5 Phylogenetic clustering can be mediated by an HLA-B*57 footprint in Gag 138
4.3.6 Phylogenetic clustering is brought about by a footprint of linked mutations 141
4.3.7 Phylogenetic clustering of Gag sequences increases with addition of sequential footprint mutations or increase in the proportion of footprint-bearing sequences 142
4.3.8 Phylogenetic clustering can be mediated by an HLA-B*27 Gag footprint in chronic infection 145

4.4 Discussion 147
4.4.1 Reasons why sites of difference between clades coincide with sites of HLA selection 147
4.4.2 Alternative approaches to identifying positive selection 148
4.4.3 Genetic determinants of clade specificity 149
4.4.4 HLA as a mediator of convergent evolution 149
4.4.5 Practical applications of these results 150

Chapter 5: Global adaptation of HIV to selection pressure imposed by HLA

5.1 Introduction 152
5.1.1 Impact of HLA selection on HIV evolution 152
5.1.2 Impact of HLA-selected mutations on disease control 153
5.1.3 Hypothesis and aims 154

5.2 Methods 155
5.2.1 Study cohorts 155
5.2.2 Selection of polymorphisms for analysis 156
5.2.3 Statistical analysis 159
5.2.4 Cellular assays 160
5.2.5 Analysis of HIV-infected subjects from Zimbabwe 160

5.3 Results 161
5.3.1 RT-I135X mutation is commonly selected by subjects with HLA-B*51 161
5.3.2 Most RT-I135X mutations do not impose a cost to viral fitness 162
5.3.3 RT-I135X mutations are associated with amino acid co-variation 163
5.3.4 RT-135X mutations allow escape from $CD8^+$ T cell recognition 165
5.3.5 The prevalence of RT-I135X mutations correlates with the phenotypic frequency of HLA-B*51 165
5.3.6 The prevalence of HLA-selected polymorphisms correlates with frequency of the selecting allele for persisting and reverting mutations 166
5.3.7 Models of HIV adaptation to HLA 168
5.3.8 The frequency of all escape mutations combined is strongly correlated with the population frequency of the selecting allele 170
5.3.9 Founder effect does not explain the relationship observed between HLA prevalence and HIV amino acid polymorphisms 171
5.3.10 RT-I135X mutations accumulate longitudinally 172
5.3.11 Class I alleles historically associated with disease control may lose their impact over time 173
5.3.12 HLA-B*5703 is enriched in HIV-infected subjects in Zimbabwe compared to other Southern African cohorts 176
5.3.13 HIV taxa from Southern Africa cluster together phylogenetically 178
5.3.14 Equivocal impact of HLA-B*5703 on viral loads in Zimbabwe 179
5.3.15 Sites of entropy difference between Zimbabwe and other Southern African cohorts coincide with HLA-B*57 footprint sites 180
5.3.16 HLA-B*57 footprints in Zimbabwe occur in proportion to the phenotypic frequency of HLA-B*57 181
5.3.17 $CD8^+$ T cell responses to Gag HLA-B*57 epitopes are preserved in Zimbabwe 182

5.4 Discussion 183
5.4.1 Evidence for HIV adaptation to HLA 183
5.4.2 HLA-B*51 selected mutation RT-I135X contributes to disease control 184
5.4.3 Challenges in identifying factors driving HIV evolution 184
5.4.4 Relationship between reverting mutations and HIV evolution 185
5.4.5 Impact of viral adaptation to HLA on disease control 186
5.4.6 Long term implications of HIV adaptation to HLA 187

Chapter 6: The effect of HLA-B*35 on disease control in HIV

6.1 Introduction 188
6.1.1 Rationale for the study of HLA-B*35 in HIV 188
6.1.2 HLA-B*35 subtypes and peptide binding 188
6.1.3 HLA-B*35 restricted $CD8^+$ T cell responses in HIV 189
6.1.4 The impact of HLA-B*35 on HIV disease outcome 190
6.1.5 Hypothesis and aims 191

6.2 Methods 192
6.2.1 Study cohorts 192
6.2.2 Generation of ELISpot data 192
6.2.3 Analysis of sequence data 192
6.2.4 Binding studies 193
6.2.5 Statistics and terminology 193

6.3 Results 193
6.3.1 HLA-B*3501 is the most frequent B*35 allele in HIV-infected cohorts in Europe, Asia and Africa 193
6.3.2 HLA-B*35 has a different impact on viraemic control in B-clade compared to C-clade HIV 194
6.3.3 The impact of HLA-B*35 on disease control may change over time 197
6.3.4 HLA-B*35 alleles present multiple HIV epitopes 198
6.3.5 Confirmation of HA9 (HPVHAGPIA, Gag 216-224) as an optimal epitope restricted by HLA-B*35 199
6.3.6 HLA-B*35-restricted epitopes HA9 and PY9 are in highly polymorphic regions of Gag 202
6.3.7 Gag-D260X is among the most frequent polymorphisms selected by HLA-B*35 206
6.3.8 Frequency of Gag-D260X polymorphism varies according to the population prevalence of HLA-B*35 207
6.3.9 Consequences of Gag-D260X polymorphism in subjects with HLA-B*35 208
6.3.10 Covariation of Gag-D260X with other Gag mutations 211
6.3.11 Subjects with HLA-B*35 make ELISpot responses weighted towards Gag/Pol in C-clade and towards Env/Nef in B-clade infection 213
6.3.12 HLA-B*35 restricted responses to OLPs containing Gag epitopes HA9 and PY9 are associated with reduction in viral load 217

6.4 Discussion 219
6.4.1 Evidence that HLA-B*35 may contribute to viraemic control and HIV sequence evolution 219
6.4.2 Identification of HLA-B*35 epitopes and escape mutations 219
6.4.3 Use of ELISpot data to determine HLA-B*35-restricted responses in B and C-clade infection 223
6.4.4 Summary comments 224

Chapter 7: *In vitro* quantification of the $CD8^+$ T cell response using Viral Suppression Assays (VSA)

7.1 Introduction 225
7.1.1 Limitations of ELISpot assays in assessing the $CD8^+$ T cell response 225
7.1.2 Use of Viral Suppression Assays (VSA) for quantifying the $CD8^+$ T cell response 226
7.1.3 Hypothesis and aims 226

7.2 Methods 227
7.2.1 Cell enrichment protocol and set-up of VSA 227
7.2.2 Refinements to methods for cell enrichment 228
7.2.3 Viral growth in culture 229
7.2.4 'Subtraction' approaches to quantify role of individual $CD8^+$ T cell responses 230
7.2.5 Analysis of p24 ELISA results 232

7.3 Results 233
7.3.1 *In vitro* HIV infection of enriched $CD4^+$ T cells results in a logarithmic growth phase up to day 4 233
7.3.2 Viral load and $CD4^+$ T cell count predict the outgrowth of autologous virus from enriched $CD4^+$ cells in culture 234
7.3.3 Inhibition of viral growth is not mediated by the addition of non-HIV-specific $CD8^+$ T cells 236
7.3.4 Inhibition of viral growth in superinfected cells is mediated by HIV-specific $CD8^+$ T cells in a dose-dependent manner 237
7.3.5 Significant viral suppression occurs *in vitro* in the presence of autologous $CD8^+$ T cells 238
7.3.6 Magnitude of viral suppression is predicted to varying extents by HLA genotype and clade of autologous infection 239
7.3.7 Subtraction assays: tetramer depletion approach 245
7.3.8 Subtraction assays: viral escape approach 248

7.4 Discussion 250
7.4.1 Summary of VSA as an *in vitro* approach to quantifying $CD8^+$ T cell responses 250
7.4.2 Conclusions from assays using bulk $CD8^+$ T cell populations 250
7.4.3 Conclusions from 'subtraction' assays 251
7.4.4 Caveats to the use of 'subtraction' approaches 252
7.4.5 Further optimisation of VSA and future work 253
7.4.6 Summary comments 254

Chapter 8: Summary and discussion

8.1 What do these studies add? 255
8.1.1 Relationship between reverting mutations and viraemic control 255
8.1.2 Influence of HLA on viral evolution within and between clades 256
8.1.3 Differences between viraemic control in association with HLA-B*35 in different populations 257
8.1.4 Use of *in vitro* assays to quantify viraemic suppression in association with bulk and individual $CD8^+$ T cell responses 258

8.2 Related studies currently in progress 259
8.2.1 The effect of HLA haplotypes on HIV viraemic control 259
8.2.2 Study of diversity in Southern African *nef* genes 259
8.2.3 *In vitro* comparison of HLA-B*35-restricted $CD8^+$ T cell responses 260
8.2.4 Epitope presentation and recognition in acute infection 260

8.3 Questions needing further investigation 260
8.3.1 What are the correlates of immune protection? 260
8.3.2 How significant is HLA in driving HIV evolution in the long term? 261
8.3.3 Why are certain HLA alleles favourable or unfavourable in effecting viraemic control? 261

8.4 The future of vaccine design 263
8.4.1 Is a preventative vaccine possible? 263
8.4.2 What modifications might be required for a successful vaccine? 265

8.5 Concluding remarks 266

References 268

Appendices 296

Appendix I: HIV-1 sequences for *gag*, *pol* and *nef* genes. C-clade consensus and Durban consensus compared to B-clade reference strain HXB2 296

Appendix II: Pairs of HLA alleles transmitted in linkage disequilibrium in 710 HIV-infected subjects in Durban, South Africa 300

Appendix III: Relationship between number of sequences analysed and mean Shannon Entropy Score per amino acid residue 301

Appendix IV: Intracellular staining of HLA-B*35 HA9 T cell lines derived from HIV-infected subjects R044 and R051 to identify IFN-γ production 302

Appendix V: Binding of HLA-B*35 optimal epitopes to the HLA-B*35 allele 303

Appendix VI: Supplementary methods for Chapter 7 304

Manuscripts and publications relevant to this thesis

Amino Acid Covariation in HIV-1 C-Clade p24 Reveals Structural Constraints. Crawford H, **Matthews P**, Leslie A, Ndung'u T, Carlson J, Heckerman D, Goulder P. Manuscript in preparation for J Virol.

I was involved in performing this analysis of co-variation in order to describe the compensatory changes that appear in conjunction with HLA escape polymorphisms in Gag.

HLA Mediated Control of HIV and HIV adaptation to HLA. Payne R, **Matthews PC**, Prado JG, Goulder P. Adv Parasitol. 2009;68:1-20. Review.

I contributed to the overall production of this review article, including text and figures, particularly with respect to the sections on HIV phylogeny and evolution.

HLA Footprints on HIV-1 are associated with Inter-Clade Polymorphisms and Intra-Clade Phylogenetic Clustering. **Matthews PC**, Leslie AJ, Katzourakis A, Crawford H, Payne R, Prendergast A, Power K, Kelleher AD, Klenerman P, Carlson J, Heckerman D, Ndung'u T, Walker BD, Allen TM, Pybus OG, Goulder PJR. J Virol. 2009;83:4605-15.

As first author on this project, I was responsible for the majority of data analysis, development of methodology and production of the final manuscript.

Adaptation of HIV-1 to HLA I. Kawashima Y*, Pfafferott K*, Frater J*, **Matthews P***, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, Kuse N, Oka S, Duda A, Prendergast A, Crawford H, Leslie A, Brumme Z, Brumme C, Allen T, Brander C, Kaslow R, Tang J, Hunter E, Allen S, Mulenga J, Branch S, Roach T, John M, Mallal S, Ogwu A, Shapiro R, Prado J, Fidler S, Weber J, Pybus O, Klenerman P, Ndung'u T, Phillips R, Heckerman D, Harrigan P, Walker B, Takiguchi M, Goulder P. Nature. 2009;458:641-5 **(* joint 1st author).**

I contributed to the large pool of sequence data used in this project, particularly adding to sequences for South Africa and Botswana cohorts. I performed a significant role in statistical analysis, including longitudinal

analysis of sequence data from Japan, and contributed to the text and figures presented in the final manuscript.

Phylogenetic Dependency Networks: Inferring Patterns of CTL Escape and Codon Covariation in HIV-1 Gag. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, **Matthews P**, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D. PLoS Comput Biol. 2008;4:e1000225.

This work was undertaken in collaboration with Microsoft Research. I contributed to the C-clade sequence data, and was involved in development and refinement of the lineage-corrected methods described.

Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. **Matthews PC***, Prendergast A*, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, Brander C, Bishop K, Mlotshwe N, Mullins JI, Coovadia H, Ndung'u T, Walker BD, Heckerman D, Goulder PJ. J Virol. 2008;82:8548-59 **(* joint 1st author).**

I contributed the significant majority of Nef sequences to this project, and – together with Andy Prendergast – undertook all of the analysis of data generated by Microsoft Research. The final manuscript and figures were all produced by Andy and myself.

HLA class-I driven evolution of HIV-1 Subtype C Proteome: Immune Escape and Viral Load. Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, Prendergast A, **Matthews P**, Payne R, Rolland M, Raugi DN, Maust BS, Learn GH, Nickle DC, Coovadia H, Ndung'u T, Frahm N, Brander C, Walker BD, Goulder PJ, Bhattacharya T, Heckerman DE, Korber BT, Mullins JI. J Virol. 2008;82:6434-46.

This project was undertaken in parallel with the manuscript listed above (J Virol. 2008;82:8548-59); I contributed sequence data and was involved in development of methodology.

Transmission of HIV-1 Gag Immune Escape Mutations is Associated with Reduced Viral Load in Linked Recipients. Goepfert PA, Lumm W, Farmer P, **Matthews P**, Prendergast A, Carlson J, Derdeyn CA, Tang J, Kaslow RA, Bansal A, Yusim K,

Heckerman D, Mulenga J, Allen S, Goulder P, Hunter E. J Exp Med. 2008 12;205:1009-17.

This paper used the data generated from our Durban analysis (J Virol. 2008;82:8548-59) to identify sites of escape mutation in Gag and Nef. In addition, I performed the phylogenetic analysis and contributed the tree figure for the final manuscript.

Fitness cost of escape mutation in p24 Gag in association with control of HIV-1. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, Thobakgale C, Honeyborne I, Crawford H, **Matthews P**, Pillay T, Mullins J, Brander C, Walker B, Stuart DI, Kiepiela P, Goulder PJR. J Virol. 2006; 80: 3617-23.

*I contributed to the analysis included in this paper whilst writing my grant application for research funding. I performed the statistical analysis of escape mutations in association with HLA-B*57 in B and C-clade study cohorts.*

Acknowledgements

First and foremost, thank you to my supervisor, Prof Philip Goulder – for taking me on as a PhD student, for advice and ideas, for creating amazing opportunities for international collaboration, and for persevering with publications through 32 drafts and beyond.

I am indebted to my colleagues in the Goulder Group. Special thanks go to post-docs Al Leslie, for his constant advice, support and provision of all-round entertainment, and Julia Prado for her mentorship, particularly in teaching me methods for cellular assays and viral culture. I am hugely grateful to my fellow PhD students Andy Prendergast, Rebecca Payne, Hayley Crawford, Henrik Kloverpris and Isobel Honeyborne, all of whom have contributed endless advice, expertise and enthusiasm and have made the past three years a delightful and unforgettable experience. Thanks also to Hannah Goodliffe, Claudia Juarez and Catherine Koofhethile for their contribution to the smooth running of the lab.

Collaborations with David Heckerman and Jonathan Carlson at Microsoft Research have contributed hugely to this work. My thanks to them for their patience, efficiency and good humour in countless re-runs of statistical algorithms, and for being constantly on-call via trans-Atlantic email. Thanks also to Bruce Walker and all at Partners AIDS Research, Boston for their help and collaboration. In particular, I am grateful to Todd Allen for contributing sequence data, and to Boris Julg for his advice in developing viral suppression assays.

International links have been fundamental to the success of these projects. I would particularly like to acknowledge Thumbi Ndung'u and colleagues at the HIV Pathogenesis Programme in Durban, Masafumi Takiguchi and team in Kumamoto, and Roger Shapiro in Gaborone, all of whom have generously contributed time and resources to recruiting patients and generating data.

Local patient recruitment would not have been possible without the dedicated efforts of clinical colleagues. In particular, Lynn Riddell, Fabian Chen, Graz Luzzi, Anne

Edwards and Thambiah Balachandran and their out-patient teams have worked tirelessly on our behalf.

In total, data from nearly 3500 HIV-infected subjects are represented in this work. Without the selfless generosity of these individuals, these studies would never have left the drawing board. Their time, effort and commitment are humbling, and hugely appreciated.

I am grateful to the UK Medical Research Council for providing me with a Clinical Research Training Fellowship grant to fund three years of full-time research.

And finally, enormous thanks to my amazing husband Andrew, for his endless patience and encouragement, and for keeping me smiling, and to my wonderful parents for their continued love and support in everything I do.

Chapter-specific contributions

The text, tables and figures in this thesis are my own work unless otherwise stated. Contributions from organizations and individual collaborators are listed below, and are also stated within methods and results sections of the relevant chapters.

Chapter 1: Introduction

Part of the text for the discussion concerning the origin of HIV clades (sections 1.4.1 and 1.4.2) is taken from my contribution to a review article published in 2009 (Payne *et al.*, *Adv Parasitol.* 68:1-20). Sources of figures are acknowledged individually in the figure legends.

Chapter 3: Effect of reverting mutations on control of viraemia in HIV-1

Amplification and sequencing of proviral *gag*, *pol* and *nef* genes was performed by the Goulder Group in Oxford (*gag* by Hayley Crawford; *pol* by Andy Prendergast and Rebecca Payne; *nef* by me, with an additional 100 sequences contributed by Alasdair Leslie). Christine Rousseau in Seattle, USA, performed RNA extraction and sequencing to contribute additional sequence data. The algorithm for lineage-corrected sequence analysis was written and run by David Heckerman and Jonathan

Carlson, Microsoft Research, USA. Data generated from this algorithm were analysed by myself and Andy Prendergast. ELISpot assays for the Durban cohort were performed at the HIV Pathogenesis Programme, Durban, South Africa, in the lab supervised by Fanny Kiepiela and subsequently by Thumbi Ndung'u. Philip Goulder performed the analysis of ELISpot data for new optimal epitopes. The majority of this chapter was published in 2008 (Matthews *et al.*, J. Virol. 2008;82:6434-46); additional analysis of methods was performed by me for inclusion in this thesis.

Chapter 4: The impact of HLA selection on HIV phylogeny

Oliver Pybus and Aris Katzourakis (Department of Zoology, University of Oxford) provided invaluable assistance in the development of methodology to quantify phylogenetic clustering and advised on the implementation of these models. Al Leslie contributed substantially to the development of methods, and to the original analysis of simulated datasets. C-clade sequence data for the Durban cohort were generated as described in acknowledgements for Chapter 3. B-clade sequence data were provided by Anthony Kelleher (University of New South Wales, Australia) and Todd Allen (Partners AIDS Research Center, Boston, USA). Todd Allen and Ursula Hempel (Partners AIDS Research Center, Boston, USA) contributed amino acid sequences for the Thai cohort.

Chapter 5: Global Adaptation of HIV-1 to selection pressure imposed by HLA

The published data in this chapter resulted from a multi-centre analysis, with data from nine cohorts provided through collaborations with the following individuals:

Vancouver, Canada: Zabrina Brumme, Chanson Brumme

Bridgetown, Barbados: Songee Branch

Oxford, UK: Katya Pfafferott, Anne Edwards

London, UK: John Frater, Anna Duda, Sarah Fidler, Rodney Phillips

Gaborone, Botswana: Roger Shapiro

Lusaka, Zambia: Eric Hunter, Susan Allen, Joseph Mulenga

Durban, South Africa: Thumbi Ndung'u

Perth, Australia: Simon Mallal

Kumamoto, Japan: Yuka Kawashima, Masafumi Takiguchi

Additional unpublished B-clade sequence data for analysis of covariance were kindly provided by Jonathan Carlson and Zabrina Brumme from the HOMER and Western

Australian cohorts. PCR and sequencing of the Botswana cohort was undertaken by Rebecca Payne (*gag* genes), Tom Strong (*pol* genes) and myself (*pol* and *nef* genes). Oliver Pybus advised on statistical methods and contributed to computational analysis. Yuka Kawashima performed *in vitro* competition and fitness assays.

Chapter 6: The effect of HLA-B*35 on disease control in HIV-1

B-clade sequencing of UK subjects was performed by Katja Pfafferott. International cohorts (South Africa, Botswana and Japan) were derived from the sources acknowledged in Chapters 3, 4 and 5. The Søren Buus lab in Copenhagen performed HLA binding studies. Jonathan Carlson (Microsoft Research) provided the lineage-corrected data from which HLA associations and co-variation analysis were undertaken.

Chapter 7: Viral Suppression Assays

Julia Prado and Rebecca Payne (Goulder Group, Oxford), and Boris Julg (Walker Group, Boston) provided invaluable advice on optimizing the viral suppression methodology. MHC Class I monomers were made by Stephen Hickling (HLA-B*57-KF11) and Matthias Hoffmann (HLA-B*35-PY9) in Rodney Phillips's group in Oxford. The KF11 mutant virus was provided by Julia Prado, and sequenced by Catherine Koofhethile. Henrik Kloverpris provided help with the use of the BD LSR II FACS machine and made helpful suggestions regarding presentation of data for this chapter.

Total word count: approx 50,000 excluding tables, figures and references

Glossary of terms

Amino acid abbreviations

Characteristics	Single letter code	Three letter code	Amino acid
Non-polar (Hydrophobic)	A	Ala	Alanine
	F	Phe	Phenylalanine
	G	Gly	Glycine
	I	Ile	Isoleucine
	L	Leu	Leucine
	M	Met	Methionine
	P	Pro	Proline
	V	Val	Valine
	W	Trp	Tryptophan
Polar (Hydrophilic)	C	Cys	Cysteine
	N	Asn	Asparagine
	Q	Gln	Glutamine
	S	Ser	Serine
	T	Thr	Threnonine
	Y	Tyr	Tyrosine
Electrically charged (Negative and hydrophilic)	D	Asp	Aspartic acid
	E	Glu	Glutamic Acid
Electrically charged (Positive and hydrophilic)	H	His	Histidine
	K	Lys	Lysine
	R	Arg	Arginine

Units of measurement

°C	Degrees centigrade
g	g-force (relative centrifugal force)
h	hours
M	Molar
ml	millilitres
µg	micrograms
µl	microlitres
pg	picograms
pmol	picomoles
V	volts

Acronyms and abbreviations

Ad5	Adenovirus serotype 5
AIDS	Acquired Immuno-Deficiency Syndrome
APC	Allophycocyanin
APC	Antigen Presenting Cell
CD	Cluster of Differentiation
CRF	Circulating Recombinant Form
CTL	Cytotoxic T Lymphocyte ($CD8^+$ T lymphocyte)
CO_2	Carbon dioxide
Cyp-A	Cyclophilin-A
DC	Dendritic Cell
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside triphosphates
EDTA	Ethylene-Diamine-Tetraacetic Acid
ELISA	Enzyme Linked ImmunoSorbent Assay
ELISpot	Enzyme Linked Immunosorbent Spot assay
ER	Endoplasmic Reticulum
ERAAP	Endoplasmic Reticulum Aminopeptidase associated with Antigen Processing
E:T ratio	Effector:Target ratio
FACS	Fluorescence Activated Cell Sorting
FCS	Foetal Calf Serum
FDR	False Detection Rate
FITC	Fluorescein Isothiocyanate
GALT	Gut Associated Lymphoid Tissue
GFP	Green Fluorescent Protein
HAART	Highly Active Anti-Retroviral Therapy
HIV	Human Immunodeficiency Virus
HLA	Human Leucocyte Antigen
HTLV	Human T-Lymphotropic Virus
ICS	Intra-Cellular Staining
IFN-γ	Interferon gamma
IL-2	Interleukin-2
IP	Inner Primer
IQR	Interquartile range
KIR	Killer Immunoglobulin-Like Receptor

LD	Linkage Disequilibrium
LTNP	Long-Term Non-Progressor
MACS	Magnetic Cell Separation
MHC	Major Histocompatibility Complex
ML	Maximum likelihood (phylogenetic tree)
MOI	Multiplicity of Infection
MSM	Men who have Sex with Men
MT	Mutant
MT-4	Human T cells isolated from a patient with adult T-cell leukaemia
MTCT	Mother To Child Transmission
NAb	Neutralising Antibody
NJ	Neighbour joining (phylogenetic tree)
NK cell	Natural Killer cell
NL4-3	Standard B-clade HIV reference strain
OLP	Overlapping Peptide
OP	Outer Primer
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PE	Phycoerythrin
PerCP	Peridinin Chlorophyll Protein
PHA	Phytohemagglutinin
PIC	Pre-Integration Complex
PMA	Phorbol myristate acetate
RNA	Ribonucleic Acid
SEM	Standard Error of the Mean
SFU	Spot Forming Units
SIV	Simian Immunodeficiency Virus
TAP	Transporter Associated with Antigen Processing
TCID	Tissue Culture Infective Dose
TCR	T Cell Receptor
VL	Viral Load
VSA	Viral Suppression Assay
WT	Wild Type

CHAPTER 1: Introduction

1.1 Historical context of HIV

In 1981, clusters of cases of opportunistic infections and rare malignancies were identified in Los Angeles and New York, initially affecting young gay men and later arising in injecting drug users (Brennan and Durack 1981). The clinical syndrome became known as Acquired Immunodeficiency Syndrome, AIDS. The aetiological agent was subsequently identified as a Retrovirus of the lentivirus group in 1983 (Barre-Sinoussi *et al.* 1983), and – following initial classification as HTLV III - is now termed Human Immunodeficiency Virus-1 (HIV-1). Since the early 1980's, HIV-1 has spread to become a global pandemic, affecting an estimated total of 60 million people worldwide and accounting for nearly 30 million deaths (UNAIDS 2008).

1.2 HIV epidemiology

1.2.1 Classification of HIV

HIV-1 belongs to the family *Retroviridae*, so named because the RNA genomes are reverse transcribed before integration with the host DNA; within this family it is further classified as a lentivirus on account of its long period of clinical latency.

1.2.2 The scale of the epidemic

Currently, 33.2 million adults and children are estimated to be infected with HIV-1. The greatest burden of disease is in Sub-Saharan Africa, where 22.5 million people are infected, 60% of whom are women. In 2007, this region accounted for 90% of

global paediatric infections and 75% of AIDS deaths (UNAIDS 2008). South Africa has an estimated 5.5 million people living with HIV, making it the country with the largest number of infections in the world (UNAIDS 2008; Fig. 1.1).

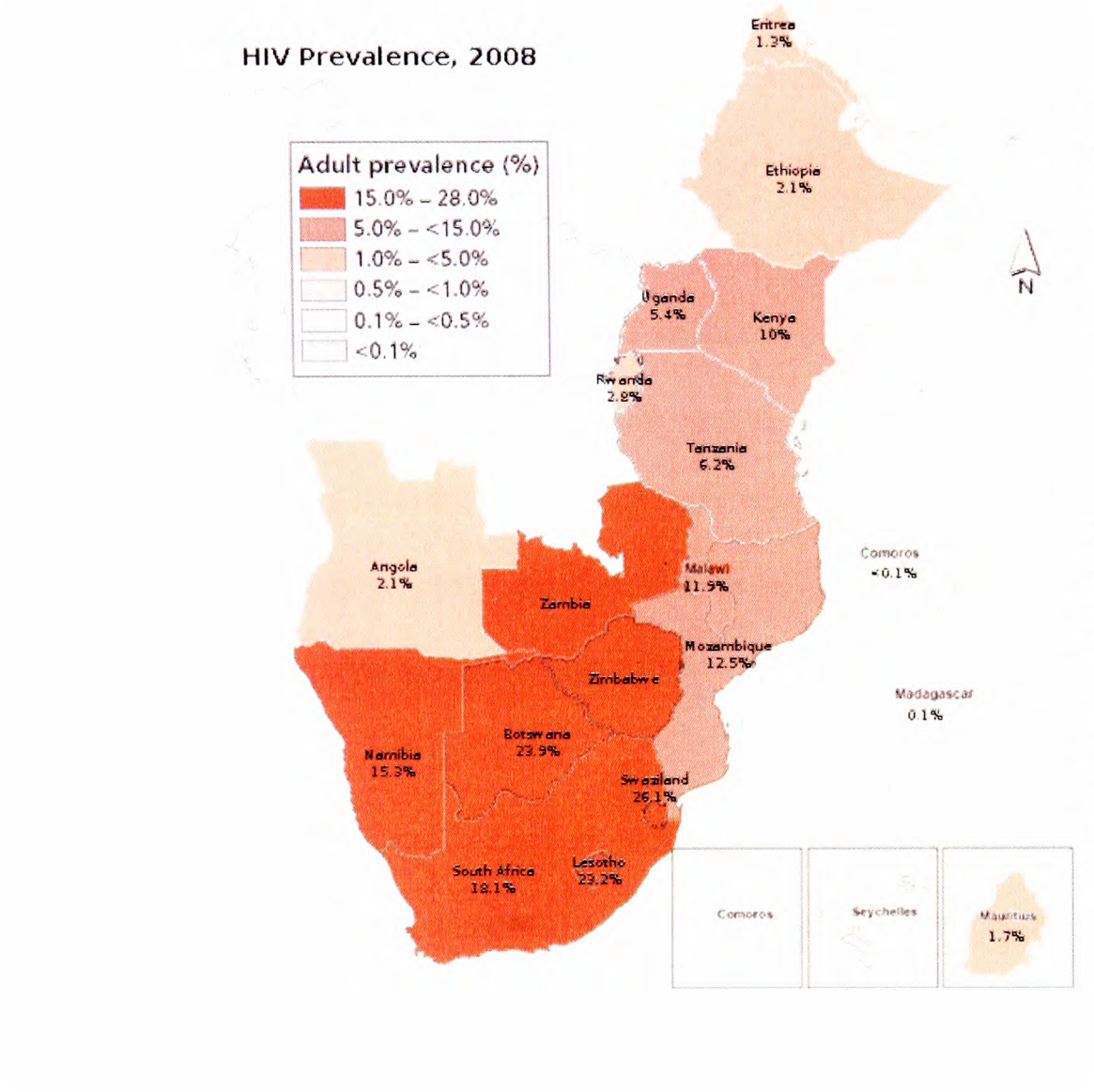


Figure 1.1: Map showing estimated prevalence of adult HIV infection in 2008 in Eastern and Southern Africa.

Image reproduced with permission from UNAIDS Regional Profile for Eastern and Southern Africa; image available at www.unaidsrstesa.org/profile-regional-profile

1.2.3 Routes and risks of HIV transmission

Although the first recognized cases of AIDS were seen in MSM (Men who have Sex with Men) (Brennan and Durack 1981), the vast burden of the global HIV epidemic is now attributable to heterosexual transmission in Africa (UNAIDS 2008). In other parts of the world, MSM remain disproportionately at risk, along with sex workers and intravenous drug users (www.cdc.gov).

In a Ugandan study, heterosexual sex carried a risk estimated at 0.00007 to 0.0028 per coital act, depending on the stage of disease and in the absence of HAART (Wawer *et al.* 2005). In sero-discordant couples in Zambia, the risk of transmission has been quantified at 8% per couple per year, even in the presence of condom provision and education (Goepfert *et al.* 2008). In parallel with the heterosexual epidemic of HIV-1 among young African adults, 90% of new paediatric infections occur in Africa, with an overall risk of 25-40% of vertical transmission from an infected mother (Prendergast *et al.* 2007). Two-thirds of Mother-to-Child-Transmission (MTCT) of HIV occurs in the peripartum period, either late in the third trimester (2/3) or intra-partum (1/3). The remaining cases are attributable to breast-feeding, with the highest risk mostly within the first few weeks of life when viral loads in breast milk are highest (Prendergast *et al.* 2007).

Small clusters of infection have also arisen from point sources disseminated by transfused blood products. The rate of infection in these cases is higher than sexual or vertical transmission, estimated at up to 95% per exposure event (Hladik and McElrath 2008), although still likely to depend on the viral sequence and viral load of the donor. One of the best characterised of such outbreaks is a subtype B Nef-deleted strain that was transmitted via the Sydney Blood Bank (Learmont *et al.* 1999; Gorry *et al.* 2007). Despite the attenuated viral strain in the donor, 8 of 10 recipients sero-converted (Gorry *et al.* 2007).

1.3 Structure and function of HIV

1.3.1 Structure of HIV virion and genome

The virion contains two copies of a 10kb single-stranded RNA genome of nine genes encoding three structural proteins, Gag, Pol and Env, and six regulatory and accessory proteins, Vpr, Vpu, Vif, Rev, Tat and Nef (Fig. 1.2).

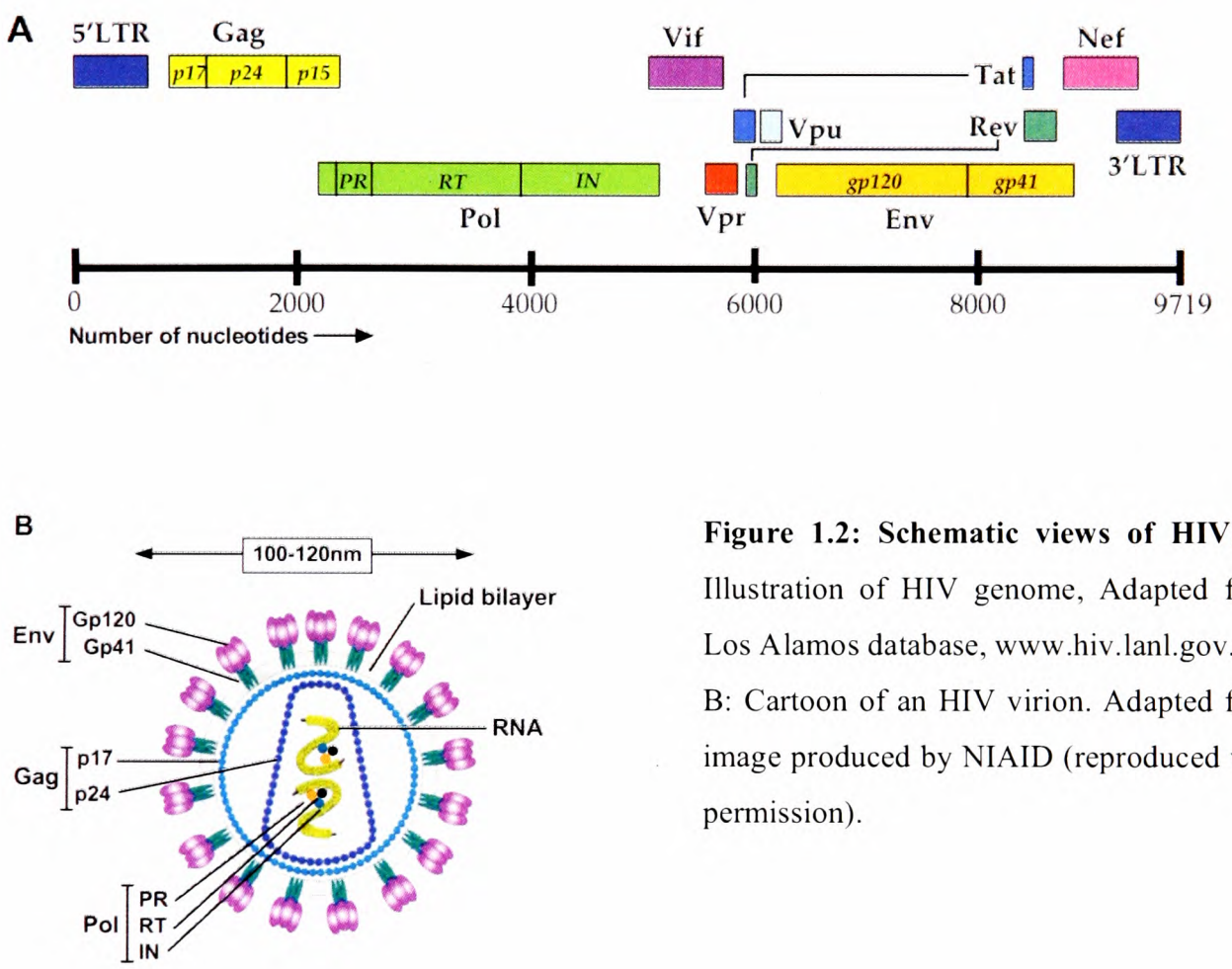


Figure 1.2: Schematic views of HIV. A: Illustration of HIV genome, Adapted from Los Alamos database, www.hiv.lanl.gov. B: Cartoon of an HIV virion. Adapted from image produced by NIAID (reproduced with permission).

Gag	Group specific antigen	**Accessory and Regulatory Proteins:**	
Pol	Viral enzymes	**Tat**	Transactivator of HIV gene expression
PR	Protease	**Rev**	Regulator of Expression of Viral proteins
RT	Reverse transcriptase	**Vif**	Viral Infectivity Factor
IN	Integrase	**Vpr**	Viral Protein R
Env	Envelope proteins	**Vpu**	Viral Protein U
Gp120	Glycoprotein 120	**Nef**	Negative Factor Protein
Gp41	Glycoprotein 41		
LTR	Long Terminal Repeat		

1.3.2 Gag

Gag ('Group Specific Antigen') is a protein component of all retroviruses, being critical for viral assembly, maturation and tertiary structure, and pre-requisite for infection (Ako-Adjei *et al.* 2005; Resh 2005; Grigorov *et al.* 2007). At the time of budding and maturation, Gag polyprotein is cleaved by viral protease into three major protein components, p17, p24 and p15, plus smaller protein fragments p1, p2 and p6 (Ako-Adjei *et al.* 2005). Regions of Gag are among the most conserved of the whole genome (Fig. 1.3), and the protein is present at high copy numbers (>1500 molecules per virion) even at the time of initial infection (Briggs *et al.* 2004).

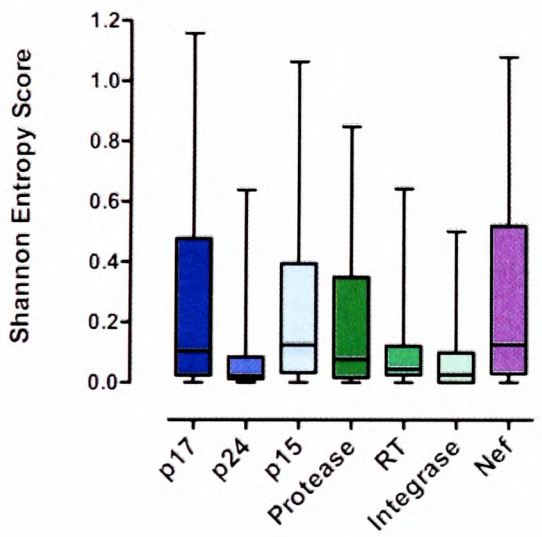


Figure 1.3: Sequence variability of HIV proteins between individuals.
Shannon entropy scores are shown for sequences from Durban, South Africa, derived from a total of 710 study subjects, reflecting the variability of constituent proteins of Gag (blue), Pol (green) and Nef (purple). Boxes show mean ± SEM, whiskers 5-95% confidence intervals. Data for non-Gag/Pol/Nef proteins is available in a 2008 review article (Goulder and Watkins 2008).

Gag polyprotein is arranged radially within immature virions, with the N-termini bound to the outer membrane and the C-termini pointing inwards to the centre of the particle (Wilk *et al.* 2001). Based on this structure, the presence of Gag is both necessary and sufficient for assembly of virus-like particles (Ako-Adjei *et al.* 2005; Resh 2005). Critical structural functions of Gag proteins are mediated via three main interactions, membrane-p17, p24-p24, and p15-RNA (Grigorov *et al.* 2007).

Gag p17, Matrix protein (MA)

Gag matrix protein is a 17kD structural protein composed of five α-helices. The varied functions of p17 are influenced by structural and biochemical changes in the protein that occur at different stages of the viral life-cycle (Hearps and Jans 2007). As new virions mature, p17 assembles into a layer beneath the envelope, forming the outer shell of the viral core (Grigorov *et al.* 2007) and guiding the assembly of envelope proteins. p17 undergoes co-translational myristoylation* at its N-terminal end (Resh 2005). Myristate is subsequently exposed during the process of Gag trimerisation, and is thus available to form a hydrophobic interaction with the membrane of the host cell (Ako-Adjei *et al.* 2005; Resh 2005; Li *et al.* 2007). In this way, viral assembly is directed to lipid-rich regions of the membrane from which progeny virions can acquire cholesterol and lipid molecules that are crucial for subsequent entry to new host cells (Hearps and Jans 2007). In order to infect terminally differentiated and non-dividing cell types, HIV must translocate its genome through the nuclear membrane of the host cell (Hearps and Jans 2007). Gag p17 contains a putative nuclear localization signal and can be detected within the cytoskeletal actin network of host cells within 20 minutes of infection (Hearps and Jans 2007).

Gag p24, Capsid protein (CA)

p24 is the most highly conserved Gag protein (Goulder and Watkins 2008) (Fig. 1.3), and its tertiary structure is conserved between all retroviruses (Ako-Adjei *et al.* 2005); in particular, a conserved stretch of 20 amino acids termed the Major Homology Region (MHR) forms the basis of a network of hydrogen bonds that produce

* Covalent binding to the 14-carbon fatty acid myristate, catalysed by the host enzyme N-myristoyl transferase (NMT), and dependent on the intitiating Met-Gly sequence.

dimerisation of p24 molecules (Gamble *et al.* 1997). Following proteolytic cleavage of p24 from other Gag components, p24 molecules dictate the formation of a hexameric lattice that is the basis of the conical capsid (Ako-Adjei *et al.* 2005; Sundquist and Hill 2007; Ganser-Pornillos *et al.* 2008). A proline-rich loop in p24 (residues 217-225) is required for binding of the host protein, Cyclophilin-A (Cyp-A), a ubiquitous eukaryotic cis/trans peptidyl-prolyl isomerase. This interaction allows the virus to evade host restriction factors that would otherwise provide resistance to infection (Towers *et al.* 2003), and also has roles in promoting the stable structure and function of HIV capsid (Gamble *et al.* 1996).

Gag p15, Nucleocapsid protein (NC)

Found at the C-terminal end of Gag polyprotein, p15 is essential for virus assembly, budding and infectivity (Grigorov *et al.* 2007). p15 is located in the inner core of the mature virion, bound to viral nucleic acid (Grigorov *et al.* 2007), where it plays a chaperoning role and stabilises the union between the two viral RNA molecules (Russell *et al.* 2003; Grigorov *et al.* 2007). Gag p15 also promotes multimerisation of p17 protein (Hearps and Jans 2007), thus contributing to the association of the immature virion with the host plasma membrane. The p6 domain, at the C-terminal end of p15 protein, is required for budding of progeny virions. Mutations in p15 therefore result in reduced production of new virus, defects in Gag multimerisation, and non-infectious viral particles (Grigorov *et al.* 2007).

1.3.3 Pol

The *pol* gene of HIV encodes three retrovirus-specific enzymes crucial to the function of the virus, Protease (PR), Reverse Transcriptase (RT) and Integrase (IN) (Jaskolski *et al.* 2009). The reading frame of the *pol* gene overlaps *gag*; expression requires a frame-shifting event in the *gag-pol* region of the genome.

Protease (PR)

HIV protease is an aspartic acid protease; in its active form it is a homodimer with a single active catalytic site (Katz and Skalka 1990). It allows maturation of the virion into a fully infectious particle by cleaving the Gag-Pol polyprotein, allowing assembly of the capsid proteins.

Reverse Transcriptase (RT)

RT has two catalytic functions that operate to convert single stranded viral RNA into double stranded proviral DNA in the cytoplasm of an infected cell (Katz and Skalka 1990; Sarafianos *et al.* 2009). First, it functions as an RNA dependent DNA polymerase, catalyzing the synthesis of a strand of proviral DNA from a single stranded viral RNA template to produce an RNA/DNA duplex. Secondly, RNase-H activity degrades the RNA from this duplex, to leave a single stranded DNA molecule that serves as a template for the synthesis of a second DNA strand (Sarafianos *et al.* 2009). Prokaryotic RT lacks a proof-reading function, such that 1 in every 1500-4000 bases is inserted incorrectly (Preston 1997).

Integrase (IN)

Proviral DNA is transported to the nucleus of the host cell, where retroviral integrase catalyses its insertion into the host genome via a two-step process (Jaskolski *et al.* 2009). In the first ('processing') step, two nucleotides are removed from the 3' end of viral DNA, followed by a second ('joining') step in which host DNA is cut with a five nucleotide stagger, such that the host DNA is free to form a covalent bond with viral nucleic acids (Jaskolski *et al.* 2009). The catalytic activity of the enzyme is all contained within the central domain of the protein, although intact N and C terminal domains are also required for full enzymatic function (Jaskolski *et al.* 2009).

1.3.4 Env

Env is the most diverse of HIV proteins, with up to 35% variation in nucleotide sequences between clades (Spira *et al.* 2003; Lynch *et al.* 2009). Env precursor protein gp160 is post-translationally cleaved by a host protease* into viral surface protein gp120, and membrane-anchored gp41 (Montero *et al.* 2008), which remain non-covalently associated as gp41-gp120 heterodimers (Roux and Taylor 2007). These heterodimers assemble in triads to form 'spikes' that protrude from the viral surface (Hartley *et al.* 2005; Roux and Taylor 2007), mediating fusion and entry to host cells, determining co-receptor tropism, and providing the primary target for the humoral immune system (Montefiori *et al.* 2007; Roux and Taylor 2007).

Gp41 anchors the envelope proteins to the viral membrane. Spanning ~345 residues, it is divided into an extracellular domain, a transmembrane region and a cytoplasmic tail. At the time of infection, structural and conformational changes in gp41 facilitate

* FURIN, a paired basic amino acid cleaving enzyme

merging of viral and host cell membranes (Montero *et al.* 2008). In an intact virion, gp41 is shielded from antibodies by overlying gp120, but portions of gp41 may be exposed during the fusion process, making this region a potential target for antibody vaccine strategies (Montero *et al.* 2008; Sun *et al.* 2008).

Gp120, a protein of 500 residues, is divided into variable regions V1-V5, and is the most polymorphic region of the virus. The V3 region is a major determinant of tropism, defining the extent of co-receptor binding and thus affecting the course of pathogenesis (Hartley *et al.* 2005). Infection is initiated by binding of gp120 to CD4 receptors on the surface of the host cell.

1.3.5 Accessory and regulatory proteins

The accessory and regulatory proteins of HIV have a variety of roles that modify the host cellular environment to maximize viral infectivity, permit viral evasion of host defences, and enhance the production of new virions (Malim and Emerman 2008).

Nef (Negative Factor protein)

Nef is an accessory protein of 200-215 amino acids in length, transcribed in large copy numbers early in infection, and performing a number of independent functions contributing to HIV-1 pathogenicity and infectivity (Foster and Garcia 2008). By reducing cell surface expression of HLA-A and –B, Nef contributes to viral evasion of $CD8^+$ T cell responses (Lewis *et al.* 2008); cells infected *in vitro* with Nef-deleted HIV express around ten-fold more MHC compared to wild-type (Ali *et al.* 2005). Nef also mediates down-regulation of host cell surface expression of CD4, CXCR4 and CCR5, reducing superinfection, and increasing the production of viable virions

(Iafrate *et al.* 2000; Foster and Garcia 2008). Other roles of Nef include mediation of cellular signalling and activation, for example by activation of host p21-activated protein kinase (PAK-2) (O'Neill *et al.* 2006).

Nef-deleted HIV and SIV variants remain viable, but are significantly attenuated. In initial studies of Rhesus macaques, Nef-deleted SIV not only appeared non-pathogenic, but also protected against subsequent intra-venous challenge with wild type SIV, promising potential hope for a Nef-deleted vaccine (Kestler *et al.* 1991; Daniel *et al.* 1992). Studies of smaller Nef deletions and point mutations in SIV also report initial attenuation of disease in parallel with an *in vitro* reduction in viral replication; however, many animals infected with these variants subsequently develop high viral loads (Iafrate *et al.* 2000; Brenner *et al.* 2006). In a study of infant macaques infected with multiply deleted SIV viruses, AIDS was likewise induced in the majority of animals (Baba *et al.* 1999).

The best characterised example of Nef-deleted HIV is the Sydney Blood Bank cohort (Gorry *et al.* 2007). Ten subjects were infected with a virus containing a deletion of ≥150 base pairs in the Nef/LTR overlap region, of whom three maintained undetectable viral loads and stable CD4 counts for 14-18 years without HAART (Learmont *et al.* 1999). However, increase in viral load was documented in one subject who died, in two surviving subjects, and in the donor (Learmont *et al.* 1999). Thus, despite the initial hopes of a Nef-attenuated vaccine, these observations confirm that ultimate progression to AIDS is seen in subjects with Nef-defective HIV (Learmont *et al.* 1999; Gorry *et al.* 2007).

Tat (Transactivator of Transcription)

Tat is a regulatory factor of up to 86 amino acid residues, essential for HIV gene expression. In infected cells, it localizes primarily to the nucleus/nucleolus, where it binds the LTR region to initiate transcription of viral RNA. Low levels of Tat can be identified both within chronically infected cells, and free in the extracellular media (HIV Sequence Compendium 2009).

Rev (Regulator of Expression of Viral proteins)

Rev is a phosphoprotein that transits between the nucleus and cytoplasm of infected cells. Highly conserved among lentiviruses, it is responsible for stabilization and nuclear export of viral mRNA (HIV Sequence Compendium 2009). It binds the Rev Response Element (RRE) in the *Env* gene of viral mRNA to accelerate transport into the cytoplasm; as such it is pre-requisite for viral propagation.

Vif (Viral Infectivity Factor), Vpr (Viral protein R) and Vpu (Viral protein U)

These three accessory proteins all recruit host ubiquitin ligases to induce polyubiquitinylation and proteasomal degradation of cellular targets that would otherwise interfere with pathogenicity (Malim and Emerman 2008). Vif is a cytoplasmic protein common to most lentiviruses, with a role in promoting the infectivity of cell-free virus (HIV Sequence Compendium 2009). Vif acts by inhibiting the action of the host restriction factor APOBEC-3G (Sheehy *et al.* 2002; Goila-Gaur and Strebel 2008); (see section 1.8.4). Vpr interacts with the p6 component of Gag, and localizes primarily to the nucleus of infected cells. It performs a variety of functions, including nuclear import of HIV pre-integration complexes, regulation of reverse transcription and alterations in host cell cycle including

induction of apoptosis (HIV Sequence Compendium 2009; Le Rouzic and Benichou 2005). Vpu contributes to ubiquitin-dependent down-regulation of CD4, enhances release of progeny virus from the host cell and plays a role in Env maturation (Nomaguchi *et al.* 2008).

1.4 HIV diversity and evolution

1.4.1 Relationship between HIV and SIV

HIV is closely related to Simian Immunodeficiency Virus (SIV). SIV is non-pathogenic in its natural hosts - African non-human primates including sooty mangabeys (*Cercocebus atys)*, African green monkeys (*Cercopithecus aethiops*) and chimpanzees (*Pan troglodytes*) - even in the presence of high viral loads. However, SIV is pathogenic in the rhesus macaque (*Macaca mulatta*), making this the most widely used animal model of HIV*.

HIV-1 infection emerged in humans after transmission of SIV from chimpanzees in central Africa, with three separate transmission events suggested by the phylogenetic division of HIV-1 into three groups M, N and O interspersed between SIV lineages (Wain *et al.* 2007) (Fig. 1.4A). The related virus, HIV-2, arose in humans after separate transmission events from sooty mangabeys. HIV-2 accounts for a minority of infections, with cases largely confined to West Africa; although it also leads to progressive $CD4^+$ T cell depletion, it is characterised by longer latency, slower disease progression and lower viral loads than HIV-1 (Chen *et al.* 1997). Throughout the rest of this thesis, the term 'HIV' is used to refer to HIV-1 infection.

* The Indian rhesus macaque is the most frequently used model, but Burmese and Chinese rhesus macaques are alternative models of HIV. Other species, including cynomolgus macaques (*M. fascicularis*) and pig-tailed macaques (*M. nemestrina*) are less commonly used models.

1.4.2 The origin of clades

The origins of HIV can be traced to a common ancestral sequence calculated to have arisen in the early 1930's (Korber *et al.* 2000). M-group viruses are the most prevalent, and the most diverse, of these groups, accounting for the vast majority of current infections. M-group viruses are phylogenetically divided into subtypes (clades) (Fig. 1.4B), genetic sub-groups that are approximately genetically equidistant, and are defined primarily by sequence differences at the nucleotide level (Robertson *et al.* 2000). Increasing numbers of Circulating Recombinant Forms (CRFs) are also documented; these are recombinant lineages that play an important role in the HIV pandemic (Robertson *et al.* 2000; Fig. 1.5).

There are several strands of evidence to suggest that these clades arose from founder strains in Africa rather than diverging subsequently as a consequence of immunological selection pressure in different human populations (Peeters and Sharp 2000; Vidal *et al.* 2000). First, the star-burst appearance of the M-group phylogenetic tree suggests near simultaneous evolution of viral subtypes from SIV transmission events (Rambaut *et al.* 2001). Second, analysis of envelope sequences from HIV-1 strains circulating in central Africa reflects the enormous genetic diversity of HIV in this region, encompassing sequences from all M-group clades (Nkengasong *et al.* 1994; Vidal *et al.* 2000; Rambaut *et al.* 2001). Additionally, study of phylogeny and geographic routes of dissemination of the different clades suggests direct transmission events from African founder viruses (McCutchan *et al.* 1996; Vidal *et al.* 2000; Gilbert *et al.* 2007; Thomson *et al.* 2007).

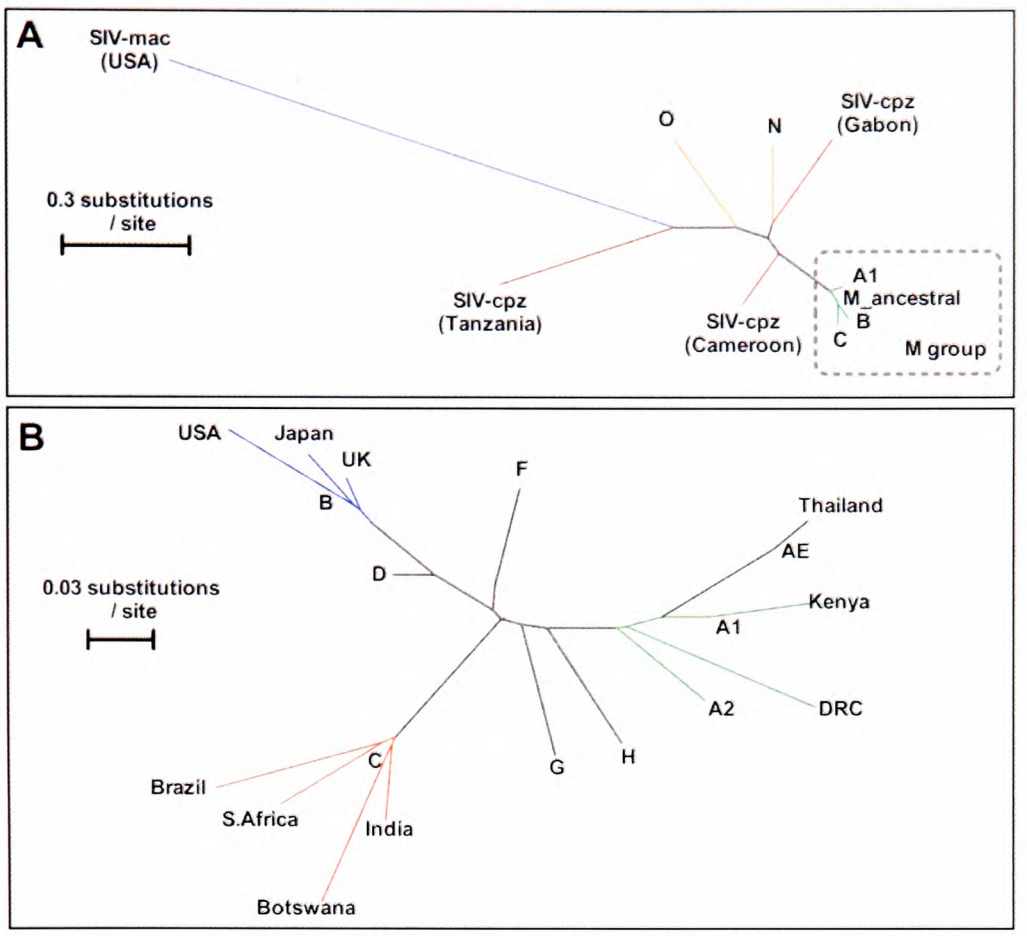


Figure 1.4: Maximum likelihood phylogenetic trees to show relationship between SIV and HIV-1, and between HIV-1 subtypes within M group. Reference strains for each subtype obtained for *gag* genes from Los Alamos Sequence Compendium 2009 (HIV Sequence Compendium 2009); trees constructed using GARLI software. A: Relationship between SIV-cpz (chimpanzee) and SIV-mac (macaque), and HIV-1 groups M (major), O (outlier) and N (non-M, non-O). B: Relationship between HIV-1 reference strains within M group. Consensus strains for each subtype are marked with individual letters; reference strains for countries are labelled by country. The subtypes accounting for the majority of global infections are sub-type A - green, B - blue, and C - red.

The presence of distinct viral sub-clades in different geographical locations – for example, the characteristic clustering of C-clade Indian viruses – is also suggestive of descent from a single common ancestor (Gaschen *et al.* 2002). Likewise, recombinant viruses circulating in Thailand form a distinct CRF_AE sub-cluster, as well as bearing similarities to strains that have been identified in central Africa (McCutchan *et al.* 1996).

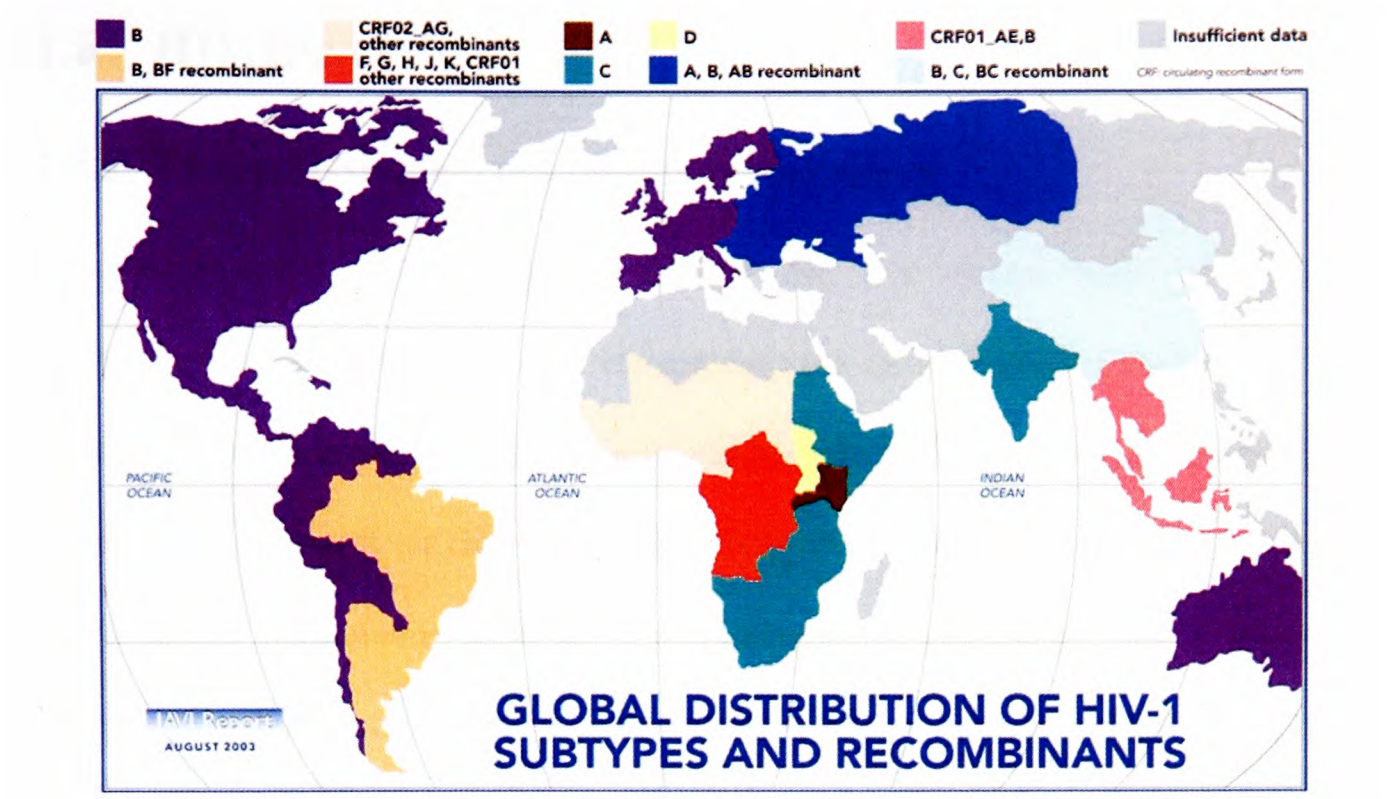


Figure 1.5: Global distribution of HIV-1 subtypes and recombinant strains (CRFs, circulating recombinant forms). Figure reproduced from International AIDS Vaccine Initiative (IAVI) at www.iavi.org.

1.4.3 Factors driving HIV-1 diversity

HIV phylogeny is strongly influenced by founder effect – that is, differences in the ancestral strains that gave rise to sub-clusters of infection within the epidemic (Bhattacharya *et al.* 2007). The continued rapid genetic diversification of HIV-1 is driven by the combined influence of selectively neutral genetic drift (Shriner *et al.* 2004), plus positive selection pressure on the virus from the combined selective pressures of host immune responses (McMichael and Klenerman 2002; Moore *et al.* 2002; Leslie *et al.* 2004; Leslie *et al.* 2005; Wain *et al.* 2007) or anti-retroviral therapy (Spira *et al.* 2003; Lemey *et al.* 2005; Little *et al.* 2008). In the context of rapid rates of HIV replication (Ho *et al.* 1995), the potential for recombination events, and lack of RT proof reading function (Katz and Skalka 1990; Preston 1997), annual diversity of quasispecies within a single country outweighs the entire annual global genetic diversity of influenza viruses (Korber *et al.* 2001).

1.5 HIV life cycle

1.5.1 Transmission across mucus membranes

In intact mucosal surfaces, models suggest that free virions may penetrate gaps between epithelial cells or invade via microabrasions in epithelial surfaces to reach the deeper layers of the mucosa (Hladik and McElrath 2008). Despite the diversity of quasispecies within a single individual, a bottleneck effect operates at transmission across mucus membranes, such that a median of one virus (range 1-5) is responsible for initiating HIV infection in a new host (Greenier *et al.* 2001; Keele *et al.* 2008; Kearney *et al.* 2009). There is conflicting evidence about the selection of viral variants for mucosal transmission; some data suggest that the process may be stochastic (Greenier *et al.* 2001), while other studies point to the preferential transmission of fitter viral variants (Miller 1998; Overbaugh *et al.* 1999).

1.5.2 Fusion and entry to host cell

Before the process of fusion and entry to host cells, virions may adhere to target cells using a variety of 'attachment receptors'. These include the lectin receptor DC-SIGN (**D**endritic **C**ell **S**pecific **I**ntercellular Adhesion Molecule-3 **G**rabbing **N**on-integrin), present on dendritic cells (DCs) and macrophages (de Witte *et al.* 2008), and integrin adhesion molecules in gut mucosa (Sattentau 2008).

HIV gains entry into a host cell via interaction between Env proteins and a primary receptor, CD4, and a co-receptor – either chemokine receptor CCR5 (R5 tropic viruses) or CXCR4 (X4 tropic viruses) (Hartley *et al.* 2005; Sattentau 2008). Following gp120 binding to CD4, conformational changes in Env protein occur, permitting co-receptor binding and fusion of the virus lipid envelope with the host cell

membrane (Hartley *et al.* 2005; Johnston and Fauci 2007; Montero *et al.* 2008; Sattentau 2008). The initial interaction prompts alterations in the cytokeleton of the host cell, recruiting further receptors to the site of virus binding (Sattentau 2008).

R5-tropic (also termed macrophage tropic) virions are usually responsible for initiation of new infections (Keele *et al.* 2008), while X4 (T cell tropic) variants generally evolve within individual patients during the course of disease (Clapham and McKnight 2001), most commonly in subtype B infection (Philpott 2003). Uncommon host CCR5 deletions (CCR5-Δ32) or mutations are associated with delayed progression to AIDS (Smith, Dean *et al.* 1997) and prevent infection by R5 viral strains. Compared to R5 strains, X4-tropic viruses are associated with faster replication *in vitro* (Pollakis *et al.* 2004), and with accelerated disease and higher infectivity *in vivo* (Philpott 2003; Tsai *et al.* 2007). These differences may be a consequence of the greater genetic complexity of X4 viruses (Tsai *et al.* 2007), or the wider spectrum of cells that are potentially susceptible to infection with X4 strains (Philpott 2003).

1.5.3 Cells targeted by HIV infection

The primary cells targeted by HIV infection are $CD4^+$ T lymphocytes (T helper, T_H cells). These cells arise in the thymus as naïve T cells, and mature upon stimulation with antigen presented by MHC Class II molecules (Fig. 1.6). Following $CD4^+$ T cell activation, autocrine secretion of IL-2 prompts further expansion of the $CD4^+$ population (Brenchley *et al.* 2004).

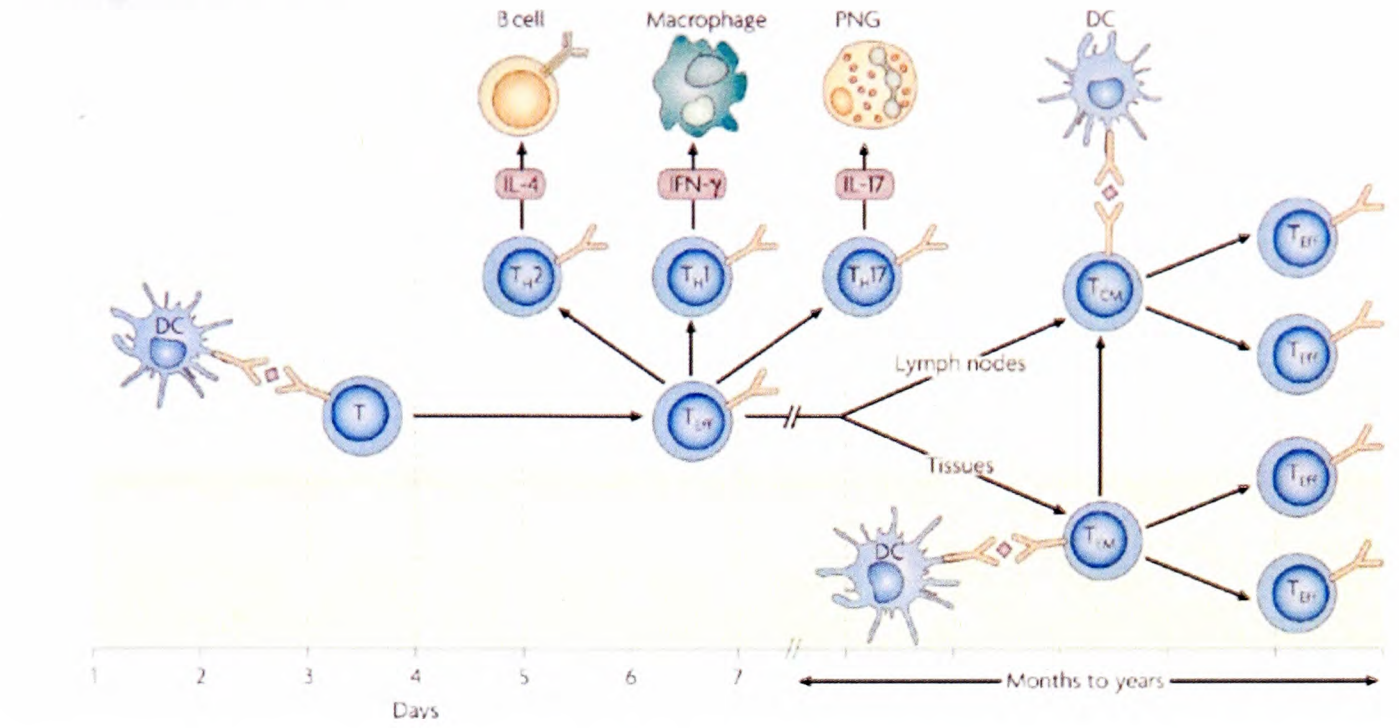


Figure 1.6: T_H cell differentiation pathways. Following contact with antigen presented by antigen presenting cells such as dendritic cells (DC), T cells differentiate into effector (T_{EFF}) cells, in three main subsets T_H1, T_H2 and T_H17. T_H1 cells produce IFN-γ, stimulate macrophages and drive proliferation of $CD8^+$ T cells. T_H2 cells stimulate the humoral immune system via the production of IL-4. The T_H17 subset is characterised by the production of IL-17 and IL-22, and performs a regulatory function, particularly critical in neutrophil recruitment and defence at mucosal surfaces (Brenchley *et al.* 2008). Memory T cell subsets also develop: central memory (T_{CM}) in lymph nodes, and effector memory (T_{EM}) in tissues. Figure reproduced from Kaufmann 2007 with permission from the Nature Publishing Group, licence number 2216960870820.

Other cell types bearing CD4 surface receptors are also susceptible to infection. Within the monocyte-macrophage lineage, a large pool of circulating peripheral monocytes can migrate into tissue and differentiate either into macrophages, or DCs (Chomarat *et al.* 2000). DCs ($CD4^+CCR5^+$DC-$SIGN^+$) are professional antigen presenting cells, constitutively expressing MHC Class II, that can either be productively infected (*cis*-infection), or can transport and disseminate infectious particles without themselves being infected (*trans*-infection) (Alexaki *et al.* 2008; Hladik and McElrath 2008; Sattentau 2008). Less commonly, cells lacking CD4 surface receptors, but possessing co-receptors CCR5 and/or CXCR4 are infected,

although the role of these cells in production and dissemination of progeny virus is uncertain (Alexaki *et al.* 2008).

1.5.4 Reverse transcription and integration of HIV genome

Each virion contains two copies of genomic RNA (Hu and Temin 1990) that are non-covalently linked at their 5' ends. Reverse transcription is initiated in mature virions, but completed in the cytosol of a host cell (Le Rouzic and Benichou 2005). An HIV 'pre-integration complex' (PIC) is then assembled from the newly reverse-transcribed HIV DNA, Gag p17, integrase and Vpr (Hearps and Jans 2007) for transport through the nuclear envelope of the host cell. The viral genome is integrated into the host DNA, and thereafter utilises host machinery for protein transcription.

1.5.5 Production and transmission of progeny virions

Within 24h of HIV-1 infection, new virions are released from the surface of infected cells; an estimated mean of $0.7x10^9$ new virions are produced daily (Ho *et al.* 1995). Unlike other retroviruses, HIV is able to replicate in terminally differentiated and non-dividing cells (monocytes and macrophages) (Le Rouzic and Benichou 2005), although active proliferation in T lymphocytes does require cellular DNA synthesis (Weinberg *et al.* 1991).

The macaque model has been used to track the passage of SIV following acute atraumatic genital tract infection (Spira *et al.* 1996; Hu *et al.* 2000): within two days of infection, abundant infected cells (mostly DCs) can be detected in the lamina propria of the vaginal epithelium and in the endocervix. Acutely infected cells then migrate to local lymph nodes via afferent lymphatics (Hladik and McElrath 2008),

and proviral DNA can be detected in the draining lymph nodes by day 2 (Spira *et al.* 1996).

The traditional model of viral dissemination is by budding from infected cells to release cell-free virions. However, direct cell-to-cell transmission also occurs (Sattentau 2008) via transient synapses that form between the plasma membranes of infected host cells and uninfected T cells, leading to clustering of cell surface receptors, release of virus into the synaptic cleft and viral fusion with the target cell (Groot *et al.* 2008; Sattentau 2008). Retroviruses can also be transmitted along filopodial bridges between cells, or via actin-mediated transport along nanotubes (Sowinski *et al.* 2008; Eugenin *et al.* 2009). These methods of viral dissemination are likely to be more rapid that spread by cell-free virus, facilitate infection of multiple target cells by a single infected host cell, disrupt healthy immunological synapses, and potentially contribute to evasion of antibody neutralisation (Sattentau 2008).

1.5.6 Gut immunopathology

Up to 80% of the total $CD4^+$ pool is found as GALT (Gut Associated Lymphoid Tissue), particularly as mucosal memory cells (Douek 2007), occurring either in focal Peyer's patches and lyphoid follicles, or as scattered diffuse $CD4^+$ T cells (Veazey *et al.* 1998). Irrespective of the original route of infection, there is evidence for early severe depletion of $CD4^+CCR5^+$ cells in the gut, leaving a damaged mucosal surface susceptible to bacterial translocation (Douek 2007; Hladik and McElrath 2008; Sattentau 2008). In the SIV model, severe CD4 depletion in the gut is apparent by day 7 after acute infection, but is also sustained in more chronically infected animals (Veazey *et al.* 1998).

1.5.7 Viral latency and reservoirs of infection

One of the mechanisms by which HIV is able to evade both natural immune responses and therapeutic strategies is through the establishment of latent reservoirs of infection in relatively long-lived target cells (Chun *et al.* 1998; Alexaki *et al.* 2008). Such reservoirs are established early in acute infection, and are not affected by the delivery of early HAART, even when peripheral HIV RNA is rapidly suppressed (Chun *et al.* 1998). Latency can occur pre-integration (reverse transcription or nuclear import is halted) or post-integration (pro-viral DNA is integrated with the host genome but translation is limited or absent) (Alexaki *et al.* 2008). Cell populations that contribute to this latent reservoir include resting $CD4^+$ T cells (naïve or memory T cells), macrophages and DCs (Groot *et al.* 2008; Sattentau 2008).

1.6 Dynamics and clinical correlates of HIV infection

1.6.1 Relationship between $CD4^+$ T cell count, viral load and disease

Following acute infection in adults, there is a rapid peak in viraemia to 10^4-10^8 RNA copies / ml plasma, which may be accompanied by 'seroconversion illness' characterised by influenza-like symptoms, rash and lymphadenopathy (Daar *et al.* 1991; Clark and Shaw 1993). This is followed by an asymptomatic period (median 10 years (Mellors *et al.* 1996)) during which viral load in plasma reaches a 'set-point' plateau (median 30 – 40,000 HIV RNA copies / ml (Kiepiela *et al.* 2004)). This set-point is one of the most reliable clinical predictors of the time to AIDS (Mellors *et al.* 1996; Schacker *et al.* 1998; Fig. 1.7).

Following loss of viraemic control, $CD4^+$ T cell count declines, rendering the host susceptible to AIDS at an estimated rate of 2.8-4.2% per annum (Tindall *et al.* 1988).

The timing of these events in different individuals is likely to reflect the balance between host immune responses and the fitness of circulating viral quasispecies. In the absence of HAART, the associated opportunistic infections and malignancies that occur when $CD4^+$ T cell counts are <250 cells/mm^3 render AIDS almost universally fatal.

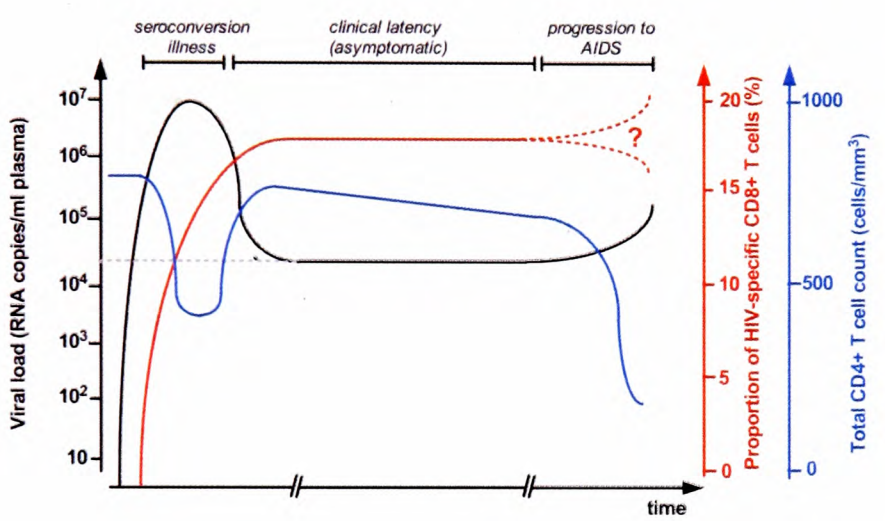


Figure 1.7: Temporal changes in HIV-1 viral load, HIV-specific $CD8^+$ T cells and total $CD4^+$ T cell count following acute adult infection. $CD4^+$ T cell count shown in blue, HIV-specific $CD8^+$ T cells in red, viral load in black. Dashed grey line indicates set-point. Figure adapted from Goulder and Watkins 2004.

1.6.2 Mechanisms of $CD4^+$ T cell decline

Progressive decline in $CD4^+$ T cell count is brought about, in part, by direct lysis of infected cells by viral infection. However, given the relatively small proportion of the total $CD4^+$ T cell pool that is productively infected, this mechanism of cell death does not exclusively account for the overall depletion (Hazenberg *et al.* 2000; Brenchley *et al.* 2004). The expression of PD-1 (Programmed-Death 1), a marker of an exhausted phenotype, is increased on the surface of $CD4^+$ T cells and HIV-specific $CD8^+$ T cells from chronically infected individuals and correlates with low CD4 counts (Day *et al.* 2006); exhaustion of $CD4^+$ T cells may occur as a direct consequence of increased turnover, although there is no shortening of telomere lengths to support this hypothesis (Wolthers and Miedema 1998).

In the context of chronic HIV infection, immune activation indirectly leads to loss of $CD4^+$ T cells, as well as B cells, NK cells and $CD8^+$ T cells (Hazenberg *et al.* 2000) (Moir and Fauci 2009). This 'bystander effect' due to immune activation may arise, at least in part, as a consequence of bacterial translocation across a leaky and damaged gut, evidenced by high systemic levels of bacterial lipopolysaccharde (LPS) (Brenchley *et al.* 2006; Douek 2007). HIV also leads to chronic dysregulation of homeostatic mechanisms that operate between $CD4^+$ T cell subsets, with a variety of consequences including increased turnover of memory T cells (Brenchley *et al.* 2004).

1.7 HLA Class I and $CD8^+$ T cell mediated immunity

1.7.1 Evidence for the role of $CD8^+$ T cells in control of viraemia

There are several strands of evidence for the importance of $CD8^+$ T cells in the control of HIV viraemia. *In vivo* studies demonstrate that, following acute infection, the decline in peak viraemia is temporally associated with a rise in the HIV-specific $CD8^+$ T cell population, suggesting that these cells play a role in the initial control of viral replication (Borrow *et al.* 1994; Koup *et al.* 1994). Subsequent studies in the SIV-macaque model have confirmed the effect of blocking this $CD8^+$ T cell response*: in acute infection, viraemia remains high and sustained, and in chronic infection viral load rebounds, suggesting uncontrolled viral replication (Jin *et al.* 1999; Schmitz *et al.* 1999).

The strong and consistent associations between HLA Class I genotype and disease outcome in large cross-sectional cohorts also points to $CD8^+$ T cell-mediated control of disease (O'Brien *et al.* 2001; Kiepiela *et al.* 2004; Brumme, Tao *et al.* 2008;

* The $CD8^+$ T cell population in macaques includes NK cells.

Matthews *et al.* 2008; Rousseau *et al.* 2008). A genome-wide study also pointed to the MHC locus as being predictive of disease control (Fellay *et al.* 2007), although this does not specifically pinpoint HLA Class I as the determinant of outcome, as many other genes are also contained in this region. The precise relationship between HLA and viraemic control, and the mechanisms underpinning this are reviewed in the following sections of this chapter.

1.7.2 Evolution and Genetics of HLA

Human Leucocyte Antigen (HLA) Class I molecules on the surface of infected cells present antigen to $CD8^+$ T cells, and thus orchestrate the adaptive immune response. HLA is a member of the immunoglobulin gene superfamily (Bodmer, 1987), encoded in the Major Histocompatibility Complex (MHC) region on Chromosome 6p.

The HLA B locus is the most diverse in the MHC region, which is itself the most diverse region of the human chromosome (Human Genome Project, www.sanger.ac.uk/HGP/Chr6/). This enormous genetic variability allows the adaptive immune system to present a vast range of antigens. The close proximity of alleles at the A, B and C loci determines transmission of combinations of alleles in linkage disequilibrium (LD), with the strongest associations between HLA-B and Cw loci as a result of the molecular distance between the genes (Bodmer, 1987; Fig. 1.8).

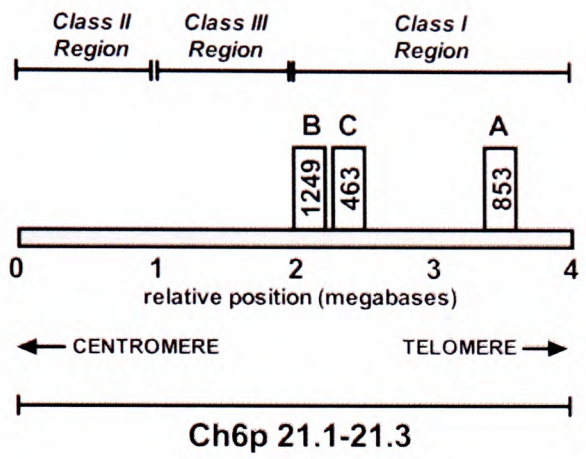


Figure 1.8: Schematic view of the MHC region of Chromosome 6.

HLA Class I Loci A, B and C are shown, with the number of alleles identified enclosed in the respective boxes. The Class II region encodes DP, DQ, DR, DM, while the Class III region contains genes for complement components and factor B. Figure adapted from references (Marsh *et al.* 2000; Mehra and Kaur 2003); number of alleles updated from the IMGT/HLA database based on data from July 2009 (http://hla.alleles.org/class1.html).

1.7.3 HLA structure and peptide binding

Heavy and light chains of HLA Class I molecules

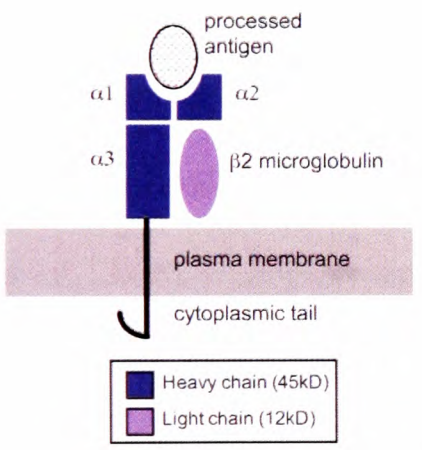


Figure 1.9: Schematic view of an HLA Class I molecule on the cell surface.

The molecule comprises a variable heavy chain composed of three α domains (362-366 amino acids encoded on Ch6p) plus an invariant light chain (β_2-microglobulin, 90 amino acids, encoded on Ch15). There is a hydrophobic trans-membrane region and a short cytoplasmic tail (Bodmer 1987; Marsh *et al.* 2000; Janeway *et al.* 2005). The α1 and α2 domains form the peptide binding site. All class I molecules at the cell surface contain bound antigen, which may be 'self' or pathogen-derived (Marsh *et al.* 2000).

A schematic structure of an HLA Class I molecule is shown (Fig. 1.9). The antigen binding repertoire of an HLA Class I molecule is determined by polymorphic residues in the α_1 and α_2 chains that form the peptide binding groove, composed of six pockets designated A-F. Typically, the most specific regions of the binding groove are the anchor positions, where pockets B and F bind position 2 and the C-terminal residue of a peptide antigen respectively, by strong networks of hydrogen bonds (Parham 1992). However, for some alleles, C, D, and other pockets may play a more important role in

determining peptide binding. Within these constraints, a given HLA allele can bind epitopes composed of different residues (at non-anchor positions) and of differing lengths, as longer peptides can bulge out of the peptide binding groove (Parham 1992).

Processing of peptides for HLA Class I binding

HIV peptides are degraded in the cytosol by the 26S proteasome and other cellular proteases, releasing peptides of between 2 and 25 amino acids length. These peptides are transported into the endoplasmic reticulum (ER) by a heterodimeric complex of TAP proteins (**T**ransporters Associated with **A**ntigen **P**rocessing, TAP1 and TAP2, encoded in the MHC region of Ch6) (Janeway *et al.* 2005; Peaper and Cresswell 2008). Within the ER, enzymes of the ERAP family (**E**ndoplasmic **R**eticulum **A**mino-**P**eptidases) trim peptide fragments to customize them to the required 8-11 amino acids required for HLA Class I binding (Serwold *et al.* 2002; Shastri *et al.* 2005; Yan *et al.* 2006). Complete assembly of the HLA-peptide complex (Fig. 1.10) occurs in association with the MHC-Class I specific chaperone tapasin before export from the ER for cell surface presentation (Janeway *et al.* 2005; Peaper and Cresswell 2008).

Processed peptides from intra-cellular pathogens must bind HLA sufficiently well to out-compete binding of self-antigens within the ER. The strong avidity of epitope binding also ensures that the peptide remains bound, stabilising the HLA molecule (Marsh *et al.* 2000), and avoiding the binding of free extra-cellular peptides to HLA molecules on healthy cells that might result in immune destruction of a healthy cell (Marsh *et al.* 2000; Janeway *et al.* 2005). Empty HLA Class I molecules on the cell surface rapidly break down, and the α chain is re-internalised.

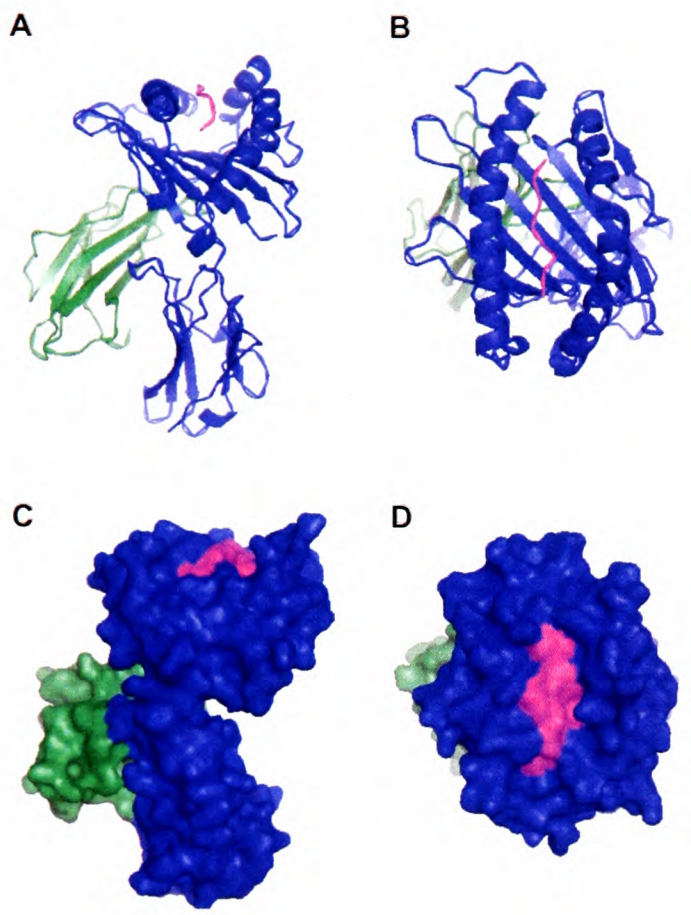


Figure 1.10: HLA-B*3501 bound to HIV epitope VY9 (VPLRPMTY, Nef 75-82).

Three alpha domains of the HLA Class I molecule are shown in blue, β_2 microglobulin in green and the Nef epitope in pink. A,B Stuctural view showing helices and β pleated sheets; C,D surface view. A,C side projection; B,D viewed from above. Images reconstructed using MacPymol from crystal structures solved by Smith et al (Smith *et al.* 1996).

Classification of HLA-A and B alleles

Despite the huge diversity of peptide binding by different HLA Class I molecules, HLA-A and -B alleles can be grouped into twelve major supertypes (Table 1.1) according to the binding specificities of the allele, determining the anchor positions of the presented peptide (Sidney *et al.* 2008). HLA-B alleles are also divided into two groups, Bw4 and Bw6, on the basis of differences in the amino acid sequences at the C-terminal end of the α_1-helix. These residues are critical for determining the peptide binding groove adjacent to the F-pocket and thus mediate a strong influence on epitope specificity (Salter and Parham 1989; Flores-Villanueva *et al.* 2001).

Supertype	Examples of alleles included in supertype
A01	A*0101, A*2601
A01 A03	A*3001, A*3008
A01 A24	A*2902
A02	A*0201, A*0202, A*6802
A03	A*0301, A*1101, A*74
A24	A*2301, A*2402
B07	B*0702, B*3501, B*3910, B*4201, B*5101, B*5301, B*8101
B08	B*0801, B*0802
B27	B*2705, B*1402, B*1503
B44	B*1801, B*4402, B*4403, B*45
B58	B*1516, B*5701, B*5703, B*5801, B*5802
B62	B*1501

Table 1.1: Twelve HLA-A and B supertypes, with examples of common alleles in each type; adapted from Sidney *et al.* 2008.

1.7.4 Antigen presentation to T Cell Receptors (TCR) by HLA

Each $CD8^+$ T cell bears about 10-30,000 identical TCRs, each composed of a heterodimer most commonly composed of α and β chains (Janeway *et al.* 2005; Fig. 1.11). Regions of hypervariability in the α and β chains are known as complementarity-determining regions, combinations of which expand the diversity of the TCR repertoire (Turner *et al.* 2006).

The TCR binds the α_1 and α_2 domains of the HLA Class I molecule in association with the exposed portion of the bound antigen, although most of the bound peptide may be buried in the heavy chain of the HLA molecule (Fig. 1.10), leaving only a few side chains accessible for TCR recognition. A single TCR clone may bind the same antigen fragment presented by different HLA Class I molecules, suggesting that changes in T cell response may be dictated by the antigen rather than by the HLA

molecule *per se* (Tomiyama *et al.* 2000; Ueno *et al.* 2002). The MHC-peptide complex presented on the surface of an infected cell is also weakly bound by a co-receptor molecule, CD8 or CD4 (Janeway *et al.* 2005).

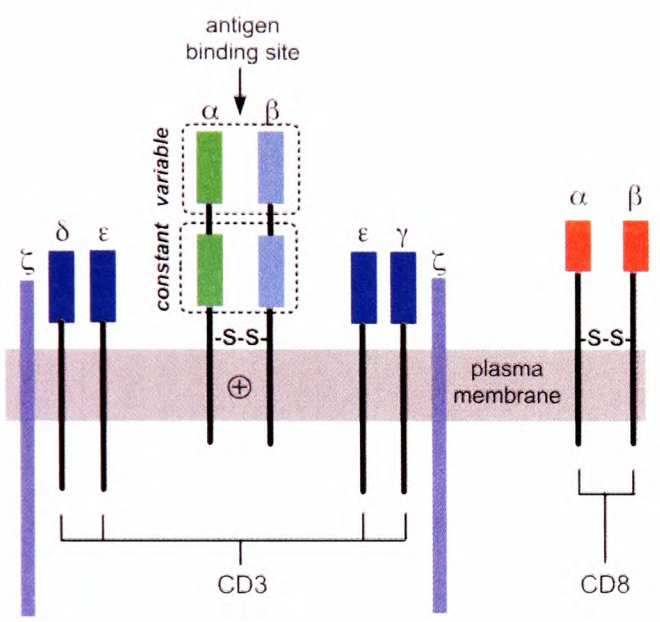


Figure 1.11: T cell receptor (TCR) αβ complex.

Each TCR is composed of a heterodimer of α and β chains, each composed of a variable and constant region, and linked by a disulphide bond (–S-S-). Positively charged trans-membrane domains allow association with membrane components. CD8 is a co-receptor for the HLA-peptide complex. Figure adapted from Janeway (Janeway *et al.* 2005).

1.7.5 $CD8^+$ T cell-mediated immune escape

Immune escape in HIV was first described in 1991 (Phillips *et al.* 1991), demonstrating that the selection of amino acid polymorphisms in viral sequence can reduce or abrogate HLA Class I presentation and subsequent TCR recognition. Thus, from a large number of circulating viral quasispecies, those bearing mutations at specific positions that interfere with epitope processing and presentation are selected to outgrow wild-type.

The timing and consequences of selection of escape varies. If escape mutations are selected early, almost all subjects with the selecting allele will carry the mutation (E.g. HLA-B*57 selection of Gag escape T242X in the TW10 epitope (Leslie *et al.* 2004; Martinez-Picado *et al.* 2006); HLA-B*51 selection of RT escape I135X in the

TI8 epitope (Tomiyama *et al.* 1999; Frater *et al.* 2007; Kawashima *et al.* 2009)). Accumulation of some of these mutations, despite allowing escape from immune responses, may confer a fitness cost, contributing to control of viraemia (Martinez-Picado *et al.* 2006; Goepfert *et al.* 2008; Matthews *et al.* 2008; Boutwell *et al.* 2009). Conversely, other mutations are not selected until late in the course of disease, but may be associated with dramatic loss of immune control (E.g. Mamu-A*01 selection of Gag escape T182A in the CM9 epitope (Barouch *et al.* 2002); HLA-B*27 selection of Gag escape R264K in the in KK10 epitope (Goulder, Phillips *et al.* 1997; Feeney *et al.* 2004; Chen *et al.* 2009)).

Escape mutations permit evasion of $CD8^+$ T cell responses, to varying extents, via a number of different mechanisms (Fig. 1.12).

Mutations that affect peptide processing for HLA Class I presentation

These mutations can bring about escape by reducing or abrogating cleavage by the 26S proteasome, TAP-mediated transport into the ER, or post-proteasomal processing in the ER (Allen *et al.* 2004; Draenert *et al.* 2004; Yokomaku *et al.* 2004; Milicic *et al.* 2005; Zimbwa *et al.* 2007). Sequence changes outside epitopes are most associated with this mechanism of escape, although changes within epitopes can also mediate processing effects (Sewell *et al.* 2002; Yokomaku *et al.* 2004). ERAP1 trimming enzymes in the ER are unable to cleave an N-terminal proline (Serwold *et al.* 2001), so substitutions to proline may particularly affect this processing, as in the case of the A146P mutation flanking the HLA-B*57 restricted epitope ISW9 (ISPRTLNAW, Gag 147-155) (Draenert *et al.* 2004).

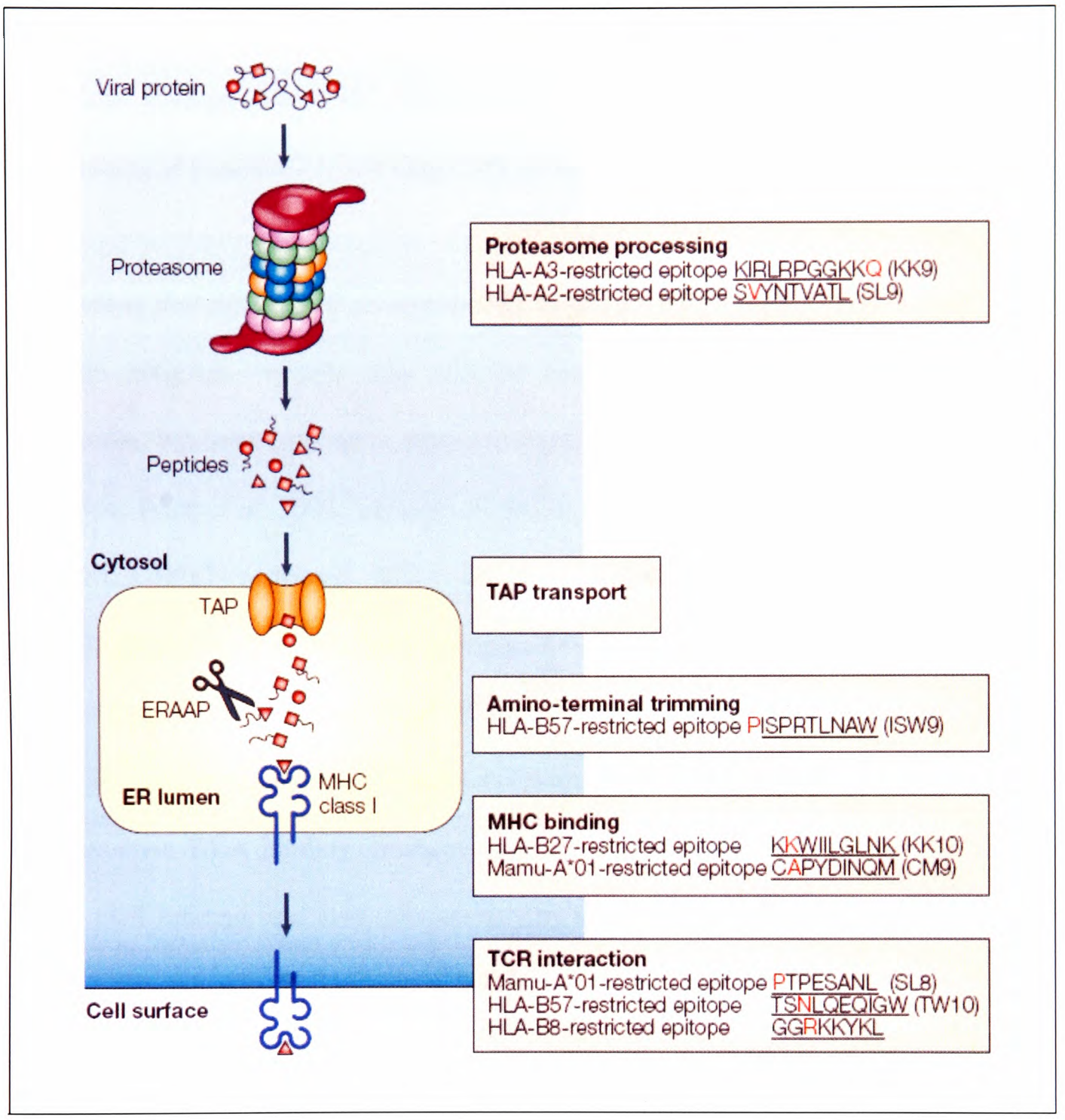


Figure 1.12: Mechanisms of mutational escape in HIV and SIV. A schematic view of antigen processing for cell surface presentation in association with MHC Class I is shown, with boxes showing steps at which mutational escape may arise. Example epitopes are listed with sites of mutation highlighted in red. ER, endoplasmic reticulum; TAP, transporter associated with antigen processing; ERAAP, Endoplasmic Reticulum Aminopeptidase associated with Antigen Processing. Figure reproduced from Goulder and Watkins, 2004, with permission from the Nature Publishing Group, licence number 2220071095739.

Mutations that affect binding of peptide to HLA Class I.

Mutations at the anchor positions (position 2 and C-terminus of the epitope) are particularly associated with this method of escape, exemplified by the HLA-B*27

R264K HIV escape mutation selected at position 2 of the Gag KK10 epitope (Goulder, Phillips *et al.* 1997; Feeney *et al.* 2004) and the Mamu-A*01 selection of SIV escape at position 2 of the Gag CM9 epitope (Barouch *et al.* 2002).

Mutations that affect TCR recognition by TCR on CD8$^+$ T cells.

Certain antigenic variants may still be bound and presented by HLA Class I molecules, but have reduced or absent recognition by circulating CD8$^+$ T cells (Lee *et al.* 2004; Price *et al.* 2004; Ueno *et al.* 2007). This is seen in the case of escape from the HLA-B*57 restricted KF11 (KAFSPEVIPMF, Gag 162-172) and TW10 (TSTLQEQIGW, Gag 240-249) epitopes, in which common escape variants reduce TCR recognition (Gillespie *et al.* 2006). However, mechanisms of escape are not mutually exclusive, and these mutations (particularly at position 2 in KF11) may also impact upon HLA binding. Immune control may be diminished initially by escape from TCR binding, and also subsequently by the generation of less effective CD8$^+$ T cells to the new variant (Ueno *et al.* 2007).

1.7.6 HLA and HIV transmission

The fate of escape mutations following transmission

Following transmission to an HLA-mismatched recipient, there are several possible consequences for escape mutations. Persistence of mutations in the absence of other sequence changes suggests that the mutation is not detrimental to viral infectivity or replicative capacity. In other cases, the original mutation is associated with a fitness cost to the virus, but the covariation of amino acid at other sites can provide structural or functional compensation (Gatanaga *et al.* 2006; Martinez-Picado *et al.* 2006; Brockman *et al.* 2007; Crawford *et al.* 2007; Schneidewind *et al.* 2008; Schneidewind

et al. 2009). Over time, these polymorphisms – both escape mutations and compensatory changes - can accumulate in populations and may eventually come to replace the original wild-type amino acid (Leslie *et al.* 2005; Kawashima *et al.* 2009).

Other HLA-selected mutations revert to the wild-type amino acid following transmission (Allen *et al.* 2004; Friedrich *et al.* 2004; Leslie *et al.* 2004; Li B 2007; Matthews *et al.* 2008; Crawford *et al.* 2009). In the SIV model, in animals infected with a virus bearing $CD8^+$ T cell escape mutations, reversion to wild type sequence has been observed as rapidly as 2 weeks (Fernandez *et al.* 2005), and progressively up to 86 weeks post-infection (Friedrich *et al.* 2004). *In vitro* fitness assays, in which mutant variants are out-competed by wild-type, support the hypothesis that reversion is driven by an uncompensated cost to viral fitness imposed by the mutation (Friedrich *et al.* 2004; Martinez-Picado *et al.* 2006; Liu *et al.* 2007).

Horizontal transmission

Studies of adult transmission pairs have highlighted the potential impact of HLA-selected escape polymorphisms arising in the donor on disease outcome in the recipient (Chopera *et al.* 2008; Goepfert *et al.* 2008; Crawford *et al.* 2009). $CD8^+$ T cell escape mutations in the transmitted virus that result in a fitness cost may lead to improved control of viraemia in the recipient, based on the acquisition of a partially crippled virus (Chopera *et al.* 2008). However, an accelerated disease course may be observed in recipients who share HLA Class I molecules with their donor, due to transmission of viral escape mutations rendering T cell epitopes unavailable for class I presentation. Accordingly, in a Zambian study, the transmission of increasing numbers of Gag mutations was associated with a progressive reduction in viral loads

only in recipients whose own HLA-B alleles do not select the same escape mutations in Gag (Goepfert *et al.* 2008).

Vertical transmission

MTCT represents a special case in the study of the impact of host HLA Class I on control of viraemia, as mother and child share at least half of their HLA Class I alleles. Increased HLA concordance between mother and child has been associated with increased risk of transmission (Mackelprang *et al.* 2008), and worse viraemic control following transmission (Thobakgale *et al.* 2009).

1.7.7 Determinants of a successful $CD8^+$ T cell response

The specific mechanism by which certain $CD8^+$ T cell responses are able to mediate sustained control of viraemia is incompletely understood (Barouch 2008). However, a growing body of literature has sought to define the correlates of successful immune control.

HLA genotype

The HLA class I genotype of the host is a significant determinant of HIV disease progression both in B-clade (O'Brien *et al.* 2001) and C-clade (Kiepiela *et al.* 2004) infection. HLA-B is the locus with the strongest association with control of viraemia (Kiepiela *et al.* 2004), potentially reflecting the increased diversity of HLA-B compared to HLA-A and HLA-Cw alleles, permitting a wider range of antigen presentation, or relating to the specific repertoire of epitopes presented by HLA-B molecules. Particular alleles, particularly in the HLA-Bw4 group, have well documented association with favourable control of viraemia (E.g. HLA-B*57,

B*5801, B*2705, B*5101). Others are characteristically associated with more rapid progression to AIDS, and mostly belong to the Bw6 group (E.g. HLA-B*1801, B*3502/03) (O'Brien *et al.* 2001). Likewise, in the macaque model, certain alleles are characteristically associated with disease control, such as Mamu-A*01 and Mamu-B*17 (Wilson *et al.* 2006; Yant *et al.* 2006).

HLA heterozygosity is associated with improved control of viraemia compared to homozygosity at class I locus (Carrington *et al.* 1999), suggesting that the breadth of $CD8^+$ responses plays an important role. In particular, possession of two different Bw4 alleles ('Bw4 homozygosity') confers the greatest advantage (Flores-Villanueva *et al.* 2001). Rare HLA alleles are also associated with improved viraemic control, as the virus is unlikely to have acquired escape mutations permitting immune evasion from these responses (Trachtenberg *et al.* 2003; Frahm *et al.* 2006).

Magnitude of $CD8^+$ T cell response

The proportion of $CD8^+$ T cells that are HIV specific has been found to vary widely, but has been quantified as up to 18-22% (Gea-Banacloche *et al.* 2000; Migueles and Connors 2001), with the majority of these cells responding to epitopes in Gag and Pol (Gea-Banacloche *et al.* 2000). However, the overall frequency of HIV-specific $CD8^+$ T cells, or the magnitude of individual responses, does not predict control of plasma viraemia (Gea-Banacloche *et al.* 2000; Addo *et al.* 2003; Chen *et al.* 2009).

Peptide specificity

An increased breadth of Gag-targeted responses correlates with lowered viral set point in chronic infection (Kiepiela *et al.* 2007; Brumme, Tao *et al.* 2008; Matthews *et al.*

2008). This may be explained by the large copy numbers of Gag that are present early in infection (Briggs *et al.* 2004), making Gag epitopes rapidly available for surface presentation by HLA class I. Studies in the SIV-macaque model support this hypothesis by demonstrating the recognition of Gag epitopes within 2 hours of acute infection, ahead of Pol and Nef epitopes and prior to translation of new HIV proteins (Sacha, Chung, Rakasz *et al.* 2007).

Another explanation for the relationship between Gag-specific responses and reduction in viraemia is that Gag (p24 in particular) is highly conserved (Schneidewind *et al.* 2007), so mutations are more likely to occur at a detriment to viral fitness (Martinez-Picado *et al.* 2006; Brockman *et al.* 2007; Crawford *et al.* 2007). Thus alleles selecting strongly for mutations in Gag are most likely to impose a viral fitness cost, reducing replicative capacity and allowing the maintenance of lowered viral loads despite escape from the selecting $CD8^+$ T cell response (Frater *et al.* 2007; Matthews *et al.* 2008). Epitope processing may also vary between different HIV proteins, accounting for more rapid cell surface presentation of Gag peptides than epitopes from other HIV proteins (Le Gall *et al.* 2007).

Immunodominance hierarchy

In natural HIV infection, the cell mediated immune response is often focused on a limited number of epitopes restricted by a minority of alleles, a phenomenon termed immunodominance (Goulder, Sewell *et al.* 1997; Watkins *et al.* 2008). This epitope preference may be advantageous (Goulder *et al.* 1996; Schneidewind *et al.* 2008), as exemplified in subjects with HLA-B*57, who produce immunodominant Gag responses that exceed the breath and magnitude of responses restricted by all other co-

expressed Class I molecules combined (Altfeld *et al.* 2003; Gillespie *et al.* 2006). However, if the dominant responses are directed at non-Gag epitopes (Ngumbela *et al.* 2008), the resulting suppression of subdominant responses may be detrimental to viraemic control (Frahm *et al.* 2006). Immundominance is determined not only by the HLA genotype of the individual, but also by the availability of epitopes; mutations in dominant epitopes may cause a switch to sub-dominant responses (Goulder, Sewell *et al.* 1997; Barouch *et al.* 2002; Loffredo *et al.* 2008).

Polyfunctionality

The quality of the $CD8^+$ T cell response has further been assessed by staining peptide-stimulated cells for cytokines (e.g. IFN-γ, TNF-α, IL-2), chemokines (e.g. MIP-1β), and degranulation markers (e.g. CD107a) (Betts *et al.* 2006). An inverse correlation between polyfunctional responses and viral load was initially described; $CD8^+$ T cell responses in non-progressors were shown preferentially to produce both IFN-γ and IL-2 compared to progressors (Betts *et al.* 2006), $CD8^+$ T cells specific for the HLA-B*27 KK10 response associated with good control have been shown to be polyfunctional, and favourable immune control in association with HLA-B has been correlated with polyfunctionality (Harari *et al.* 2007). Vaccine studies in the SIV-macaque model have also demonstrated improved disease control in association with polyfunctional responses (Liu *et al.* 2008). However, a subset of non-progressors do not make these HIV-specific $CD8^+$ T cell responses, and more recent prospective studies have identified no association between polyfunctionality and duration of AIDS-free survival, suggesting that persistence of viral antigen may be a cause, rather than an effect, of functional impairment of $CD8^+$ T cell responses (Emu *et al.* 2008; Schellens *et al.* 2008; Streeck *et al.* 2008).

Avidity

Avidity has been defined as 'a composite measure of antigen binding at the cellular level' (Appay *et al.* 2008), reflecting the biological outcome of TCR binding of its cognate MHC-peptide complex. High levels of avidity (or 'antigen sensitivity') correspond to the recognition of MHC-peptide at low densities on the surface of infected cells. These higher avidity interactions have been associated with superior anti-viral activity in the context of disease control by alleles such as HLA-B*27 (Almeida *et al.* 2007). Reduced avidity interactions observed in subjects with chronic infection suggests a possible decline in avidity over the course of disease in response to high viraemia or persistent antigen exposure (Harari *et al.* 2007). However other studies have found no consistent link between TCR avidity and antiviral potency or disease progression (Yang, Sarkis, Trocha *et al.* 2003; Chen *et al.* 2009).

Proliferative capacity

The proliferative capacity of HIV-specific $CD8^+$ T cells has been inversely correlated with viral load (Day *et al.* 2007), demonstrating a relationship between preserved proliferation of effector cell populations and disease control in LTNP subjects (Migueles *et al.* 2002). Proliferation may also be driven by high avidity interactions, with rapid expansion of T cell populations that are highly antigen sensitive (Appay *et al.* 2008). However, this is another instance in which cause and effect of antigen load remain to be elucidated (Streeck *et al.* 2008).

Cytotoxic capacity

$CD8^+$ T cells effect killing of infected cells by the release of effector molecules, including granzyme B and perforin, from lytic granules (Migueles *et al.* 2008; Harari *et al.* 2009). In SIV, the capacity for rapid lysis has been shown to correlate with

improved outcomes following immunization with live attenuated vaccine compared to prime-boost regimens (Rollman *et al.* 2007). In LTNP subjects with HIV, superior elimination of infected cells has also been linked to lytic granule loading and rapid release of cytotoxic molecules (Migueles *et al.* 2008).

1.8 Other immune responses to HIV

1.8.1 B cells and antibody mediated immunity

Viral antigen is presented to B cell receptors (surface immunoglobulin molecules on the surface of naïve B cells) in lymphoid tissue (Janeway *et al.* 2005). The antigen is degraded and epitopes are displayed on the cell surface by MHC Class II molecule. Recognition of this MHC-peptide complex by an antigen-specific T_H cell, combined with an additional B–T_H cell interaction (via CD40-CD40 ligand) stimulates clonal proliferation of B cells, which differentiate either into resting memory B cells or antibody-producing plasma cells (Janeway *et al.* 2005; Moir and Fauci 2009).

A successful B cell response to HIV is thwarted by a combination of factors. Firstly, chronic immune activation leads to dysregulation of B cell responses, such that immature B cells become chronically activated, but these cells make few, low-specificity antibodies (Moir and Fauci 2009). In response to chronicity of infection, the pool of circulating B cells becomes functionally exhausted (Moir and Fauci 2009); they have poor proliferative responses, and express high levels of inhibitory receptors such as PD-1 (Day *et al.* 2006). Secondly, the antibodies that are produced by plasma cells do not have broadly neutralizing capacity, due to the antigenic diversity of

exposed viral peptide and the high degree of Env glycosylation (Mascola *et al.* 1996; Hartley *et al.* 2005; Montefiori *et al.* 2007; Barouch 2008; Lynch *et al.* 2009).

Stable binding of antibody to Envelope spikes may be sufficient to impair infection (Hartley *et al.* 2005). Neutralizing antibodies (NAbs) to HIV-1 gp120 and gp41 have been identified (Mascola *et al.* 2000; Parren *et al.* 2001), including the b12 antibody that binds a structurally conserved domain overlapping the CD4 binding site (Zhou *et al.* 2007). The potential benefit of a broad NAb response is highlighted by macaque studies, in which protection has been elicited by passive administration of antibodies 24h prior to challenge with SHIV (SIV-HIV chimera) (Mascola *et al.* 2000). In these instances, antibodies bind highly conserved regions of the Env proteins and can thus neutralise multiple HIV quasispecies. However, there is no evidence of a NAb response contributing to natural viraemic control in a study of elite controllers, in whom antibody responses were equivalent to those seen in subjects with persistent viraemia (Lambotte *et al.* 2009).

1.8.2 Natural Killer (NK) cells

NK cells are large granular lymphocytes that make up 5-15% of PBMCs, and share similarities in surface markers, profile of cytokine production, and cell function with T cells (Parham 2005). NK cells undergo expansion early in the course of viral infections in response to IFNα and IL-15. This activated NK population produces pro-inflammatory cytokines that drive a strong T_{H1} response, promote rapid $CD8^+$ T cell proliferation, and produce large quantities of perforin and granzyme (Alter and Altfeld 2009).

Several pieces of evidence suggest a role for NK cells in HIV disease control. HLA Class I molecules complexed with viral antigen serve as a ligand for KIR (Killer Immunoglobulin Like Receptors) on NK cells, as well as for TCRs on $CD8^+$ T cells, suggesting that the strong protection associated with some HLA-epitope combinations (e.g. HLA-B*27-KK10) might be accounted for by NK activation as well as by $CD8^+$ T cell responses (Stewart-Jones, di Gleria *et al.* 2005). The interplay between HLA and KIR genotype is also exemplified by the HLA-Bw4/KIR3DS1 haplotype; individuals with combinations of these genes progress more slowly to AIDS than individuals with only one of these alleles (Carrington *et al.* 2008). Furthermore, the link between gene expression at the HLA-Cw locus and lowered HIV viraemia in a genome-wide study (Fellay *et al.* 2007) has pointed to a possible mechanism via increased recruitment of NK cells, as HLA-Cw is the main ligand for the KIR2D receptors (Alter and Altfeld 2009).

1.8.3 $CD4^+$ T cells

The role of $CD4^+$ T cells in control of HIV replication is much less well characterised than that of $CD8^+$ cells. However, $CD4^+$ T cells are known to be crucial in control of viral infections, in stimulating and maintaining both $CD8^+$ and B cell populations, and potentially in contributing a direct cytolytic effector function (Sacha *et al.* 2009). Preliminary evidence for the role of these cells comes from observations that the preservation of HIV-specific $CD4^+$ T cell populations and sustained IL-2 production are correlated with improved immune control (Rosenberg *et al.* 1997), and that certain MHC Class II molecules in Rhesus macaques are associated with enhanced protection (Giraldo-Vela *et al.* 2008). More recently, $CD4^+$ T cell responses directed towards Gag and Nef specific macrophages have been demonstrated in CD8-depleted

macaques, suggesting their primary role may be in suppressing infection in the macrophage population (Sacha *et al.* 2009).

1.8.4 Host restriction factors

A variety of intra-cellular proteins that provide innate immune defence against viral infections are termed restriction factors (Towers 2007). TRIM5α, a protein identified in 2004 as part of the **Tri**partite **M**otif protein family (Stremlau *et al.* 2004), is one such factor that blocks retroviral infection of mammalian cells (Towers 2007). The precise mechanism of action of TRIM5α remains unknown, but it interacts with HIV-1 hexameric capsids, potentially disrupting the structural arrangement required for uncoating or trafficking, or tagging them for ubiquitin-dependent proteasomal degradation (Towers 2007). In order to evade this anti-viral mechanism, the virus recruits host Cyp-A protein which prevents the interaction between TRIM5α and Gag capsid.

Host APOBEC (**Apo**lipoprotein **B** mRNA **E**diting **C**atalytic polypeptide) proteins are a family of cellular cytidine deaminases that provide anti-retroviral immunity by causing extensive cytidine to uridine substitutions (through G to A 'hypermutation') or by restricting replication through deaminase-independent pathways (Sheehy *et al.* 2002; Goila-Gaur and Strebel 2008). APOBEC-3G operates as an anti-HIV restriction factor at the reverse transcription stage, introducing multiple errors into the newly synthesized DNA strand (Mangeat *et al.* 2003). HIV-1 Vif counters the activity of APOBEC-3G, removing it from host cells and preventing its incorporation into new virions (Goila-Gaur and Strebel 2008).

1.9 HIV vaccine strategies

1.9.1 Current options for HIV vaccines

Broadly speaking, vaccine strategies can be divided into the following four categories (Berkley and Koff 2007):

i. ***Vaccines from HIV proteins:*** despite promising data from animal studies (Berman *et al.* 1996), subunit vaccines using recombinant envelope proteins (rgp120) failed to elicit protective immunity in humans in Phase III trials (Flynn *et al.* 2005).

ii. ***Vaccines from whole HIV:*** these rely either on immunization with whole killed virus, or use live attenuated strains. Killed virus is insufficiently immunogenic, while the use of live viruses – although informative in animal models (Daniel *et al.* 1992; Wyand *et al.* 1996) - has been precluded in humans due to safety concerns.

iii. ***Vaccines from HIV genes:*** vaccines can use naked DNA, or DNA transmitted by a viral or bacterial vector. Some animal data suggest superior responses to $CD8^+$ T cell vaccines following DNA priming (Casimiro *et al.* 2005).

iv. ***Combination vaccines:*** any of the above strategies combined.

1.9.2 Goals of a successful vaccine

The ultimate aim of an HIV vaccine is to elicit durable sterilising immunity across multiple subtypes of infection (Barouch 2008). However, given the significant obstacles to achieving this aim (Table 1.2), the current focus is on producing a vaccine that reduces viraemia, thus delaying disease in infected individuals, and

reducing onward transmission (Johnston and Fauci 2007; Barouch 2008; Walker and Burton 2008; Watkins *et al.* 2008).

Watkins *et al.* propose a threshold of ≥1.5 log reduction in viraemia as a criterion for vaccine efficacy (Watkins *et al.* 2008), based on a substantial reduction in risk of HIV-1 transmission by individuals with viral loads suppressed from a median viral load of 30,000 RNA copies/ml to <1500 copies/ml (Quinn *et al.* 2000; Gray *et al.* 2001). In order to achieve these aims, a successful HIV vaccine is likely to need to induce both broad cell mediated immunity and a broad NAb response, to preserve memory $CD4^+$ lymphocytes, and possibly to recruit innate immune responses, with a view to boosting local mucosal immunity at the site of infection and targeting subsequent viral replication (Johnston and Fauci 2007; Montefiori *et al.* 2007; Steinbrook 2007; Barouch 2008; Alter and Altfeld 2009).

Obstacle to vaccine design	Contributing factors
Marked antigenic diversity of virus	Differences in sequence between and within clades (Gaschen *et al.* 2002; Allen *et al.* 2005) and diversity of quasispecies within an individual host alter epitope availability and allow immune escape. Variability compounded by recombination (Korber *et al.* 2001) and superinfection. Rapid mutation rate allows escape from humoral (Montefiori *et al.* 2007; Moir and Fauci 2009) and adaptive (Goulder, Phillips *et al.* 1997; Allen *et al.* 2004; Draenert *et al.* 2004; Feeney *et al.* 2004) immune responses.
Cell-to-cell transmission and latent reservoirs	Virus can be transmitted between cells without existing in a cell-free form (Sattentau 2008), allowing it to escape antibody-mediated neutralization. Early establishment of reservoirs of virus allow HIV to evade immune detection and HAART (Montefiori *et al.* 2007).
Uncertain correlates of immune protection	Magnitude and quality of $CD8^+$ T cell responses fail to correlate consistently with disease outcome (Addo *et al.* 2003).
Failure of natural disease or vaccines to induce broadly neutralizing antibodies	Induced antibodies neutralize only a minority of circulating strains of virus; exposed regions of Env are diverse, and extensive envelope glycosylation shields the virus from neutralization (Montefiori *et al.* 2007; Moir and Fauci 2009). Env trimers are unstable and difficult to synthesise in recombinant form. Functional exhaustion of B cells occurs in association with HIV progression (Moir and Fauci 2009).
Attenuated viruses unsafe for use in humans	Eventual progression to disease occurs even in the context of infection with an attenuated virus (Learmont *et al.* 1999; Whitney and Ruprecht 2004).
Damage/dysfunction of immune system	HIV infects and destroys cells that play crucial roles in the immune system. HIV mediates MHC down-regulation.
Lack of a small animal model	SIV-macaque is the only animal model, but MHC types and antigen specificity differ between this model and human infection (Watkins *et al.* 2008). Vaccine trials in animals are expensive, and restricted to small numbers.
Insufficient pharmaceutical investment	Lack of initiative for financial investment (Berkley and Koff 2007).

Table 1.2: Obstacles to design of a successful HIV Vaccine.

1.9.3 CD8$^+$ T cell vaccines

Vaccines in the SIV-macaque model

CTL vaccines in the SIV-macaque model have sought attenuation of, or protection against, infection with clonal (e.g. SIV_{mac239}) or swarm (e.g. SIV_{mac251}) viruses with variable success (Watkins *et al.* 2008). The most significant protection from infection has been achieved using vaccination with live attenuated SIV strains (Koff *et al.* 2006); robust protection can be achieved by vaccination with homologous strains (Daniel *et al.* 1992).

Other SIV studies have focused on the use of vectors, deemed safer for subsequent transfer into the human host. Animals vaccinated with a vaccinia virus vector expressing SIV Gag, Pol and Env were able to clear viraemia following mucosal challenge with SIV_{mac251}, but not following intra-venous infection (Benson *et al.* 1998).

Studies in Burmese macaques have yielded data that are promising for the ultimate success of a CD8$^+$ T cell vaccine: immunisation with a DNA prime / Gag-expressing Sendai virus* boost regimen produced viraemic control in the majority of animals, followed by the emergence of viral strains with reduced viral fitness due to the selection of CD8$^+$ T cell escape mutations (Matano *et al.* 2004). Subsequently, vaccinees were infected with viruses bearing mutations in immunodominant Gag epitopes, and viraemia was not controlled (Kawada *et al.* 2008), highlighting the crucial role of these CD8$^+$ T cell responses in containing disease. Together, these data

* Murine parainfluenza virus type 1

underscore the potential for successful vaccine-mediated immune control derived from Gag-specific $CD8^+$ T cell responses.

Protection has also been demonstrated in macaques vaccinated with a DNA-prime / Ad5 (Adenovirus 5) boost regimen incorporating viral Gag, Tat, Rev and Nef: potential for reduction in peak viraemia and in viral set-point at 12 months was demonstrated, but was significant only in animals with the protective class I allele Mamu-A*01 (Wilson *et al.* 2006). More recently, a trial in Mamu-A*01-negative animals showed disease attenuation following a heterologous Ad26 prime / Ad5 boost regimen delivering Gag antigen, prior to challenge with SIV_{mac251} (Liu *et al.* 2009).

Vaccines in HIV infection

Vaccine constructs are listed in Table 1.3.

Vaccine (Manufacturer)	Vector	Antigens
Trivalent T cell Vaccine (Merck)	Replication incompetent Ad5	HIV-1 subtype B Gag, Pol, Nef
ALVAC-HIV (Aventis Pasteur)	Recombinant canarypox	CRF01_AE HIV-1 gp120 + transmembrane-anchoring portion of subtype B gp41 + HIV-1 subtype B Gag and protease (Nitayaphan *et al.* 2004)
AIDSVAX (VaxGen)	Nil (protein only)	HIV-1 subtype B gp120 + CRF01_AE envelope (Nitayaphan *et al.* 2004)

Table 1.3: HIV-1 Vaccine strategies recently trialled in humans

Recent phase I (Priddy *et al.* 2008) and phase II (Buchbinder *et al.* 2008; McElrath *et al.* 2008) clinical trials in humans have been undertaken using the Merck Ad5 T cell vaccine in homologous prime/boost regimens. In safety and immunogenicity trials, 72% of participants receiving higher doses of virus ($3x10^{10}$ viral particles), had

ELISpot responses directed against ≥2 peptide pools, and 44% had responses to all three proteins, with responses sustained to ≥78 weeks post-vaccine (Priddy *et al.* 2008). However, responses were diminished by prior Ad5 immunity; in particular, inverse associations were found between ELISpot responses to Gag peptides and baseline Ad5 antibody titre (Priddy *et al.* 2008).

In 2007, the multicentre phase II STEP and Phambili trials were discontinued early due to futility at interim analysis; 3% of participants in both vaccine and placebo groups became infected, with no difference in plasma RNA viral between groups (Buchbinder *et al.* 2008). The hazard ratio for infection was greater in uncircumcised men, and in men with prior Ad5 immunity (Buchbinder *et al.* 2008). Furthermore, HIV-1 infection in men with prior Ad5 immunity was higher in vaccinees than in placebo recipients over 40-100 weeks post-immunisation (Barouch 2008; McElrath *et al.* 2008). The explanation for this outcome remains uncertain, but may relate to the expansion of Ad5-specific $CD4^+$ T cell populations in subjects with prior Ad5 immunity, that act as a reservoir of cells for HIV infection (Barouch 2008). An alternative hypothesis is that Ad5-specific antibodies opsonise rAd5 vectors (Barouch 2008).

1.9.4 Antibody vaccines

Most anti-viral vaccines depend on the induction of NAbs, and a successful HIV vaccine is likely to need to include this approach (Montefiori *et al.* 2007; Watkins *et al.* 2008). Studies in the SIV-macaque model have provided proof of principle that sterilising immunity can be induced by passive transfer of broadly NAbs (Mascola *et al.* 2000; Mascola 2002). Human phase I/II trials in Bangkok demonstrated safety and

immunogenicity of a combined vaccine protocol incorporating ALVAC-HIV and AIDSVAX (Nitayaphan *et al.* 2004). The majority of HIV-negative vaccine recipients developed neutralizing antibodies to gp120, and nearly 50% also developed antibodies to p24. In addition, 24% of vaccinees developed $CD8^+$ T cell responses to ALVAC antigens (Nitayaphan *et al.* 2004). However, in subsequent phase III trials, the vaccine failed to prevent infection (Pitisuttithum *et al.* 2004), corroborating *in vitro* findings that induced antibodies lack broadly neutralizing capacity (Mascola *et al.* 1996; Montefiori *et al.* 2007). Results of a further large trial in Thailand combining ALVAC and AIDSVAX are due late in 2009.

Reasons for difficulties in eliciting protection using antibody-based vaccines are summarised in Table 1.2.

1.10 Aims of these studies

In the setting of the failed Merck T cell vaccine trial, a return to basic science approaches has been advocated to advance an understanding of the immune correlates of protection (Barouch 2008; Plotkin 2009). These studies therefore set out to improve characterisation of $CD8^+$ T cell responses to HIV, in particular focusing on responses that are associated with viraemic control. Given the strong HLA-selection pressure for $CD8^+$ T cell escape mutations, consideration has also been given to the evolutionary impact of viral escape in determining the future direction of the epidemic. The particular focus of this work is on subtype C infection, as this accounts for over two thirds of the total global burden of HIV, and affects regions of the developing world least able to deliver effective HAART.

Specifically, the following aims were pursued:

i. To define the role of reverting $CD8^+$ T cell escape mutations (i.e. mutations that are detrimental to viral fitness) in control of HIV viraemia.

ii. To determine the impact of $CD8^+$ T cell escape mutations on HIV evolution, both within and among clades.

iii. To study $CD8^+$ T cell responses that are associated with a lowered viral set point, focusing on viral suppression in association with HLA-B*35.

iv. To optimise *in vitro* assays in order to quantify the magnitude of viral suppression effected by different $CD8^+$ T cell responses.

CHAPTER 2: Materials and Methods

Methods relevant to multiple chapters are detailed here; further optimisation and development of specific methods are discussed in the methods section of individual chapters. Constituents of media and details of on-line tools are included in Supplementary Methods (section 2.14).

2.1 Patient cohorts

2.1.1 Durban, South Africa

Treatment-naïve adults with C-clade HIV infection were recruited from Durban, South Africa, following voluntary counselling and testing in either antenatal or medical outpatient clinics at Cato Manor Clinic, Sinikithemba Clinic, St. Mary's Hospital and King Edward VIII Hospital. Subjects were almost exclusively of Zulu or Xhosa origin, and were infected by heterosexual transmission. Informed consent was obtained from all individuals. Blood samples were processed at the HIV Pathogenesis Programme (HPP) laboratory, Durban. Viral load was measured from plasma using the Roche Amplicor v.1.5 assay. CD4 count was measured by flow cytometry on BD FACS Calibur (Table 2.1). The study was approved by the University of KwaZulu-Natal Review Board, the Oxford Research Ethics Committee and the Massachusetts General Hospital Review Board.

Study cohort	n	Viral load median (IQR) (RNA copies/ml plasma)	CD4 count median, (IQR) (cells/mm^3)
Durban, S. Africa	1213	38200 (7209-155000)	374 (233-519)
Gaborone, Botswana	365	15300 (3245-66350)	342 (226-477)
Thames Valley, UK	264	15000 (3662-65405)	415 (305-542)

Table 2.1: Baseline viral load and CD4 count, median and interquartile range (IQR), for HLA-typed subjects from cohorts in South Africa, Botswana (antenatal cohort only) and UK.

2.1.2 Gaborone, Botswana

Pregnant women infected with HIV were recruited from ante-natal clinics in Gaborone following voluntary counselling and testing. CD4 and viral load data are shown in table 2.1. Ethics approval was given by the Office of Human Research Administration (Harvard School of Public Health), and the Health Research Development Committee (Botswana Ministry of Health). Viral loads were performed by Roche Amplicor v.1.5. An extended Botswana cohort including an additional 93 adult subjects with chronic infection, previously recruited by V. Novitsky, were included for analysis of HLA types only.

2.1.3 Thames Valley, UK

HAART-naïve adult subjects attending HIV outpatient clinics in the Thames Valley region were recruited. Participating centres were Oxford, Northampton, Slough, Reading, High Wycombe, and Luton. The majority of these subjects were chronically HIV-infected. Ethnic origin is shown in Fig. 2.1; B-clade subjects were predominantly infected by MSM, and C-clade subjects by heterosexual transmission. All subjects provided informed consent for participation. For each subject, 30-50ml of whole blood was collected in EDTA at enrollment, and, when available, at follow-up time points of 3-6 month intervals. Viral load and CD4 count data were provided by

the recruiting centres (Table 2.1). The study was approved by local Research and Development Committees at each recruitment site, and by the Oxford Research Ethics Committee.

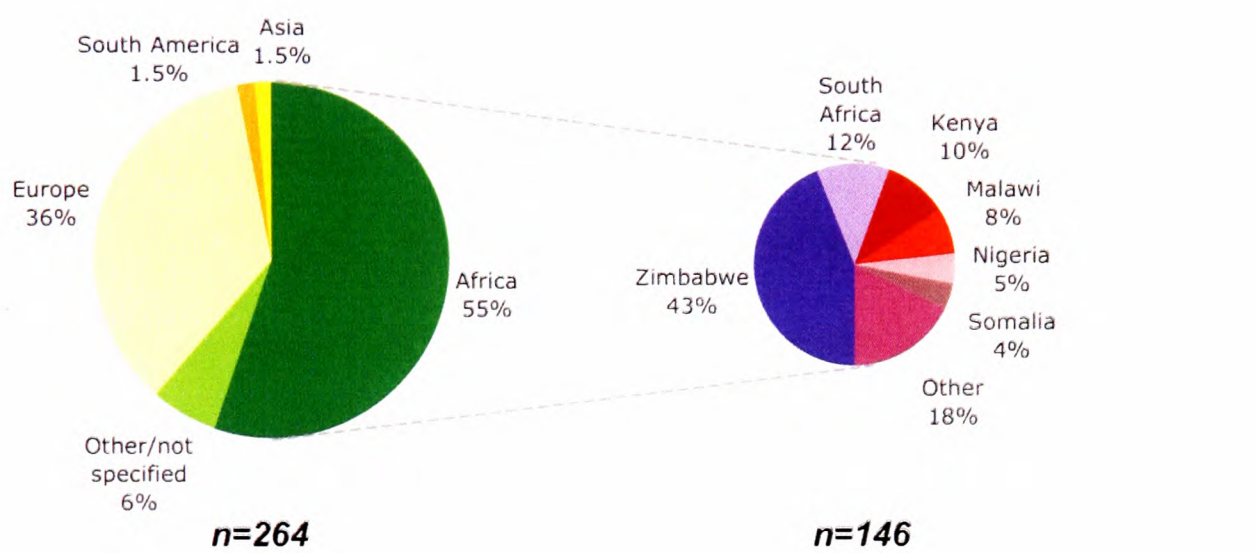


Figure 2.1: Ethnic origin of 264 adult subjects recruited in the Thames Valley Cohort.

2.1.4 Additional sources of sequence data

Japanese subjects with acute and chronic HIV infection were recruited in Kumamoto, Japan, and sequence data were provided by M. Takiguchi. Thai sequences were provided by T. Allen and U. Hempel, from chronically infected subjects enrolled in The King Chulalongkorn Memorial Hospital and Thai Red Cross AIDS Research Centre in Bangkok, Thailand. The local Institutional Research Boards approved these cohorts, and all subjects provided written informed consent.

Additional B-clade sequences were sourced from sequences uploaded to Genbank, available at Los Alamos HIV database, www.hiv.lanl.gov). Clade consensus sequences used in chapter 4 were also sourced from Los Alamos HIV database.

2.2 Processing of fresh whole blood

2.2.1 Harvesting PBMCs from whole blood

Whole blood was collected in EDTA. Samples were spun at 500g for 10 minutes to separate plasma; this was frozen at -80°C in 1.5 ml aliquots. The remaining sample was transferred to a 50 ml Falcon tube, diluted with phosphate-buffered saline (PBS, Invitrogen, UK) at a ratio of 1:1 and layered onto 15 ml of Lymphoprep sucrose gradient (Axis-Shield, Norway). The gradient was spun at 500g for 30 minutes with slow brakes on the centrifuge. The PBMC layer was collected using a Pasteur pipette. Cells were washed three times in PBS, and re-suspended in R10.

2.2.2 Freezing and thawing cells

Cells were prepared for freezing by spinning at 500g for 5 minutes and pouring off the supernatant. The cells were resuspended in freezing media (see section 2.14), transferred to 2 ml cryovials and frozen slowly to -80°C before transferring to -180°C for long term storage. Cells were thawed by incubating the cryovial in a waterbath at 37°C.

2.2.3 DNA Extraction

Genomic DNA extraction was performed from fresh whole blood, using Puregene reagents (Qiagen, UK). To extract DNA from 3 ml whole blood, 9 ml of red cell lysis buffer was added, and mixed gently. The mixture was incubated at room temperature for 10 minutes, then spun at 500g for 10 minutes. The supernatant was discarded and the pellet resuspended in 3 ml cell lysis buffer. After a minimum of 48 hours at room

temperature, 1 ml of protein precipitation solution was added and the mixture was vortexed for 20 seconds, then spun at 500g for 10 minutes. The supernatant was decanted into 3 ml of isopropanol, and inverted 50 times until white threads became visible. The suspension was spun for 3 minutes at 500g, then the supernatant carefully discarded. The DNA was washed with 3 ml 70% ethanol, spun at 500g for 1 minute, and then inverted and left to dry. 250 µl DNAse free H_2O was added. Samples were stored at -20°C.

2.2.4 RNA Extraction

Viral RNA was extracted from plasma using RNA extraction minikit (Qiagen UK) in accordance with the manufacturer's instructions. The protocol was amended to add 10 µl DNAse per sample (RNAse free DNAse set, Qiagen UK) prior to washing the column with AW1 buffer, and to add 1 µl RNAse inhibitor (RNAsin Plus, Promega) per column prior to elution of RNA. Following RNA extraction, samples were stored at -20°C.

2.3 HLA Typing

Four-digit HLA typing of the Class I locus was undertaken using genomic DNA by sequence-based typing at the ASHI* accredited HLA typing laboratory, University of Oklahoma Health Sciences Centre, USA. Locus specific PCR of exons 2 and 3 of Class I were amplified and DNA sequenced. Ambiguity resolution was undertaken according to the ASHI committee recommendations (Cano *et al.* 2007).

---

* American Society for Histocompatibility and Immunogenetics

2.4 Interferon-gamma (IFN-γ) ELISpot assays

2.4.1 General methods for ELISpot assay

ELISpot assays were performed using sterile 96 well plates with a Polyvinylidene Fluoride membrane (Millipore, USA). The plate was pre-coated with anti-human IFN-γ monoclonal antibody (Mabtech, Sweden) and incubated at 4°C for 24 hours. Prior to use, the plate was washed 6 times with blocking buffer (5% FCS in PBS), and 50 μl R10 was added to each well. Peptide was added at a final concentration of 2 μg/mL per well. For each plate, two positive wells (PHA at 250 μg/mL (Phytohaemagglutinin, Roche)) and four negative wells (containing cells but no peptide/PHA) were included. PBMCs were added at $5.0x10^4$/well.

Following 12-14 hours incubation at 37°C, the wells were washed six times with PBS. The plate was processed by addition of biotinylated IFN-γ antibody at 1 μg/μl (Mabtech, Sweden) and incubated for 90 minutes at room temperature in the dark, followed by re-washing and addition of 100ul streptavidin-alkaline phosphatase (ALP) conjugate (Mabtech, Sweden) and incubation for a further 45 minutes. Development was performed by addition of substrate colour solution (Bio-Rad Laboratories, UK). The plate was read using AID ELISpot plate reader (ELISpot v.4.0, Autoimmune Diagnotiska, Germany). Positive responses were calculated by subtracting the background (mean of the values from the four negative control wells). A significant positive response was then defined as anything above the background + (3 x standard deviation) of the control wells, and the final results were expressed as number of spot forming cells (SFCs) per million PBMC.

2.4.2 Screening for HIV-specific responses by megamatrix

To screen subjects for $CD8^+$ T cell responses to any HIV proteins, a panel of 410 overlapping 18-mer peptides (OLPs), spanning the HIV proteome, were synthesised based on the 2001 B and C-clade consensus sequences (Kiepiela *et al.* 2007) (provided by Massachusetts General Hospital Peptide Core Facility). These were used in a matrix system of 11-12 peptides/pool in 72 pools, with each peptide represented in two different pools. The pools were pre-mixed and stored at 4°C.

To confirm $CD8^+$ T cell responses to individual OLPs (suggested by positive responses to both pools containing the individual peptide), a confirmation was undertaken using individual OLPs in place of peptide pools. In addition, optimal peptides of the relevant HLA restriction were added to the megamatrix or confirmation plate at a final concentration of 2 μg/ml. For the purposes of data analysis, subjects making ELISpot responses to adjacent OLPs, only one of these was counted as a positive response (the bigger magnitude response).

2.4.3 ELISpot confirmation of optimal epitopes

Sites of putative epitopes were initially determined based on sites of HLA-associated HIV sequence polymorphism, at sites where no epitope of the relevant restriction had been previously defined. The optimal epitope was sought on the basis of previously published HLA binding motifs defining characteristic amino acids present at position 2 and the C-terminus (Marsh *et al.* 2000; Honeyborne *et al.* 2006). For each putative optimal, five peptides were synthesized: the optimal epitope plus four truncations (adding and subtracting one amino acid residue from the C- and N-terminus of the peptide). Serial dilutions of each of these peptide variants were performed in R10,

starting at 10^{-4}M and proceeding with tenfold dilutions to 10^{-10}M. Cells were added at $1x10^5$/well, and the plate was incubated and developed as described above.

2.5 Amplification of HIV genes by Polymerase Chain Reaction (PCR)

2.5.1 Amplification from proviral DNA

Genomic DNA was extracted from PBMC and amplified by nested PCR to obtain population sequences. Regions of the virus amplified were the three most immunogenic proteins, Gag, Pol and Nef. A mastermix was generated for each PCR reaction, using 20 μmol unmodified oligonucleotide primers (MWG Biotech, Germany), as below (Table 2.2 and table 2.4). Primer sequences were designed by A. Leslie and K. Pfafferott, based on division of the HIV genome into eight approximately equal fragments of around 1200 base pairs each (Tables 2.4, 2.5).

2.5.2 Amplification from RNA

Reverse-transcription of RNA to make cDNA was undertaken using Superscript Reverse Transcriptase III (Invitrogen, UK) as a one-step reaction combined with outer PCR (RT-PCR) (Table 2.3 and table 2.4).

2.5.3 Visualisation of PCR product

Following amplification by PCR, 5 μl of the PCR product was loaded onto a 4% agarose gel (Sigma, UK) in TBE buffer (Tris-borate-EDTA, Promega, UK). Ethidium bromide was added at 5 μl per 400 ml gel. To determine the size of the amplified

fragment, samples were run in parallel with Hyperladder I (Bioline, UK). After running at 240V for 40 minutes, bands were visualized under UV light.

Reagent	Volume (ul)
H_2O	37
Buffer (NH_4)*	5
$MgCl_2$ (50mM)*	3
5'-primer	1
3'-primer	1
dNTP	0.5
Bio*Taq* (5u/ml)*	0.5
DNA	2
Total	**50**

Table 2.2: Components of PCR reaction mixture. * Reagents from Bioline, UK

Reagent	Volume (ul)
H_2O	7
Reaction mix (2x) *	12.5
5'- outer primer	0.5
3'- outer primer	0.5
Superscript enzyme *	0.5
RNA	5
Total	**26**

Table 2.3 Components of RT-PCR reaction mixture. * Reagents from Invitrogen, UK

2.5.4 Caveats of population sequencing

Taq DNA polymerase is a heat stable enzyme derived from the thermophilic organism *Thermus aquaticus*. The enzyme lacks proof-reading function (3'-5' exonuclease activity), leading to nucleotide substitution errors estimated at $0.2\text{-}2x10^{-4}$ misincorporations per base pair per PCR cycle, with an overall error rate estimated at 10 per 10,000 base pairs (Eckert and Kunkel 1991; Smith, McAllister *et al.* 1997; Bracho *et al.* 1998). Other PCR errors can occur as a result of frameshift mutation (estimated at $2.0x10^{-5}$ per base pair) and strand switching (whereby the polymerase amplifies nucleotide sequences from more than one taxon causing *in vitro* recombination). In addition, a bias in amplification can arise by 'allelic preference', as

a result of primers binding preferentially to a small number of quasispecies (Eckert and Kunkel 1991).

For the purposes of the studies presented here, which focus largely on sequences derived from large study populations, random errors at these low frequencies were deemed unlikely to interfere with statistical analysis (Eckert and Kunkel 1991).

Fragment	Clade	Region of genome	Primer	Sequence
1	B	Gag p17 + p24	Frag1_5'OP	AAATCTCTAGCAGTGGCGCC
			Frag1_3'OP	TGTTGGCTCTGGTCTGCTCT
			Frag1_5'IP	ACTAGCGGAGGCTAGAA
			Frag1_3'IP	GCCACAATTGAAACACTT
1	C	Gag p17 + p24	Frag1_5'OP	CTAGCAGTGGCGCCCGAACA
			Frag1_3'OP	GCAGTCTTTCATTTGGTGTCCTTC
			Frag1_5'IP	TCTCTCGACGCAGGACTC
			Frag1_3'IP	TTTCCACATTTCCAACAGCC
2	C	Gag p15 + Protease	Frag2_5'OP	ATAGTAAGAATGTATAGCCCTGT
			Frag2_3'OP	ATGGAGTTCATACCCCATCC
			Frag2_5'IP	GAGTGGGAGGACCTAGC
			Frag2_3'IP	TACAAGTCATCCATATATTGATAG
3	C	Protease + Reverse Transcriptase	Frag3_5'OP	GAAGGACAGTACTAAGTGGAG
			Frag3_3'OP	CTGGCTACATGGACTGCTAC
			Frag3_5'IP	ACTGCATTCACCATACCT
			Frag3_3'IP	TTATGTGCTGGTACCCATGA
4	C	Integrase	Frag4_5'OP	CATGGTGGACAGACTATTGGC
			Frag4_3'OP	CCTGCCATCTGTTTTCCATA
			Frag4_5'IP	CCTCCCCTAGTAAAATTATGG
			Frag4_3'IP	GCTGTCTCTGTAATAAACCCG
8	C	Nef	Frag8_5'OP	GATCCATTCGATTAGTGAACGGA
			Frag8_3'OP	CTTATATGCAGCATCTGAGG
			Frag8_5'IP	TTCAGCTACCACCGATTGAGA
			Frag8_3'IP	TGAGGGTTGGCCACTCC

Table 2.4: PCR Primers for amplification of HIV *gag*, *pol* and *nef* genes

Fragment	Clade	Region of genome	Primer	Sequence
1	B	Gag p17 + p24	Frag1_5'IP	ACTAGCGGAGGCTAGAA
			Frag1_2	CCTGGAGGTTCTGCAC
			Frag1_3	GACACCAAGGAAGCTTTAG
			Frag1_4	ATCCCAGGATTATCCATC
			Frag1_5	CCAAGGGAAGTGACAT
			Frag1_3'IP	GCCACAATTGAAACACTT
1	C	Gag p17 + p24	Frag1_5'IP	ACTCGGCTTGCTGAAGTGC
			Frag1_2	CTGCACTATAGGATAATTTTGAC
			Frag1_3	GACACCAAGGAAGCCTTAG
			Frag1_4	CTCCCACTGGAACAGGTG
			Frag1_5	GGAACAAATAGCATGGATGAC
			Frag1_3'IP	CAATTTCTGGCTATGTGCCC
2	C	Gag p15 + Protease	Frag2_5'IP	GAGTGGGAGGACCTAGC
			Frag2_2	ATCATCTGCTCCTGTGTCTA
			Frag2_3	AGGCCAGGGAATTTCCTTCA
			Frag2_4	CATTGTTTAACCTTTGGGCCATC
			Frag2_5	CCCATTGAAACTGTACCAG
			Frag2_3'IP	TACAAGTCATCCATATATTGATAG
3	C	Protease + Reverse Transcriptase	Frag3_5'IP	ACTGCATTCACCATACCT
			Frag3_2	TGGGTCATAATATACTCCATGTA
			Frag3_3	CAGACATAGTACCACTAACTGAA
			Frag3_4	CATAGAAAGTTTCTGCTCCT
			Frag3_5	CTCCCCTAGTAAAATTATGGTA
			Frag3_3'IP	TTATGTGCTGGTACCCATGA
4	C	Integrase	Frag4_5'IP	CCTCCCCTAGTAAAATTATGG
			Frag4_2	TCACTAGCCATTGCTCTCC
			Frag4_3	ATTGGAGGAAATGAACAAGTAGA
			Frag4_4	GCCCACCAACAGGCTGC
			Frag4_5	GAAACAGGACAAGAAACAGC
			Frag4_3'IP	GCTGTCTCTGTAATAAACCCG
8	C	Nef	Frag8_5'IP	TTCAGCTACCACCGATTGAGA
			Frag8_2	ACCTGAGGTCTGACTGGAAAG
			Frag8_3	AGATAAACATGGGGCACTTAC
			Frag8_4	AGCAGTCTTTGTAATACTCCG
			Frag8_5	ACCTCAGGTGCCTTTAAGAC
			Frag8_3'IP	TGAGGGTTGGCCACTCC

Table 2.5: Sequencing primers for HIV *gag*, *pol* and *nef* genes

Step number	Temperature (°C)	Time
1	94	2 minutes
2	94	15 seconds
3	Temperature 1*	30 seconds
4	72	1 minute
5	*go to step 2, 19 times*	
6	94	15 seconds
7	Temperature 2*	30 seconds
8	72	1 minute
9	*go to step 6, 19 times*	
10	72	7 minutes
11	4	for ever

Table 2.6: PCR cycle. * Temperatures 1 and 2 defined in table 2.8.

Step number	Temperature (°C)	Time
1	50	30 minutes
2	94	2 minutes
3	94	15 seconds
4	Temperature 3*	30 seconds
5	72	1 minute
6	*go to step 3, 34 times*	
7	72	7 minutes
8	4	for ever

Table 2.7: RT-PCR cycle. * Temperature 3 defined as 5°C less than lowest melting temperature of 2 primers (range 55-60°C).

PCR reaction	Temperature 1 (°C)	Temperature 2 (°C)
Frag1 OP	64	61
Frag1 IP	60	57
Frag2 OP	59	56
Frag2 IP	59	56
Frag3 OP	62	59
Frag3 IP	58	55
Frag8 OP	61	58
Frag8 IP	61	58

Table 2.8: Annealing temperatures for PCR.

2.5.5 Quality control of PCR

At lease one blank (negative) control was run for each PCR reaction; any band in this well suggested PCR contamination and the entire batch of samples was repeated. Following attainment of sequence data, taxa were assembled into a phylogenetic tree to ensure each sequence was distinct.

2.6 Purification and sequencing of amplified HIV genes

2.6.1 Purification of PCR product

For subjects from the Durban cohort, the PCR product was purified by Polyethylene Glycol (PEG) precipitation using a QIAquick PCR purification kit (Qiagen, UK) according to manufacturer's instructions. For subjects from Botswana and the Thames Valley Cohort, the PCR product was purified using ExoSAP-IT, containing Exonuclease I and Shrimp Alkaline Phosphatase, (GE Healthcare Life Sciences, UK) in accordance with the manufacturer's instructions.

2.6.2 Sequencing of PCR product

Sequencing of the Durban cohort was undertaken using the Big Dye ready reaction terminator mix (V3) (Applied Biosystems, UK), and sequence data were analysed using Fugu (University of Oxford). For other cohorts, sequencing was undertaken by Macrogen Inc., South Korea. Sequence data were analysed using Sequencher v.4.8 (Gene Codes Corporation). Nucleotides for each gene were aligned manually in Se-Al v.2.0a11 (A. Rambaut, Department of Zoology, University of Oxford).

Published Sequences have been uploaded to GenBank under the following accession numbers: (www.ncbi.nlm.nih.gov/Genbank)

Durban: Gag: FJ198407-FJ199088
Pol: FJ199532-FJ199992
Nef: FJ199089-FJ199531

Gaborone: Gag p17+p24: FJ497801-FJ497950
Gag p15: FJ497951-FJ498243
Pol: FJ498244-FJ498543
Nef: FJ498544-FJ498779

2.7 Bioinformatics and Phylogenetic Analysis

2.7.1 Statistical Tests

Fisher's Exact Test was used to seek correlations in categorical data presented as a 2x2 contingency table. The test was performed using software available at www.langsrud.com/fisher.htm.

Correcting statistical significance for multiple comparisons was undertaken either using a Bonferroni correction, or a *q*-value approach. A *q*-value corrects for multiple comparisons by calculation of a false discovery rate (FDR) for each *p*-value (Storey 2003; Storey and Tibshirani 2003). This method has been advocated in genome-wide studies, in which a Bonferroni approach to correcting for multiple comparisons is likely to be over-cautious, and thus to miss many true positive associations (Storey and Tibshirani 2003).

Other statistical tests are described in the methods sections of individual chapters where pertinent, and were performed using Graphpad Prism v.5.0 or Microsoft Excel v.11.2, unless otherwise stated.

2.7.2 Identification of HLA class I associations with HIV polymorphisms

Fisher's Exact Test was used to detect associations between HIV amino acid polymorphisms and expression of particular HLA Class 1 alleles, corrected for multiple comparisons. However, this approach is potentially subject to false positive associations on account of a founder effect (Moore *et al.* 2002). Bhattacharya *et al.* proposed a method to correct for this effect (Bhattacharya *et al.* 2007), in which a phylogenetic correction is undertaken. This approach was adopted for analysis of sequence data from Durban and Botswana, using an algorithm developed at Microsoft Research. Further details are included in the methods section of Chapter 3.

2.7.3 Calculation of Shannon Entropy

Shannon entropy (*H*) is a measure of variability or conservation (Shannon 1948), accounting for both the frequency and diversity of amino acid variation at a given residue. Entropy at a given residue is maximal when all amino acids occur at equal probability, and is zero when there is no variation (Schmitt and Herzel 1997). The maximum Shannon entropy is dependent on the possible number of discrete variables; for proteins this is 20 variables. Shannon entropy was computed using the on-line tools available at Los Alamos databases (www.hiv.lanl.gov).

2.7.4 Phylogenetic analysis

All phylogenetic trees were derived from nucleotide sequences aligned in Se-Al. Neighbour joining (NJ) trees were computed using PAUP (Sinauer Associates, MA, USA) using the HKY85 substitution model (**H**asegawa-**K**ishino-**Y**ano, 19**85**, (Hasegawa *et al.* 1985)) and estimating a gamma parameter (Γ). The HKY85 model

accounts for non-uniform base frequencies and for difference between transition and transversion rates (Fig. 2.2).

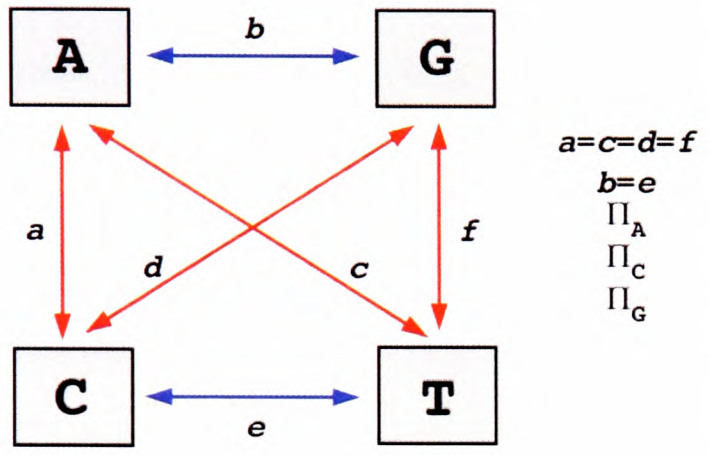


Figure 2.2: HKY85 model of nucleotide substitution. Arrows show six possible substitution patterns for nucleotide data. Transitions shown in blue arrows; transversions in red. The frequencies of each nucleotide add up to 100% ($\prod A+\prod C+\prod G+\prod T = 1.0$). Five parameters assumed by the model are listed.

Bootstrap resampling was undertaken to provide statistical support for individual nodes on a tree (Felsenstein 1985). GARLI (Genetic Algorithm for Rapid Likelihood Inference (Lewis 1998)) software was used to construct maximum likelihood (ML) trees, as described by Yang *et al.* (Yang 2000).

The extent of phylogenetic clustering of taxa was quantified using the maximum parsimony algorithm implemented in MacClade. This computes the mimimum number of mutations (smallest number of evolutionary changes) required to explain the phylogenetic distribution of taxa. This methodology is discussed in more detail in Chapter 4.

2.8 Enrichment of $CD4^+$ and $CD8^+$ T cells from PBMC

2.8.1 Positive / negative selection using Magnetic Cell Separation (MACS)

Fresh or frozen PBMC were prepared for enrichment by re-suspending in MACS buffer (see section 2.14) and passing through a 100um filter to achieve a single cell suspension. $CD4^+$ and $CD8^+$ T cell populations were enriched using magnetic columns and microbeads (Miltenyi Biotech, UK), using the principle illustrated (Fig. 2.3), according to the manufacturer's instructions.

Cell populations were sorted by either positive or negative selection, or a combination of both methods. Positive selection relies on the direct labelling of a cell population with an antibody-labelled magnetic bead, such that these cells can then be selected on passage through a magnetic column. Negative selection leaves the target cells untouched, and labels the majority of other cells using an antibody cocktail. For negative selection of $CD4^+$ T cells, a cocktail of biotin-conjugated anti-CD8, CD14, CD16, CD19, CD36, CD56, CD123, TCR γ/δ and Glycophorin-A antibodies was added, followed by the addition of anti-biotin microbeads. Optimisation of these methods is further discussed in Chapter 7.

2.8.2 Activation of enriched $CD4^+$ and $CD8^+$ T cell populations

The $CD4^-$ T cell fraction was activated in media containing R10 plus PHA at 3 μg/ml plus recombinant IL-2 (Roche, Switzerland) at 5 U/ml. $CD8^+$ T cells were cultured in R10 plus 10% Cellkines™ (Natural Human T cell Growth Factor, Zeptometrix, USA, section 2.14).

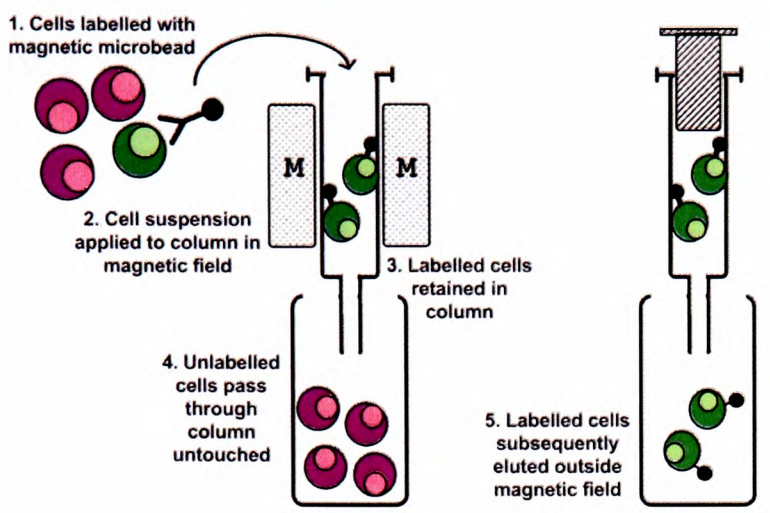


Figure 2.3: Use of magnetic microbeads to sort cell populations. The microbeads are magnetic particles made up of polysaccharide and iron oxide, of around 50 nm diameter. M represents magnetic field.
Figure adapted from product literature: www.miltenyibiotec.com.

2.9 Fluorescence Activated Cell Sorting (FACS)

Aliquots of 2.0-5.0x10^5 cells were spun down in a 96 well plate for surface staining with the panel of antibodies shown (table 2.9) for 15 minutes at room temperature in the dark. Cells were then washed twice in 200 µl PBS and resuspended in 200 µl 2% paraformaldehyde for FACS analysis. FACS was undertaken using a BD FACS Calibur, unless otherwise stated. FACS data were analysed using Flow-jo v. 8.6.3.

Antibody	Fluorochrome	Origin and isotype	Volume of antibody added (in total 50 µl volume)
CD3	FITC	Mouse anti-human IgG_1, κ	2.0 µl
CD4	APC		0.7 µl
CD8	PerCP		3.0 µl

Table 2.9: Antibodies for FACS staining (BD Biosciences immunocytometry systems)

2.10 Use of MHC Class I Tetramers

2.10.1 Tetramerisation of MHC Class I monomer

Monomer stocks were preserved at -80°C in aliquots of 25 μg / 50μl. Tetramerisation was undertaken using Extravidin conjugated to a phycoerythrin (PE) fluorochrome at 0.2mg/ml (Sigma), using a 1:20 molar ratio of extravidin conjugate to monomer. The volume of extravidin conjugate to be added was divided in half. Working on ice and keeping the solution in the dark, the first half of the extravidin conjugate was added. The remaining half was divided into five equal aliquots, added at subsequent intervals of 15 minutes.

2.10.2 Testing tetramer using compensation beads

Following the tetramerisation process, MHC Class I tetramers were tested prior to use. BD comp beads (BD Biosciences, UK) were incubated with 2 μl of mouse anti-human MHC antibody conjugated to FITC (W632, A&B Serotech, UK) for 20 minutes at room temperature in the dark. The beads were washed with PBS by spinning for 5 minutes at 500g. The tetramer was added (2 μl) and the mixture incubated for a further 20 minutes at room temperature in the dark; a further wash was performed and the mixture was resuspended in 200 μl 2% paraformaldehyde.

The samples were run on the BD FACS LSR2. Flo-jo software was used to identify a shift in the PE peak in the FITC positive population, corresponding to FITC-conjugated beads binding the tetramer.

2.10.3 Selection of $CD8^+$ T cell populations using tetramers

Fresh or frozen PBMC were used, with a confirmed optimal response on ELISpot assay. The cells were counted and spun down to a pellet. Tetramer was added at 1 μl per $1x10^6$ PBMC, plus MACS buffer at 50 μl per $1x10^7$ PBMC, and the mixture was incubated for 10 minutes at 4°C. The cells were washed in MACS buffer. Anti-PE microbeads were added at 20 μl per $1x10^7$ cells, and further incubation was performed for 15 minutes at 4°C, followed by washing as previously. The cells were resuspended in 0.5 ml MACS buffer and applied to a pre-washed Miltenyi column in a magnetic field, in accordance with the manufacturer's instructions. The untouched cells passed through the column were collected for use in experiments requiring the tetramer-depleted fraction.

2.11 Generation of $CD8^+$ T cell lines

2.11.1 Derivation of T cell lines from PBMC

$CD8^+$ T cell lines were derived in two ways. First, $5\text{-}10x10^6$ PBMCs from a subject making a known response to an optimal epitope were spun down to a pellet, pulsed with the relevant optimal peptide (10 μl of 200 μg / ml) and resuspended in R10 plus 10% Cellkines™ (as described in section 2.8.2). Alternatively, the tetramer-positive fraction from the MACS process described above (section 2.10.3) was eluted from the column, washed, counted and fed with peptide-pulsed B cells and irradiated feeders (mixed donor lymphocytes) at a ratio of 1:1:1 in the same culture media. Cell lines were tested for specificity by intra-cellular staining for IFN-γ in the presence of PHA (positive control), no stimulation (negative control), and the relevant optimal peptide.

2.11.2 Testing T cell lines by intra-cellular staining (ICS)

T cell lines in culture were tested for specificity by ICS for IFN-γ (Appay and Rowland-Jones 2002). $2.5x10^5$ cells were placed in three wells of a round-bottom 96-well plate. Working at a total volume of 200 µl per well, cells were pulsed with matched peptide (final concentration 200 µg / ml), stimulated with PMA (0.5 µl per well) and ionomycin (0.2 µl per well) (Sigma), or left unstimulated. Anti-CD28 and anti-CD49d (BD Biosciences) were added to each well (final concentration 1 µg / ml). The cells were incubated for 1.5 hours at 37°C in 5% CO_2. Brefeldin A* (Sigma) was added (final concentration 10 µg/ml), and the cells were incubated for a further 5 hours in the same conditions.

The cells were washed in PBS and surface stained for CD3, CD4 and CD8 (see section 2.9.1). Following a further wash, they were re-suspended in cytofix/cytoperm (BD Biosciences) and incubated at 4°C for 20 minutes. Cells were then permeabilised by resuspending in permwash buffer (BD Biosciences), and stained using mouse anti-human IFN-γ on a PE fluorochrome for 20 minutes at room temperature in the dark. A final wash was performed using PBS, and the cells were resuspended in 2% paraformaldehyde for FACS analysis.

2.12 Culture and estimation of titre of HIV-1 *in vitro*

2.12.1 Laboratory viruses

Wild-type HIV-1 subtype B strain NL4-3 (obtained from the AIDS Research and Reference Reagent Program, NIAID, NIH), with a GFP reporter gene in-frame with

* A Golgi transport inhibitor that leads to protein accumulation in the ER

Nef, was used for *in vitro* infection. For the experiments in Chapter 7, a mutant virus was made by Dr Julia Prado, with two mutations in the HLA-B*57-restricted epitope, KF11 (Gag 162-172) (for details see Chapter 7).

2.12.2 Generation of viral stocks

Viral stocks were grown in MT-4 cells (obtained through the AIDS Research and Reference Reagent Program, NIAID, NIH) suspended in R10 at $1.0x10^6$/ml and spun down to a pellet. Virus was added to infect at an MOI of 0.01. The volume was made up to 200 µl / $1x10^6$ MT-4 cells, and the cell suspension incubated at 37°C in 5% CO_2. The cells were washed to remove cell-free virus, and re-suspended in R10 at $1x10^6$ / ml. Growth was monitored by the detection of GFP by FACS over 5-10 days at 37°C in 5% CO_2. To increase yield, additional MT-4 cells were added in aliquots of $5x10^6$ at intervals of 3-4 days. Supernatant was harvested when >80% of cells were infected (Fig. 2.4) and was frozen at -80°C in aliquots of 0.5 ml.

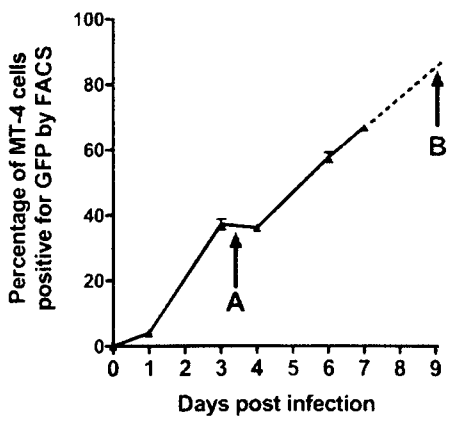


Figure 2.4: GFP production following infection of MT-4 cells.

Mean ± SEM of three parallel experiments is shown, with background GFP subtracted. A: $5x10^6$ additional MT-4 cells added on day 3. B: Supernatant to be harvested on day 9 based on extrapolated growth curve (dashed line).

2.12.3 Estimation of viral titre

Viral titre assays were set up as shown (Fig. 2.5). When different viruses were being used simultaneously in the same assay (e.g. wild-type and mutant virus), titre plates

were run in parallel, in order to control for possible differences in culture conditions that might arise from non-simultaneous experiments.

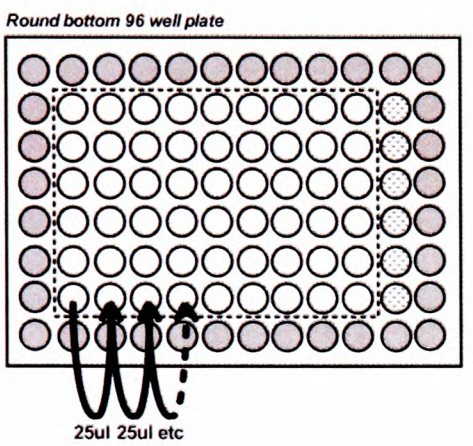


Figure 2.5: Measuring titre of HIV

Titre assays were set up in a 96 well culture plate. The outer wells (grey) were filled with 300 µl PBS. The central wells were filled with 100 µl R10. Virus was added to the wells enclosed within the dotted line, starting with 25 µl to the first column. Serial dilutions were performed using a multi-channel pipette, taking 25 µl from each column and adding to the next. The final row (stippled) was left uninfected.

MT-4 cells were added to each well at 30,000 cells per well in 100 µl R10. The plate was read after 5 days, counting the number of wells positive for GFP in each column using fluorescence microscopy.

The titre of the virus was calculated using the Spearman-Karber method, based on an interpolation of values to find the dilution where 50% of the wells would have been positive, to derive the 50% Tissue Culture Infective Dose, $TCID_{50}$. Using this $TCID_{50}$, the volume of viral stock required for a given MOI can be calculated (see box below).

Tissue Culture Infective Dose ($TCID_{50}$): number of infectious particles required to produce pathological change in 50% of cells (log units/ml).

Multiplicity of infection (MOI): ratio of infectious particles to target cells.

To calculate volume of viral stock to achieve a given MOI:

$$\mathbf{MOI = volume \times (TCID_{50})/(number\ of\ cells\ to\ be\ infected)}$$

E.g. to calculate volume of viral stock needed to infect 1.0×10^6 cells at MOI of 0.01, using viral stock with titre 1.0×10^5 units/ml:

$0.01 = volume \times 1.0 \times 10^5 / 1.0 \times 10^6$

$volume = 0.1$ ml

2.13 Viral Suppression Assays

2.13.1 *In vitro* infection of target cells

The desired number of target cells was spun in a 15 ml falcon tube at 500g for 5 minutes. The supernatant was pipetted off completely. Virus was defrosted at room temperature, and was added to target cells at the desired MOI. R10 was added to make up the final volume to 200 µl per $5.0x10^5$ cells. The cell suspension was incubated at 37°C, 5% CO_2, for 4 hours. The cells were then washed three times in PBS, pipetting all supernatant from the pellet after each wash to remove all cell-free virus. Cells were then resuspended in CD4 culture media at $1.0x10^6$ / ml.

2.13.2 Seven day viral suppression assay

Viral suppression assays were used to quantify the suppression of HIV replication in a CD4-enriched target fraction by autologous HIV-specific $CD8^+$ T cells. A schematic view of the assay is shown (Fig. 2.6), based on previously published methods (Yang *et al.* 1997; Chen *et al.* 2009).

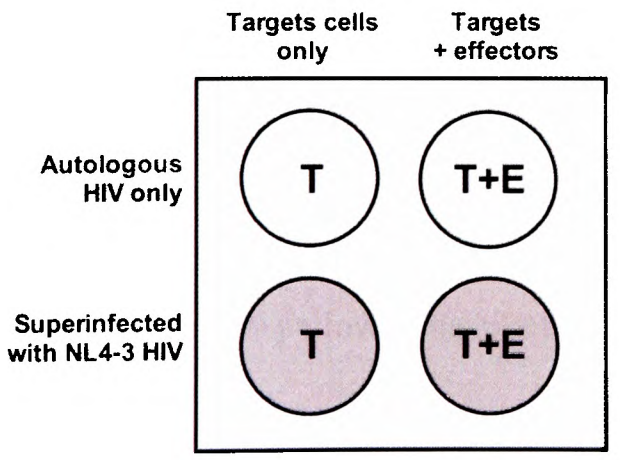


Figure 2.6: Schematic view of viral suppression assay set up in a 96-well round-bottom microtitre plate. T = target cells ($CD4^+$ enriched fraction); E = effector cells ($CD8^+$ enriched fraction). Total number of target cells per well $1x10^5$. Effectors were added at 1:1 ratio. Total volume per well was made up to 200 µl. Clear wells represent those infected with autologous virus only; grey wells represent those superinfected with laboratory virus.

At the required time-points, aliquots of 100 µl of supernatant were removed and frozen for subsequent analysis by p24 ELISA. The well was then topped up with the same volume of fresh media to keep the total volume constant. Further optimization and development of these methods is included in the methods section of Chapter 7.

2.13.3 ELISA assay to measure p24 concentration

Frozen aliquots of supernatant were defrosted at room temperature. Serial ten-fold dilutions of each sample were performed in RPMI in a 96-well plate. The maximum dilution was based on the anticipated p24 concentration (undiluted or 1:10 for target cells in the absence of superinfection, up to a maximum of 1:100,000 for superinfected cells at day 7). p24 ELISA kits (Innotest HIV Antigen mAb, Microgen Bioproducts) were used in accordance with the manufacturer's instructions.

Briefly, biotinylated monoclonal anti-p24 antibody plus NP40* was incubated with the sample for 60 minutes at 37°C. The plate was washed and peroxidase-conjugated streptavidin added for a further 30 minutes at 37°C. The plate was washed and substrate buffer (peroxidase substrate plus chromogen) was added for a further 30 minute incubation at room temperature in the dark, producing a blue colour (Fig. 2.7).

The reaction was terminated by adding 50ul sulphuric acid to each well, producing a colour change to yellow. The plate was read using a Multiskan platereader (Thermo Scientific).

* Nonyl phenoxylpolyethoxylethanol, a detergent that disrupts the cytoplasmic membrane, allowing release of cytoplasmic contents of the cell.

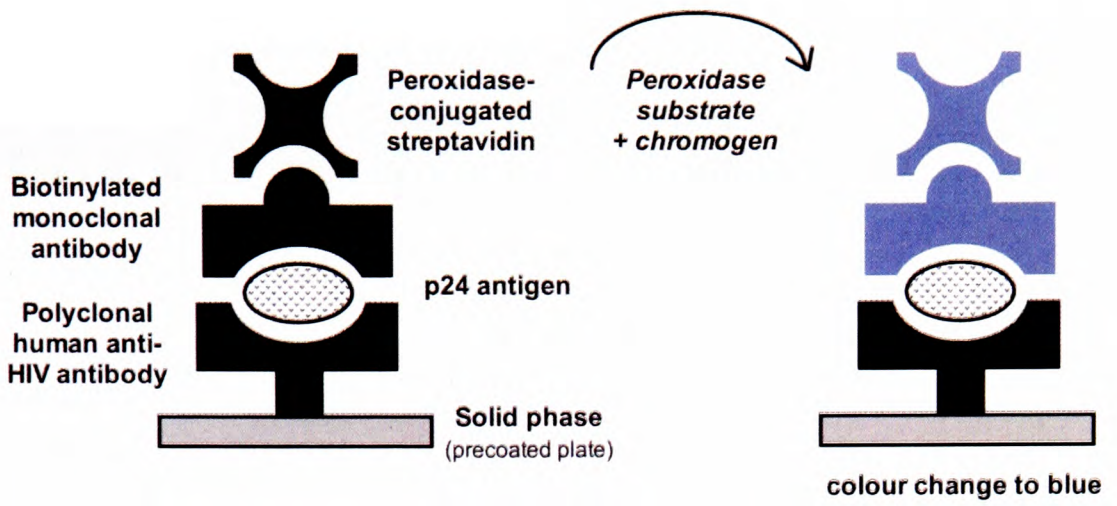


Figure 2.7: Schematic view of Innotest p24 ELISA assay. Figure adapted from product literature, www.microgenbioproducts.com.

Controls were set up for each plate, composed of serial dilutions of p24 antigen positive control (supplied by the manufacturer). A linear regression was calculated using Ascent software (Fig. 2.8); the results were only accepted if $r^2>0.96$.

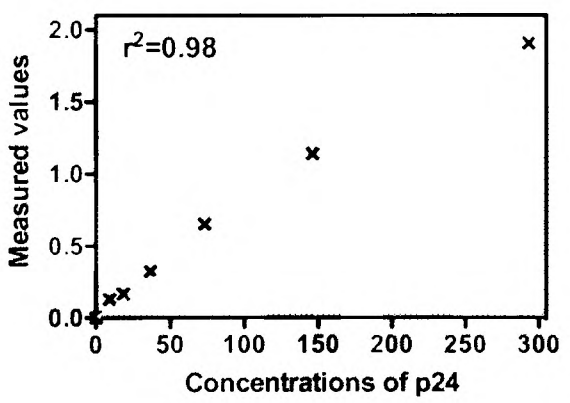


Figure 2.8: Standard curve plotted from positive controls in p24 ELISA assay using serial two-fold dilutions of p24 antigen. Values for absorbance ('measured values' output from Multiskan platereader) were read at 450nm and were used to construct a standard curve, from which the results of the sample wells can be read.

2.14 Supplementary methods tables

Reagent	Components		Manufacturer
R10	RPMI	500 ml	Sigma
	Foetal calf serum*	50 ml	HyClone
	Penicillin-streptomycin	5 ml	Sigma
	L-glutamine	5 ml	Sigma
CD4 culture media	R10	50 ml	As above
	PHA (250 µg / ml)	600 µl	Sigma
	Recombinant IL-2 (10^4u/ml)	250 µl	Roche
CD8 culture media (10% 'T-stim')	R10	45 ml	As above
	Cellkines	5 ml	Zeptometrix
Freezing Solution	Foetal calf serum*	9 ml	HyClone
	DMSO	1 ml	Sigma, UK
MACS buffer	PBS	500 ml	Invitrogen
	EDTA (0.5M)	2 ml	Sigma
	Foetal calf serum*	2.5 ml	HyClone

Table 2.10: Laboratory media. * FCS heat-inactivated at 56°C for 30 minutes, and filtered through 0.45 µm filter.

Software / tool	Function	Source / manufacturer
Ascent v.2.6	Microplate reader for p24 ELISA	Thermoscientific
Bootscore v.3.1.1	Reading of phylogenetic bootstrap values	http://sourceforge.net/projects/bootscore
Entropy-One, Entropy-Two	Calculation of Shannon Entropy	Los Alamos HIV databases www.hiv.lanl.gov/content/sequence/ENTROPY
Fig tree v. 1.1.2	Tree figure drawing tool	Institute of Evolutionary Biology, University of Edinburgh http://tree.bio.ed.ac.uk/
Flow-Jo v. 8.6.3	Analysis of FACS data	Stanford University www.flowjo.com
GARLI (Genetic Algorithm for Rapid Likelihood Inference)	Generation of maximum likelihood phylogenetic trees	National Evolutionary Synthesis Center (NESCent), N. Carolina, USA, http://www.nescent.org/informatics
Los Alamos HIV database	HIV sequence database and analysis tools	www.hiv.lanl.gov
MacClade v.4.0 (Maddison and Maddison 2000)	Phylogenetic analysis; maximum parsimony algorithm	Sinauer Associates, USA http://macclade.org/macclade.html
MacPyMol 2006	Viewing 3D molecular structures	DeLano Scientific LLC
PAUP* v.4.0	Construction of neighbour joining phylogenetic trees	Sinauer Associates, USA http://paup.csit.fsu.edu/
Prism v. 5.0	Statistical analysis and figure production	Prism for Mac, December 2007 Graphpad Software Inc.
Protein Data Bank	Download of 3D protein files	Research Collaboratory for Structural Bioinformatics (RCSB) http://www.rcsb.org/pdb
Se-Al v. 2.0a11	Alignment of DNA / protein sequences	Andrew Rambaut, Department of Zoology, University of Oxford
Sequencher v. 4.8	Analysis of sequence data	Gene Codes Corporation 1991-2007

Table 2.11: Software and on-line tools

CHAPTER 3: Effect of reverting mutations in control of viraemia in HIV

The majority of the work presented in this chapter has been published in *J. Virol.* 2008 82(17):8548-59.
Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.
Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, Brander C, Bishop K, Mlotshwa N, Mullins JI, Coovadia H, Ndung'u T, Walker BD, Heckerman D, Goulder PJ.

3.1 INTRODUCTION

3.1.1 Relationship between $CD8^+$ T cell responses that target Gag epitopes and control of HIV viraemia

The failure of the Merck T cell-based HIV vaccine (Steinbrook 2007; Moore *et al.* 2008) has emphasised the need to refocus on the question of what $CD8^+$ cell responses need to be induced by a successful vaccine. One of the strongest clues to immune control of HIV comes from the consistent associations observed between expression of particular HLA class I alleles, such as HLA-B*5703 and B*2705, and low viral setpoint, and between other alleles, such as HLA-B*3502 and B*5802, and high viral setpoint (Kaslow *et al.* 1996; Migueles *et al.* 2000; O'Brien *et al.* 2001; Goulder and Watkins 2004; Geldmacher *et al.* 2007). The HIV-specific $CD8^+$ T cell responses generated in infected individuals who have HLA-B*57 or B*27 are dominated, both in acute and chronic infection, by $CD8^+$ T cells that target Gag epitopes (Nixon *et al.* 1988; Goulder *et al.* 1996; Goulder, Phillips *et al.* 1997; Altfeld *et al.* 2003; Migueles *et al.* 2003; Friedrich *et al.* 2004).

A critical role of Gag-specific responses in control of viraemia has been established in studies of both SIV and HIV (Klein *et al.* 1995; Riviere *et al.* 1995; Ogg *et al.* 1998;

Masemola *et al.* 2004; Zuniga *et al.* 2006; Geldmacher *et al.* 2007), and population-based studies show that, irrespective of HLA type, broad Gag-specific $CD8^+$ T cell responses are strongly associated with decreasing viral load (Borghans *et al.* 2007; Kiepiela *et al.* 2007). In addition to HLA-B and broad Gag-specific $CD8^+$ T cell responses, a genome-wide association study indicated that HLA-C-restricted responses may also play an important role in immune control of HIV (Fellay *et al.* 2007). Viral kinetic data from the SIV model have shown that not only Gag-specific (Sacha, Chung, Rakasz *et al.* 2007), but also Pol-specific $CD8^+$ T cells are activated by virus-infected cells prior to *de novo* synthesis of viral proteins and Nef-mediated MHC class I down-regulation (Sacha, Chung, Reed *et al.* 2007).

3.1.2 The effect of reverting mutations on control of viraemia

Alleles such as HLA-B*57 and B*27 select escape mutations within Gag epitopes that impose high fitness costs on the virus, and are therefore likely to revert back to wildtype following transmission to HLA-mismatched recipients (Martinez-Picado *et al.* 2006; Brockman *et al.* 2007; Schneidewind *et al.* 2007). Such mutations may constrain viral replication or reduce infectivity; thus, despite allowing immune escape from the $CD8^+$ T cell response, these polymorphisms may play an important role in disease control (Matano *et al.* 2004; Fernandez *et al.* 2005; Kobayashi *et al.* 2005).

3.1.3 Hypothesis and aims

The aim of the analysis presented in this chapter was to address the role of HLA-mediated selection of viral escape mutants that revert post-transmission, based on the hypothesis that the selection of reverting mutations (due to the fitness cost of the polymorphism) may contribute to control of viraemia.

This study aimed to assess the impact of all Class I loci and to analyse responses across Gag, Pol and Nef. HLA-C restricted $CD8^+$ T cell responses have been understudied compared to HLA-A and HLA-B (Goulder, Bunce *et al.* 1997). Likewise, $CD8^+$ T cell responses to Pol have been less well studied than Gag and Nef. Therefore, it was important to employ methodology free of any bias that would result from studying associations that fall only within previously defined epitopes. In addition, the analysis aimed to account for viral lineage, in order to reduce bias resulting from founder effect (Bhattacharya *et al.* 2007).

3.2 METHODS

3.2.1 Study cohort and origin of sequence data

Sequence analysis was undertaken on sequence data from 710 adult patients from the Durban cohort. ELISpot data, viral loads (VL) and CD4 counts were available for 681 study subjects. The median VL of this cohort was 37,200 HIV RNA copies/ml plasma (IQR 6,485-133,5000), and median CD4 count was 387 cells/mm^3 (IQR 257-520).

Gag, Pol, and Nef sequences were obtained from proviral DNA for 616, 446, and 436 study subjects, respectively. In addition, viral sequences from plasma RNA were contributed by C. Rousseau (Seattle, USA) for an additional 57, 15, and 7 study subjects, respectively. Thus, from the total of 710 study subjects, 673 Gag, 461 Pol, and 443 Nef sequences were available.

3.2.2 HLA typing

This analysis was performed using high resolution HLA typing, with the exceptions of HLA-A*74, HLA-B*41, and HLA-Cw*02 in which all alleles were grouped by the two digit type in order to remove ambiguous four-digit data arising as a result of different HLA nomenclature.

3.2.3 Methods for identification of sites of HLA-mediated selection

Fisher's Exact Test

Fisher's Exact Test has been used as a simple statistical approach to identifying amino acid polymorphisms in HIV that are associated with host HLA Class I expression (Moore *et al.* 2002). However, this method carries the potential risk of false positive associations arising as a consequence of founder effect: that is, viral genetic polymorphisms are artefactually associated with the expression of a common HLA allele in the population, but are in fact due to the descent of viruses from a common ancestor and not due to positive selection by the allele in question (Bhattacharya *et al.* 2007).

Lineage-corrected approach

Computational analysis of sequence and HLA data was developed and performed by D. Heckerman and J. Carlson at Microsoft, based on methods previously described by Bhattacharya *et al.* (Bhattacharya *et al.* 2007). This method takes into account the phylogenetic structure of the taxa to be examined in order to correct for founder effect. Briefly, a maximum likelihood phylogenetic tree was constructed from all the available sequences for each individual protein. For every HLA allele, amino acid position, and amino acid at that position, two models of the observed presence or

absence of the amino acid in each sequence were created - one representing the null hypothesis that the observations are generated by the phylogenetic tree alone, and the other representing the alternative hypothesis that escape or reversion takes place due to HLA-selection pressure (Fig. 3.1).

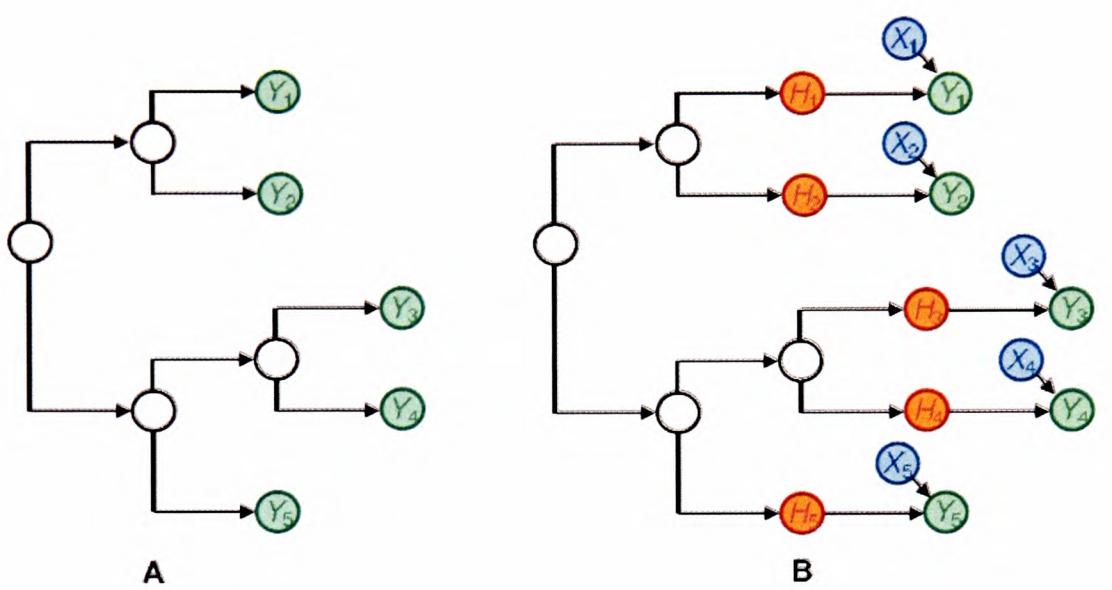


Figure 3.1: Two univariate models of the observed presence or absence of a specific amino acid residue in a given HIV sequence. A: The null model, in which an amino acid evolves independently down the tree until it reaches a leaf. B: The alternate model, in which an amino acid evolves independently down the tree until it reaches an individual, where it is influenced by selection pressure from the predictor (X_i). The variable H_i (for the i^{th} individual) represents the variable Y_i had there been no influence from X_i. Only the Y_i and X_i are observed. Figure reproduced from Carlson et. al, (Carlson et al. 2008) under the terms of the Creative Commons Attribution Licence granted by PLoS journals. doi:10.1371/journal.pcbi.1000225.g002.

The likelihood of the observations was then maximized over the parameters of the null model and the alternative model (as described in Fig. 3.1). A p value was computed using a likelihood ratio test:

Pr(data | alternative model) / Pr(data | null model).

3.2.4 Optimisation of a lineage-corrected approach to identification of HLA-selected polymorphisms

In order to optimise this approach to the detection of HLA-associations, further analysis was made of this model (in collaboration with Microsoft) in order to optimise the final analysis.

(i) Length of sequences analysed

In running analysis of full gene sequences, missing data in some taxa was a possible limitation in the accurate reconstruction of phylogeny. In order to establish the optimum approach, lineage-corrected analysis was performed to compare the identification of polymorphisms in p17 Gag, first using p17 sequences alone and subsequently using p17+p24+p15. A total of 16 possible HLA-associations were identified at 11 amino acid residues in p17 ($q<0.2$). Of these, 13 (81%) were identified both by analysis of p17 alone, and by the analysis of whole Gag (p17+p24+p15) sequences. Given the equivalent performance of these two methods, final analysis was undertaken from separate phylogenetic trees constructed for the proteins or protein components shown (Table 3.1).

(ii) Calculation of q-value, false detection rate (FDR)

A q-value was computed for each association, and values were estimated separately for Gag, Pol and Nef associations. The final data were presented initially based on a cut-off of $q<0.2$ (estimating 20% FDR), and also on the more stringent $q<0.05$ (5% FDR).

Protein fragment	Number of taxa	Number of amino acid residues*
p17 Gag	598	
p24 Gag	666	541
p15 Gag	433	
Protease	394	
RT	417	1044
Integrase	262	
Nef	443	225

Table 3.1: Protein fragments used for lineage-corrected analysis.

*Number of nucleotides / residues includes insertions.

(iii) Correction for linkage disequilibrium (LD) between Class I HLA alleles

The univariate model described above (Fig. 3.1) does not account for the possibility that there may be statistical correlations between predictor attributes. However, given that HLA alleles are transmitted in LD, this assumption is flawed. Therefore, the algorithm was further adapted to a multivariate model, correcting for LD between HLA alleles by inclusion of multiple HLA predictors to maximize elimination of spurious associations. For every amino acid at each position, the HLA-allele with the strongest association (and its corresponding p and q-values) was added to the list of identified associations. The analysis was then repeated after removing individuals having this HLA allele. This procedure was iterated until no HLA allele yielded an association with $p<0.05$.

3.2.5 Interpretation of statistical data from lineage-corrected algorithm

The analysis performed by Microsoft identified associations termed escape, attraction, reversion and repulsion, all reflecting a statistical association between HLA Class I expression and HIV sequence polymorphisms. The data generated were then re-classified to identify each association according to the biological processes of escape and reversion, as described below.

Definitions of escape and reversion

Broadly speaking, 'escape' describes amino acid selection away from consensus in the presence of a specific HLA allele, and 'reversion' describes amino acid selection towards consensus in the absence of a specific HLA allele.

ESCAPE was defined as the statistical association between the expression of a particular HLA allele with either:

i. A decrease in number of the consensus amino acid, or

ii. An increase in number of a variant amino acid.

Within the category of escape polymorphisms, 'negative' associations were also included, defined as the presence of an HLA allele in statistical association with the consensus residue. These associations are accounted for by escape mutations that do not revert, such that wild-type is eventually replaced by the HLA-selected variant, and a new population consensus is established (Moore *et al.* 2002; Leslie *et al.* 2005).

REVERSION was detected statistically by association between the absence of a particular HLA allele with either:

i. An increase in number of the consensus amino acid, or

ii. A decrease in the number of the variant amino acid.

These associations suggest that a mutation selected by a particular HLA type returns to consensus following transmission to an HLA-mismatched host. These mutations can be inferred to impose a fitness cost that is not compensated by the selection of further mutations.

In some instances, reversion was identified in the absence of escape. This can be accounted for by instances in which multiple different escape variants can be selected, but reversion always selects for return to the wild-type amino acid. Thus the statistical association between the selecting HLA allele and each escape variant may be too weak to reach the threshold for significance, but reversion to a single residue will reach statistical significance. For the purposes of analysis, 'reversion only' associations were assumed to occur in association with escape.

In order to process the results generated by the algorithm, the following steps were undertaken (example shown in Table 3.2). First, the site of the association was corrected to correspond to HXB2. In this example, there is an insertion of ten amino acids downstream of the site of association, causing the difference between alignment position and HXB2 position, so position 252 in the sequence alignment corresponds to HXB2-242. Second, the consensus amino acid at each position was identified, based on the sequences input to the algorithm (Appendix I).

Association	HLA Allele	Alignment position	HXB2 position	Amino acid	Consensus amino acid	TT	TF	FT	FF	p value	q value
Attraction	B*5703	252	242	N	T	35	4	75	531	1.49E-16	0
Escape	B*5703	252	242	T	T	4	35	516	90	3.67E-15	0
Repulsion	B*5703	252	242	N	T	35	4	75	531	6.12E-13	0
Reversion	B*5703	252	242	T	T	4	35	516	90	4.04E-12	0

Table 3.2: Sample output from lineage-corrected algorithm to detect associations between HLA Class I expression and HIV sequence polymorphisms in Durban gag sequences.

In this example, the algorithm returns all columns headed in grey. For the purposes of further analysis, the columns headed in black have been added. The statistical association is defined by the model as attraction, escape, repulsion or reversion, and is attributed to a given amino acid (irrespective of consensus, which is not computed by the algorithm). The results of a 2x2 contingency table are presented; the first letter in each pair pertains to the presence (T, true) or absence (F, false) of the HLA allele in question, and the second to the presence (T) or absence (F) of a given amino acid residue.

This example demonstrates how data generated by the algorithm may define multiple statistical associations that all reflect the same biological effect. Here, the association between HLA-B*5703 and escape from threonine (T) to asparagine (N), that reverts following transmission to an HLA-mismatched host, has been detected in four ways:

- **'Escape'** from threonine: this identifies the process of escape from T in 35 of 39 subjects with HLA-B*5703, but the mutant amino acid is not specified. From the contingency table, it can also be seen that an amino acid other than T is also detected at this position in 90 of 606 HLA-B*5703-negative subjects.
- **'Attraction'** to N: duplicates the identification of escape, but now identifies the selected mutant as N in all 35 of the HLA-B*5703 subjects. In the HLA-B*5703 negative population, only 75 subjects have N, explaining why this association is even stronger than that detected for escape.
- **'Reversion'** to T: identifies that in the HLA-B*5703 negative population, T is statistically favoured. Having identified T as consensus, from this association,

we can infer reversion. As the only escape variant seen in association with HLA-B*5703 is N, the numbers in the 2x2 table are the same as for escape.

- **'Repulsion'** from N: duplicates the identification of reversion from N to T, with the same numbers as for attraction.

Such instances of multiple detection of a single HLA selection event were corrected, such that this association would be reported only once, as an escape mutation that is statistically predicted to revert following transmission to an HLA-mismatched recipient. Further de-duplication was undertaken to remove repeat associations defined for two digit and four digit associations, reporting the four digit association when available.

3.2.6 Evaluation of lineage-corrected methodology

Additional exploration of these methods was undertaken in order to assess the performance of this approach to identification of HLA-associated polymorphisms.

Comparison of two methods used for identification of HLA-associated polymorphisms

Comparison of Fisher's Exact Test and lineage-corrected methodology for detection of HLA-associated polymorphisms demonstrated that the majority of associations (62%) were detected by both methods (Fig. 3.2). However, a further 29% were detected only by Fisher's Exact Test; these associations were lost following phylogenetic correction, reflecting the increased specificity of the latter method.

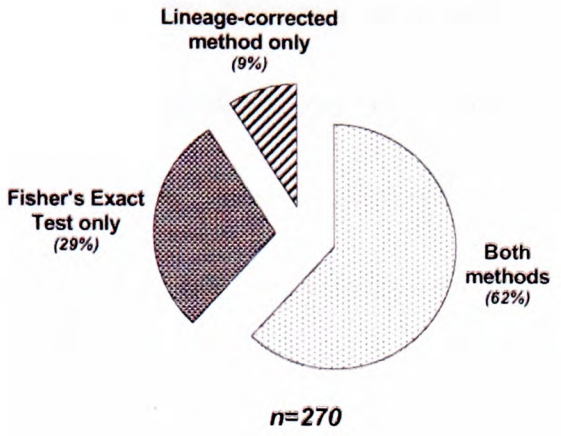
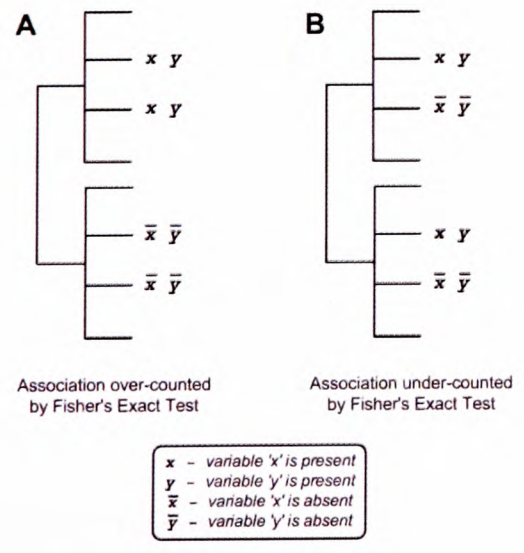


Figure 3.2: Pie chart to show the proportion of HLA-associations with HIV sequence polymorphisms detected by different statistical approaches.

Data generated from analysis of p24 Gag sequences from Durban cohort.

These findings are in keeping with analysis of synthetic data subsequently published by Carlson *et al.*, in which failure to account for phylogeny leads to a substantial over-estimate of significance (Carlson *et al.* 2008). However, an additional 9% of associations were detected only by lineage-corrected methodology, reflecting the additional sensitivity of this approach in detecting true associations (Carlson *et al.* 2008). The explanation for the differences in associations detected by the two approaches is shown in Fig. 3.3.

Figure 3.3: Model phylogenetic trees showing the reasons for over and under-estimation of association between two variables, *x* and *y*, using Fisher's Exact Test.



A: Using Fisher's Exact Test alone, variables *x* and *y* appear statistically related. However, the presence or absence of amino acid *x* is well accounted for by the phylogeny, and its appearance in adjacent branches should not be considered as two independent observations. Ignoring the phylogenetic structure leads to over-counting of such associations (false positives).

B: The presence and absence of amino acid *x* in adjacent branches of the phylogeny is unexpected, until variable *y* is accounted for. Ignoring the phylogenetic structure leads to under-counting of such associations (false negatives).

Relationship between number of sequences used for analysis and sensitivity of algorithm in detecting HLA associations

Up to a maximum of around 2000 sequences, the relationship between the number of sequences and the number of HLA associations detected is approximately linear (J. Carlson, personal communication; Fig. 3.4). The statistical power to detect polymorphisms in Gag in this analysis was therefore greater than that in Nef and Pol. For this reason, an analysis of Gag sequences was repeated using 430 taxa (selected at random from the total pool), in order to make sequence numbers more comparable to Nef and Pol.

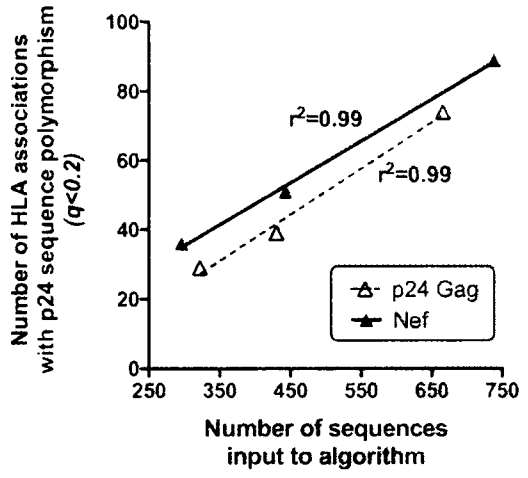


Figure 3.4: Relationship between number of sequences input to lineage-corrected algorithm and number of HLA associations detected.
Data are all taken from Durban cohort, with the exception of 739 Nef sequences that were taken from pooled South Africa (Durban), Botswana (Gaborone) and Zimbabwe cohorts. r^2 values by linear regression. The same pattern is described by Rousseau *et al.* in the analysis of *Gag, Pol, Env* and *Nef* genes (Rousseau *et al.* 2008).

Relationship between frequency of HLA allele and number of HLA associations detected

Although, inevitably, the algorithm is better powered to detect associations selected by HLA alleles occurring at a higher phenotypic frequency, it was important to rule out this relationship as a primary determinant of the outcome of sequence analysis. No such correlation was observed (Fig. 3.5), confirming that the potency of selection by certain HLA-alleles is independent of the population frequency of the allele.

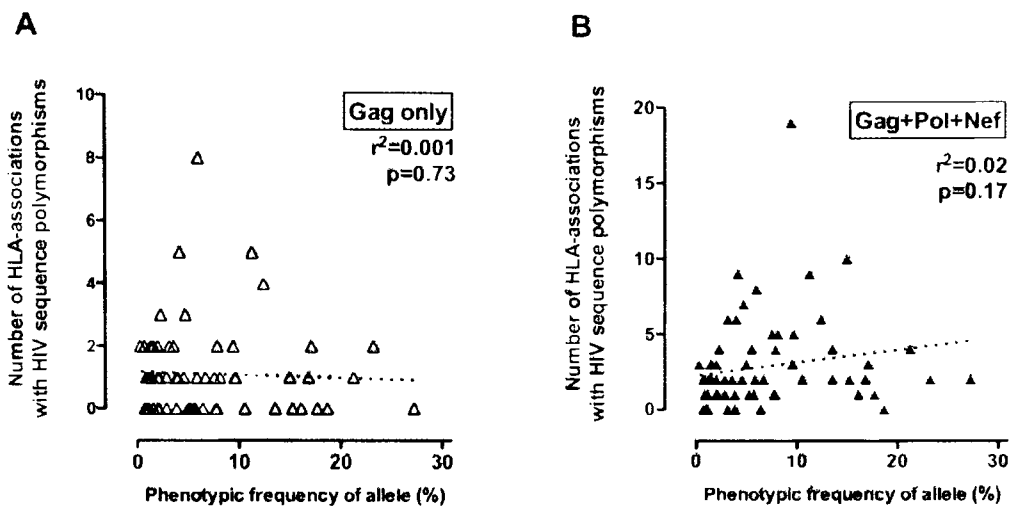


Figure 3.5: Relationship between phenotypic frequency of all HLA-A, B and C alleles in the Durban cohort with number of associated HIV sequence polymorphisms. A: Polymorphisms in Gag only. B: Total polymorphisms in Gag, Pol and Nef. r^2 and p values by linear regression.

3.2.7 Identification of HLA associations with 18-mer peptide recognition

IFN-γ ELISpot assays were performed for 681 subjects from the Durban cohort, using a screening approach by megamatrix using overlapping peptides (OLPs) spanning the entire HIV proteome. Associations between recognition of individual 18-mer OLPs and expression of particular HLA class I alleles was sought using Fisher's Exact Test. 65 HLA class I alleles expressed at a phenotypic frequency of $\geq$0.5% were included in the analysis. To correct for multiple comparisons, the conservative Bonferroni correction was used, in order to make this analysis comparable to previous studies (Kiepiela *et al.* 2007). Thus, associations of $0.05>p>0.00075$ were lost following correction for the 65 alleles analysed.

3.2.8 Statistical analysis of data

For subjects with each allele, the median VL and $CD4^+$ T cell count were calculated. In order to quantify the relationship between sequence polymorphisms, or ELISpot responses, associated with each allele and the median VL or $CD4^+$ T cell count for than allele, a Spearman rank correlation was performed.

3.3 RESULTS

3.3.1 HLA Class I expression in 710 C-clade infected South African subjects

A total of 65 HLA class I molecules were expressed at ≥0.5% phenotypic frequency in this cohort (Table 3.3).

HLA Class I allele	Phenotypic frequency of allele (%)	Median VL (RNA copies/ ml plasma)	HLA Class I allele	Phenotypic frequency of allele (%)	Median VL (RNA copies/ ml plasma)	HLA Class I allele	Phenotypic frequency of allele (%)	Median VL (RNA copies/ ml plasma)
A*0101	4.7	10878	B*0702	6.7	27900	Cw*02	18.7	50800
A*0201	6.4	20000	B*0705	0.9	86600	Cw*0302	2.0	20000
A*0202	2.1	32250	B*0801	8.1	56900	Cw*0304	0.3	39700
A*0205	7.9	24250	B*1302	1.6	23050	Cw*04(01)	10.6	30750
A*0301	7.6	66550	B*1401	3.9	21900	Cw*0501	1.1	124000
A*2301	13.6	29300	B*1402	1.1	76200	Cw*0602	27.3	49875
A*2402	5.3	38950	B*1503	15.3	55300	Cw*0701	17.7	33800
A*26(01)	2.1	38300	B*1510	17.1	51700	Cw*0702	5.7	31100
A*2902	1.6	19620	B*1516	1.1	15017	Cw*0704	2.0	152500
A*3001	16.1	32700	B*1801	5.6	84900	Cw*0801	3.6	50450
A*3002	13.6	55550	B*3501	3.1	40000	Cw*0802	3.9	37800
A*3004	2.9	57650	B*3910	4.1	23850	Cw*0804	1.4	*no data*
A*32(01)	1.0	3890	B*41	0.7	139000	Cw*1203	2.9	20700
A*33	3.9	33800	B*4201	15.0	25600	Cw*1601	9.6	56900
A*34(02)	5.0	40900	B*4202	2.0	109000	Cw*1604	1.3	22500
A*3601	0.9	32500	B*4403	9.4	27600	Cw*1701	23.3	31000
A*4301	4.6	46600	B*4501	7.9	41000	Cw*1801	7.7	28300
A*6601	5.9	37750	B*4901	1.4	157500			
A*6602	1.1	73100	B*51(01)	0.7	135500			
A*6801	5.9	72600	B*5301	4.0	21700			
A*6802	16.8	41000	B*5702	2.3	15190			
A*74	9.7	8180	B*5703	6.0	7642			
A*8001	1.1	74500	B*5801	12.4	16700			
			B*5802	21.3	60200			
			B*8101	11.3	12800			
n=23			*n=25*			*n=17*		

Table 3.3: 65 HLA Class I molecules expressed at ≥0.5% phenotypic frequency in Durban, South Africa. Phenotypic frequency of the allele and median viral load (VL) for subjects expressing this allele is shown. Four digit types shown in brackets represent instances where high resolution genotype was inferred in some subjects based on known HLA types in this population.

3.3.2 HLA-associated HIV amino acid polymorphisms selected by HLA-B and located in p24 Gag predominate

In order to take account of the possibility that certain HLA class I molecules are associated with a characteristic viral set-point as a result of the selection of particular escape mutants, statistical associations between amino acid polymorphisms in the three most immunogenic proteins, Gag, Pol and Nef and expression of any of these 65 HLA class I molecules were determined. In total, HLA-associated polymorphisms were identified at 218 different amino acid residues (57 in Gag, 125 in Pol and 36 in Nef), with HLA-B associations and p24 Gag mutations predominating (Fig. 3.6; Tables 3.4-3.6).

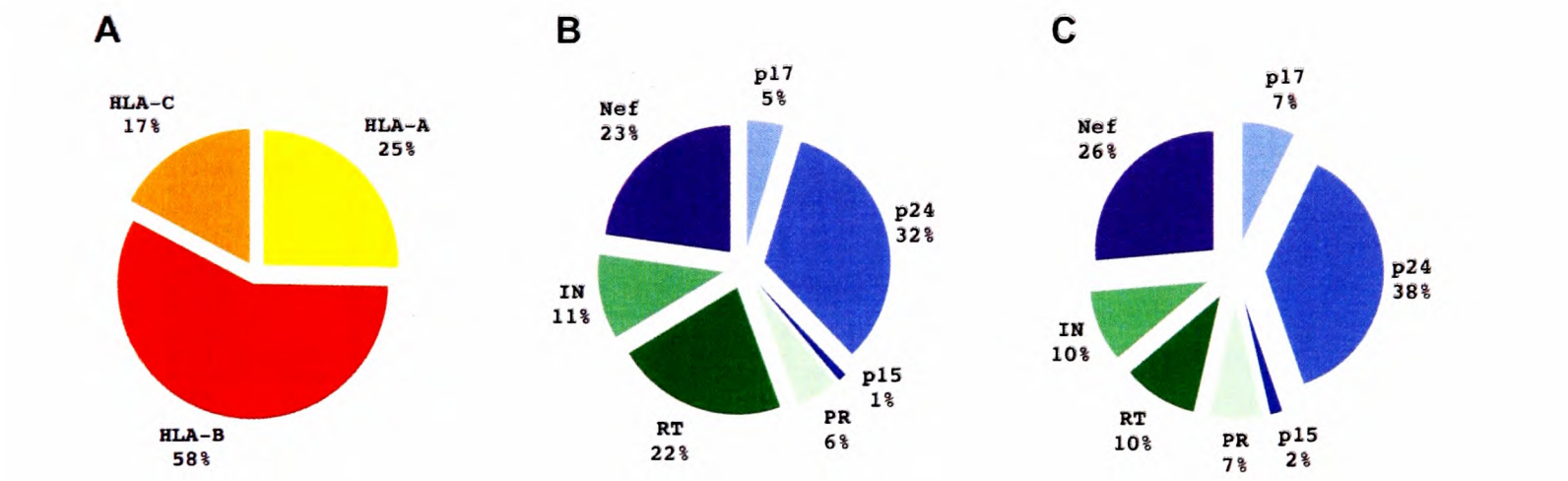


Figure 3.6: Breakdown of all HLA associations with HIV sequence polymorphisms in Gag, Pol and Nef (q<0.2). A: Proportion of all associations accounted for by HLA-A, B and C. Ambiguities due to linkage disequilibria removed from analysis. B: Total proportion of all HLA associations in each protein. C: Proportion of HLA associations corrected for number of amino acid residues in the protein.

Reducing the significance threshold to q<0.05 (5% FDR), 84 associations were identified between expression of particular HLA class I alleles and particular HIV amino acid polymorphisms, of which 61 (73%) were HLA-B associated.

Protein	HLA	Polymorphism location																				p value	q value	Reversion
p17 Gag	A*74, B*5703	G	A	R	A	S	I	L	R	G	E	K	L	D	K	W	E	K	I	R	L	R	2.71E-14	0.00	R
	A*74	G	E	K	L	D	K	W	E	K	I	R	L	R	P	G	G	K	K	H	Y	M	6.47E-10	0.00	R
	A*6801	L	R	P	G	G	K	K	H	Y	M	L	K	H	L	V	W	A	S	R	E	L	1.87E-04	0.13	R
	A*01(01)	G	K	K	H	Y	M	L	K	H	L	V	W	A	S	R	E	L	E	R	F	A	5.17E-04	0.16	
	A*01(01)	Q	T	G	T	E	E	L	K	S	L	F	N	T	V	A	T	L	Y	C	V	H	8.13E-05	0.05	
p24 Gag	A*3004							P	I	V	Q	N	L	Q	G	Q	M	V	H	Q	A		2.24E-04	0.19	R
	A*0301	P	R	T	L	N	A	W	V	K	V	I	E	E	K	A	F	S	P	E	V	I	1.39E-03	0.11	
	A*0202	N	A	W	V	K	V	I	E	E	K	A	F	S	P	E	V	I	P	M	F	T	4.26E-04	0.06	R
	A*3601	P	Q	D	L	N	T	M	L	N	T	V	G	G	H	Q	A	A	M	Q	M	L	2.75E-03	0.15	
	A*01(01)	T	V	G	G	H	Q	A	A	M	Q	M	L	K	D	T	I	N	E	E	A	A	2.77E-03	0.18	R
	A*6602	E	A	A	E	W	D	R	L	H	P	V	H	A	G	P	I	A	P	G	Q	M	5.86E-04	0.07	
	A*02	P	I	A	P	G	Q	M	R	E	P	R	G	S	D	I	A	G	T	T	S	T	1.17E-03	0.15	R
	A*6802	R	M	Y	S	P	V	S	I	L	D	I	R	Q	G	P	K	E	P	F	R	D	1.42E-03	0.11	
	A*02	R	F	F	K	T	L	R	A	E	Q	A	T	Q	D	V	K	N	W	M	T	D	2.28E-03	0.18	R
	A*03	R	F	F	K	T	L	R	A	E	Q	A	T	Q	D	V	K	N	W	M	T	D	2.16E-03	0.18	R
	A*4301	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	M	T	A	C	2.44E-03	0.15	
	A*6601	G	V	G	G	P	G	H	K	A	R	V	L										1.06E-03	0.10	
Protease	A*74			F	F	R	E	N	L	A	F	P	Q	Q	G	G	E	A	R	E	F	P	4.83E-05	0.02	
	A*0301	I	T	L	W	Q	R	P	L	V	S	I	K	V	G	G	Q	I	K	E	A	L	3.75E-04	0.13	R
RT	A*3402	N	P	Y	N	T	P	I	F	A	I	K	K	K	D	S	T	K	W	R	K	L	4.28E-06	0.00	
	A*0205	S	P	A	I	F	Q	S	S	M	T	K	I	L	E	P	F	R	A	K	N	P	1.08E-04	0.05	R
	A*0301	S	P	A	I	F	Q	S	S	M	T	K	I	L	E	P	F	R	A	K	N	P	3.87E-04	0.14	
	A*0301	M	T	K	I	L	E	P	F	R	A	K	N	P	E	I	V	I	Y	Q	Y	M	4.99E-06	0.00	R
	A*3601	H	P	D	K	W	T	V	Q	P	I	Q	L	P	E	K	D	S	W	T	V	N	3.83E-04	0.14	R
	A*74	A	S	Q	I	Y	P	G	I	K	V	R	Q	L	C	K	L	L	R	G	A	K	3.17E-04	0.13	
	A*2301	K	A	L	T	D	I	V	P	L	T	E	E	A	E	L	E	L	A	E	N	R	6.11E-04	0.17	
	A*03	H	T	N	D	V	K	Q	L	T	E	A	V	Q	K	I	A	L	E	S	I	V	4.29E-04	0.13	R
	A*66	R	L	P	I	Q	K	E	T	W	E	T	W	W	T	D	Y	V	Q	A	T	W	4.04E-04	0.13	
	A*6601	Y	Q	L	E	K	E	P	I	A	G	A	E	T	F	Y	V	D	G	A	A	N	1.70E-04	0.08	
	A*26(01)	Y	V	D	G	A	A	N	R	E	T	K	I	G	K	A	G	Y	V	T	D	R	4.72E-04	0.14	R
	A*3402	R	E	T	K	I	G	K	A	G	Y	V	T	D	R	G	R	Q	K	I	V	S	4.16E-04	0.14	
	A*68	Y	V	T	D	R	G	R	Q	K	I	V	S	L	T	E	T	T	N	Q	K	T	5.38E-04	0.15	
	A*74	I	V	S	L	T	E	T	T	N	Q	K	T	E	L	Q	A	I	Q	L	A	L	6.03E-04	0.15	
Integrase	A*0101	E	I	V	A	S	C	D	K	C	Q	L	K	G	E	A	I	H	G	Q	V	D	2.90E-03	0.15	R
	A*0101	C	D	K	C	Q	L	K	G	E	A	I	H	G	Q	V	D	C	S	P	G	I	4.71E-03	0.18	R
	A*3402	I	H	G	Q	V	D	C	S	P	G	I	W	Q	L	D	C	T	H	L	E	G	5.95E-03	0.18	R
	A*3002	S	M	N	K	E	L	K	K	I	I	G	Q	V	R	D	Q	A	E	H	L	K	5.79E-04	0.10	
	A*3002	Q	T	K	E	L	Q	K	Q	I	I	K	I	Q	N	F	R	V	Y	Y	R	D	1.82E-04	0.05	R
	A*3001	S	D	I	K	V	V	P	R	R	K	V	K	I	I	K	D	Y	G	K	Q	M	6.97E-04	0.12	
	A*29(02)	I	K	D	Y	G	K	Q	M	A	G	A	D	C	V	A	V	R	Q	D	E	D	2.50E-03	0.18	R
	A*6802	A	G	A	D	C	V	A	G	R	Q	D	E	D									1.67E-03	0.17	R
Nef	A*6801	W	S	K	S	S	I	V	G	W	P	A	V	R	E	R	I	R	R	T	R	P	1.47E-04	0.07	
	A*32	W	S	K	S	S	I	V	G	W	P	A	V	R	E	R	I	R	R	T	R	P	1.17E-03	0.17	
	A*01(01)	G	A	L	T	S	S	N	T	A	H	N	N	A	D	C	A	W	L	Q	A	Q	9.59E-04	0.16	
	A*2902	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V	1.13E-03	0.19	R
	A*2402, B*40	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V #	0.06	R	7
	A*0301	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	K	E	1.17E-03	0.19	
	A*4301	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	K	E	K	G	G	L	4.53E-04	0.16	R
	A*01(01)	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	K	E	K	G	G	L	E	1.63E-04	0.08	
	A*2301	S	F	F	L	K	E	K	G	G	L	E	G	L	I	Y	S	K	K	R	Q	E	5.59E-04	0.12	
	A*3004	G	F	F	P	D	W	Q	N	Y	T	P	G	P	G	V	R	Y	P	L	T	F	3.06E-04	0.13	
	A*2402	Q	N	Y	T	P	G	P	G	V	R	Y	P	L	T	F	G	W	C	F	K	L	3.16E-10	0.00	R
	A*2301	V	R	Y	P	L	T	F	G	W	C	F	K	L	V	P	V	D	P	R	E	V	6.81E-12	0.00	
	A*2301	W	C	F	K	L	V	P	V	D	P	R	E	V	E	E	A	N	K	G	E	N	3.08E-04	0.13	
	A*23, B*08(01)	D	P	R	E	V	E	E	A	N	K	G	E	N	N	C	L	L	H	P	M	S	6.31E-04	0.18	R
	A*2301, B*4403	H	P	M	S	Q	H	G	M	E	D	E	E	R	E	V	L	K	W	K	F	D	1.69E-06	0.00	R
	A*66	E	D	E	E	R	E	V	L	K	W	K	F	D	S	S	L	A	R	R	H	L	5.21E-05	0.03	
	A*74	K	W	K	F	D	S	S	L	A	R	R	H	L	A	R	E	L	H	P	E	Y	4.89E-05	0.00	R

Table 3.4: Statistical associations between HLA-A alleles and HIV-1 polymorphisms in Gag, Pol and Nef proteins (q<0.2). Black bar shows site of polymorphism, (listed in order of amino acid position) with ten flanking amino acids (Durban consensus sequence). Known epitopes restricted by the relevant allele are highlighted in grey. R=reversion identified at this site. Where two HLA alleles are identified as operating at the same position and are transmitted in linkage disequilibrium (Appendix II), the HLA-A and HLA-B associations are listed on the same row in this table (separated by comma). If one of the linkage pair is known to restrict an epitope in the region of the polymorphism, the association is attributed to this allele and the linked molecule is listed in square brackets.

Protein	HLA	Polymorphism location	p value	q value	Reversion
p17 Gag	B*42 [Cw*1701]	K I R L R P G G K K **H** Y M L K H L V W A S	1.05E-04	0.07	R
	B*4201	R L R P G G K K H Y **M** L K H L V W A S R E	1.72E-04	0.17	
	B*45(01)	T V A T L Y C V H E **K** I E V R D T K E A L	1.43E-04	0.17	R
	B*41	A T L Y C V H E K I **E** V R D T K E A L D K	9.41E-05	0.05	
	B*44(03)	A A D K D A G A D S **G** K V S Q N Y	2.77E-04	0.17	R
p24 Gag	B*5801 [Cw*1604]	P I V **Q** N L Q G Q M V H Q A	1.45E-04	0.08	R
	B*5703	P I V Q N **L** Q G Q M V H Q A I S	2.98E-03	0.15	
	B*1510 [A*6802]	Q N L Q G Q M V H Q **A** I S P R T L N A W V	3.67E-04	0.05	R
	B*5702	Q N L Q G Q M V H Q **A** I S P R T L N A W V	2.46E-08	0.00	R
	B*5703 [C*0701]	Q N L Q G Q M V H Q **A** I S P R T L N A W V	2.51E-14	0.00	R
	B*4202	Q N L Q G Q M V H Q **A** I S P R T L N A W V	2.99E-03	0.16	
	B*1302	N L Q G Q M V H Q A **I** S P R T L N A W V K	1.33E-03	0.11	
	B*5702	N L Q G Q M V H Q A **I** S P R T L N A W V K	7.71E-05	0.01	R
	B*5703	N L Q G Q M V H Q A **I** S P R T L N A W V K	1.62E-08	0.00	R
	B*8101	P R T L N A W V K V **I** E E K A F S P E V I	2.76E-03	0.15	
	B*39(10)	N A W V K V I E E K **A** F S P E V I P M F T	5.42E-04	0.06	
	B*5703 [A*0202]	N A W V K V I E E K **A** F S P E V I P M F T	1.39E-15	0.00	
	B*5703	W V K V I E E K A F **S** P E V I P M F T A L	2.92E-04	0.04	
	B*57	V I E E K A F S P E **V** I P M F T A L S E G	7.92E-04	0.08	
	B*81(01) [A*0101]	E V I P M F T A L S **E** G A T P Q D L N T M	4.26E-07	0.00	R
	B*39(10)	E V I P M F T A L S **E** G A T P Q D L N T M	1.58E-03	0.12	R
	B*8101	F T A L S E G A T P **Q** D L N T M L N T V G	3.59E-05	0.01	R
	B*8101 [A*01(01)]	A L S E G A T P Q D **L** N T M L N T V G G H	1.90E-03	0.11	R
	B*8101	S E G A T P Q D L N **T** M L N T V G G H Q A	5.03E-07	0.00	R
	B*39(10), Cw*1203	S E G A T P Q D L N **T** M L N T V G G H Q A	7.48E-04	0.08	
	B*1503, Cw*02	I N E E A A E W D R **L** H P V H A G P I A P	2.10E-03	0.14	R
	B*1302	I N E E A A E W D R **L** H P V H A G P I A P	2.86E-03	0.15	
	B*1510	I N E E A A E W D R **L** H P V H A G P I A P	1.32E-03	0.13	R
	B*39(10)	V H A G P I A P G Q **M** R E P R G S D I A G	1.44E-03	0.13	
	B*35(01)	P I A P G Q M R E P **R** G S D I A G T T S T	5.64E-04	0.10	R
	B*5802	P I A P G Q M R E P **R** G S D I A G T T S T	2.71E-03	0.15	
	B*5702	R G S D I A G T T S **T** L Q E Q I A W M T S	4.27E-13	0.00	R
	B*5703	R G S D I A G T T S **T** L Q E Q I A W M T S	4.04E-23	0.00	R
	B*5801	R G S D I A G T T S **T** L Q E Q I A W M T S	1.88E-28	0.00	R
	B*5801	G S D I A G T T S T **L** Q E Q I A W M T S N	2.52E-03	0.15	
	B*5703	A G T T S T L Q E Q **I** A W M T S N P P I P	1.73E-05	0.00	R
	B*41	G T T S T L Q E Q I **A** W M T S N P P I P V	1.62E-03	0.12	
	B*53(01)	T L Q E Q I A W M T **S** N P P I P V G D I Y	2.01E-03	0.14	
	B*4901	E Q I A W M T S N P **P** I P V G D I Y K R W	1.63E-03	0.19	R
	B*57	Q I A W M T S N P P **I** P V G D I Y K R W I	2.95E-03	0.18	
	B*35(01)	M T S N P P I P V G **D** I Y K R W I I L G L	1.09E-07	0.00	
	B*5703	P V G D I Y K R W I **I** L G L N K I V R M Y	1.78E-03	0.11	R
	B*1401	P F R D Y V D R F F **K** T L R A E Q A T Q D	2.54E-11	0.00	R
	B*5801	F F K T L R A E Q A **T** Q D V K N W M T D T	3.73E-07	0.00	R
	B*4403	K T L R A E Q A T Q **D** V K N W M T D T L L	3.03E-09	0.00	
	B*0801	L L V Q N A N P D C **K** T I L R A L G P G A	1.52E-03	0.11	R
	B*3910	L V Q N A N P D C K **T** I L R A L G P G A T	1.36E-04	0.01	
	B*4501	L V Q N A N P D C K **T** I L R A L G P G A T	2.35E-03	0.15	
	B*0702	M T A C Q G V G G P **G** H K A R V L	6.72E-10	0.03	
	B*0705	M T A C Q G V G G P **G** H K A R V L	1.39E-03	0.17	
p15 Gag	B*4501	A E A M S Q A N N **T** S N N N I M M Q R S	1.69E-04	0.18	R
	B*3910	Q N R P E P R P E P **T** A P P A E N F R E S	2.03E-04	0.00	R

Table 3.5A: Statistical associations between HLA-B alleles and HIV-1 polymorphisms in Gag (q<0.2). Black bar shows site of polymorphism (listed in order of amino acid position), with ten flanking amino acids (Durban consensus sequence). Known epitopes restricted by the relevant allele are highlighted in grey. R=reversion identified at this site. Where two HLA alleles are identified as operating at the same position and are transmitted in linkage disequilibrium (Appendix II), the HLA-B and HLA-C associations are listed on the same row in this table (separated by comma). If one of the linkage pair is known to restrict an epitope in the region of the polymorphism, the association is attributed to this allele and the linked molecule is listed in brackets.

Protein	HLA	Polymorphism location	p value	q value	Reversion
Protease	B*44(03)	E N L A F P Q Q G G **E** A R E F P S E Q T R	2.84E-04	0.07	
	B*5301	L A F P Q Q G G E A **R** E F P S E Q T R A N	6.62E-06	0.00	
	B*18(01)	G E A R E F P S E Q **T** R A N S P T S R E L	6.49E-04	0.14	
	B*39(10)	Q I T L W Q R P L V **S** I K V G G Q I K E A	5.15E-04	0.12	
	B*4403	D T G A D D T V L E **E** I N L P G K W K P K	3.13E-08	0.00	
	B*44(03)	G A D D T V L E E I **N** L P G K W K P K M I	1.55E-04	0.04	
	B*44(03)	T V L E E I N L P G **K** W K P K M I G G I G	3.10E-04	0.08	
RT	B*8101	L G C T L N F P I S **P** I E T V P V K L K P	7.40E-10	0.00	R
	B*8101 [Cw*18]	C T L N F P I S P I **E** T V P V K L K P G M	1.53E-05	0.00	R
	B*35(01)	G D A Y F S V P L D **E** G F R K Y T A F T I	5.39E-07	0.00	
	B*8101	F R K Y T A F T I P **S** I N N E T P G I R Y	4.51E-04	0.19	R
	B*1801	T A F T I P S I N N **E** T P G I R Y Q Y N V	4.86E-05	0.05	
	B*0702	G W K G S P A I F Q **S** S M T K I L E P F R	1.93E-06	0.00	
	B*35(01)	L E P F R A K N P E **I** V I Y Q Y M D L Y	5.61E-04	0.15	
	B*4403	Q H R A K I E E L R **E** H L L K W G F T T P	9.00E-07	0.00	R
	B*5301	K W T V Q P I Q L P **E** K D S W T V N D I Q	3.95E-04	0.14	R
	B*4201 [Cw*1701]	G K L N W A S Q I Y **P** G I K V R Q L C K L	1.45E-07	0.00	R
	B*42	S Q I Y P G I K V R **Q** L C K L L R G A K A	4.48E-04	0.14	R
	B*4403	K D L I A E I Q K Q **G** H D Q W T Y Q I Y Q	2.45E-04	0.18	
	B*14, Cw*04	L K T G K Y A K M R **T** A H T N D V K Q L T	1.84E-04	0.10	R
	B*58	Q L T E A V Q K I A **L** E S I V I W G K T P	9.56E-04	0.19	
	B*4201	R L P I Q K E T W E **T** W W T D Y V Q A T W	2.73E-04	0.13	R
	B*5702	T D Y W Q A T W I P **E** W E F V N T P P L V	5.49E-05	0.05	
	B*4403	V D G A A N R E T K **I** G K A G Y V T D R G	5.27E-04	0.17	
	B*0801	K I G K A G Y V T D **R** G R Q K I V S L T E	4.79E-04	0.16	
	B*0801	V T D R G R Q K I V **S** L T E T T N Q K T E	8.48E-07	0.00	
	B*18(01)	T N Q K T E L Q A I **Q** L A L D S G S E V N	1.50E-05	0.00	
	B*4403	A I Q L A L Q D S G **S** E V N I V T D S Q Y	2.56E-09	0.00	R
	B*39(10) [Cw*1203]	A L Q D S G S E V N **I** V T D S Q Y A L G I	3.37E-05	0.05	R
	B*3910	A L G I I Q A Q P D **K** S E S E L V N Q I I	1.81E-05	0.02	R
	B*3910	Q A Q P D K S E S E **L** V N Q I I E Q L I K	1.63E-04	0.08	R
Integrase	B*4403	L F L D G I D K A Q **E** E H E K Y H S N W R	2.28E-05	0.00	R
	B*4403	F L D G I D K A Q E **E** H E K Y H S N W R A	6.04E-11	0.00	R
	B*4901	Y H S N W R A M A S **E** F N L P P I V A K E	2.30E-04	0.06	
	B*4201	A M A S E F N L P P **I** V A K E I V A S C D	4.65E-06	0.00	R
	B*4201 [Cw*1701]	S E F N L P P I V A **K** E I V A S C D K C Q	1.38E-06	0.00	R
	B*42(01) [Cw*1701]	E I V A S C D K C Q **L** K G E A I H G Q V D	3.95E-06	0.01	R
	B*1510	Q L D C T H L E G K **V** I L V A V H V A S G	1.64E-04	0.05	
	B*4403	S G Y I E A E V I P **A** E T G Q E T A Y Y I	3.01E-04	0.06	R
	B*5801 [A*02]	H T D N G S N F T S **A** A V K A A C W W A G	9.77E-04	0.09	R
	B*1503 [Cw*02]	V K A A C W W A G I **Q** Q E F G I P Y N P Q	2.30E-07	0.00	R
	B*1503 [Cw*02]	M A V F I H N F K R **K** G G I G G Y S A G E	2.07E-09	0.00	R
	B*14, Cw*0802	A G E R I I D I I A **T** D I Q T K E L Q K Q	2.54E-05	0.00	
	B*5802	E R I I D I I A T D **I** Q T K E L Q K Q I I	1.07E-03	0.17	R
	B*08(01)	Q T K E L Q K Q I I **K** I Q N F R V Y Y R D	1.28E-03	0.18	
	B*4201	V V P R R K V K I I **K** D Y G K Q M A G A D	4.64E-04	0.10	R
	B*15	R K V K I I K D Y G **K** Q M A G A D C V A G	7.98E-04	0.08	R

Table 3.5B: Statistical associations between HLA-B alleles and HIV-1 polymorphisms in Pol (q<0.2). Black bar shows site of polymorphism (listed in order of amino acid position), with ten flanking amino acids (Durban consensus sequence). Known epitopes restricted by the relevant allele are highlighted in grey. R=reversion identified at this site. Where two HLA alleles are identified as operating at the same position and are transmitted in linkage disequilibrium (Appendix II), the HLA-B and HLA-C associations are listed on the same row in this table (separated by comma). If one of the linkage pair is known to restrict an epitope in the region of the polymorphism, the association is attributed to this allele and the linked molecule is listed in square brackets.

Protein	HLA	Polymorphism location																				p value	q value	Reversion
Nef	B*45(01)	V	R	E	R	I	R	R	T	R	P	A	A	E	R	R	A	A	R	T	R	R	4.21E-04	0.11	
	B*5802	A	S	Q	D	L	D	K	H	G	A	L	T	S	S	N	T	A	H	N	N	A	9.13E-04	0.19	
	B*53(01)	K	H	G	A	L	T	S	S	N	T	A	H	N	N	A	D	C	A	W	L	Q	7.77E-04	0.19	R
	B*53(01)	H	G	A	L	T	S	S	N	T	A	H	N	N	A	D	C	A	W	L	Q	A	1.03E-03	0.17	
	B*5802 [A*43]	N	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	5.37E-04	0.12	R
	B*4403	N	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	1.13E-03	0.19	R
	B*40, A*2402	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V	2.57E-04	0.06	R
	B*4501	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V	R	P	Q	8.26E-09	0.00	R
	B*4403	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V	R	P	5.04E-06	0.07	
	B*07 [Cw*0702]	E	E	E	E	G	V	G	F	P	V	R	P	Q	V	P	L	R	P	M	T	Y	1.02E-09	0.00	R
	B*8101	V	G	F	P	V	R	P	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	1.78E-10	0.00	R
	B*35(01)	R	P	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	1.13E-07	0.00	R
	B*5801	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	K	E	7.32E-06	0.00	R
	B*57	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	K	E	4.11E-04	0.08	
	B*4403	A	A	F	D	L	S	F	F	L	K	E	K	G	G	L	E	G	L	I	Y	S	2.78E-07	0.00	R
	B*4403	K	E	K	G	G	L	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	1.89E-08	0.00	R
	B*0801	K	G	G	L	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	W	V	7.37E-04	0.18	
	B*18(01)	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	W	V	Y	H	T	Q	1.44E-09	0.00	R
	B*4403	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	W	V	Y	H	T	Q	1.82E-08	0.00	R
	B*4403	K	R	Q	E	I	L	D	L	W	V	Y	H	T	Q	G	F	F	P	D	W	Q	1.05E-03	0.19	R
	B*35(01)	D	W	Q	N	Y	T	P	G	P	G	V	R	Y	P	L	T	F	G	W	C	F	6.46E-06	0.00	R
	B*53(01)	D	W	Q	N	Y	T	P	G	P	G	V	R	Y	P	L	T	F	G	W	C	F	1.71E-04	0.02	
	B*13	P	V	D	P	R	E	V	E	E	A	N	K	G	E	N	N	C	L	L	H	P	1.43E-03	0.18	
	B*4201	K	G	E	N	N	C	L	L	H	P	M	S	Q	H	G	M	E	D	E	E	R	3.79E-04	0.09	R
	B*4201	C	L	L	H	P	M	S	Q	H	G	M	E	D	E	E	R	E	V	L	K	W	7.07E-04	0.16	
	B*4201	R	H	L	A	R	E	L	H	P	E	Y	Y	K	D	C							9.96E-04	0.19	

Table 3.5C: Statistical associations between HLA-B alleles and HIV-1 polymorphisms in Nef (q<0.2). Black bar shows site of polymorphism (listed in order of amino acid position), with ten flanking amino acids (Durban consensus sequence). Known epitopes restricted by the relevant allele are highlighted in grey. R=reversion identified at this site. Where two HLA alleles are identified as operating at the same position and are transmitted in linkage disequilibrium (Appendix II), the HLA-B and HLA-C associations are listed on the same row in this table (separated by comma). If one of the linkage pair is known to restrict an epitope in the region of the polymorphism, the association is attributed to this allele and the linked molecule is listed in square brackets.

Protein	HLA	Polymorphism location																			p value	q value	Reversion
p24 Gag	Cw*0304	F	T	A	L	S	E	G	A	T	P	Q	D	L	N	T	M	L	N	T	V	G	2.16E-03	0.14	R
	Cw*17(01)	F	T	A	L	S	E	G	A	T	P	Q	D	L	N	T	M	L	N	T	V	G	2.69E-03	0.18	
	Cw*1601	L	K	D	T	I	N	E	E	A	A	E	W	D	R	L	H	P	V	H	A	G	9.55E-04	0.09	R
	Cw*0501	K	D	T	I	N	E	E	A	A	E	W	D	R	L	H	P	V	H	A	G	P	1.62E-03	0.11	
	Cw*0704	A	A	E	W	D	R	L	H	P	V	H	A	G	P	I	A	P	G	Q	M	R	3.48E-04	0.04	R
	Cw*1701	P	I	A	P	G	Q	M	R	E	P	R	G	S	D	I	A	G	T	T	S	T	8.99E-04	0.17	R
	Cw*1604	R	G	S	D	I	A	G	T	T	S	T	L	Q	E	Q	I	A	W	M	T	S	5.89E-05	0.01	
	Cw*1604	I	R	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	F	K	T	L	R	3.15E-03	0.15	
	Cw*0304	F	R	D	Y	V	D	R	F	F	K	T	L	R	A	E	Q	A	T	Q	D	V	2.95E-04	0.04	R
	Cw*0804	A	N	P	D	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	9.33E-04	0.09	
	Cw*0801	A	N	P	D	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	1.92E-03	0.11	
	Cw*03	D	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	M	T	A	9.68E-05	0.02	R
	Cw*0804	D	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	M	T	A	8.46E-04	0.09	
	Cw*0801 [B*1401, B*1510]	D	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	M	T	A	9.07E-04	0.09	R
	Cw*0704	C	K	T	I	L	R	A	L	G	P	G	A	T	L	E	E	M	M	T	A	C	2.51E-03	0.15	
Protease	Cw*06						F	F	R	E	N	L	A	F	P	Q	Q	G	G	E	A		1.10E-03	0.19	
	Cw*16	E	Q	T	R	A	N	S	P	T	S	R	E	L	E	R	E	Q	R	R	E	P	4.38E-04	0.10	R
	Cw*1801	F	I	K	V	R	Q	Y	D	Q	I	L	I	E	I	C	G	K	K	A	I	G	4.35E-05	0.02	
RT	Cw*04(01)	D	A	Y	F	S	V	P	L	D	E	G	F	R	K	Y	T	A	F	T	I	P	1.46E-04	0.03	
	Cw*08	K	I	E	E	L	R	E	H	L	L	K	W	G	F	T	T	P	D	K	K	H	8.69E-04	0.18	
	Cw*06	L	R	E	H	L	L	L	K	W	G	F	T	T	P	D	K	K	H	Q	K	E	7.51E-04	0.19	R
	Cw*0602	N	L	K	T	G	K	Y	A	K	M	R	T	A	H	T	N	D	V	K	Q	L	3.80E-04	0.14	
	Cw*1203	K	Q	L	T	E	A	V	Q	K	I	A	L	E	S	I	V	I	W	G	K	T	5.37E-06	0.00	R
	Cw*0302	Q	K	E	T	W	E	T	W	W	T	D	Y	W	Q	A	T	W	I	P	E	W	3.74E-05	0.05	R
	Cw*04(01)	L	W	Y	Q	L	E	K	E	P	I	A	G	A	E	T	F	Y	V	D	G	A	1.79E-05	0.00	
	Cw*0302	R	G	R	Q	K	I	V	S	L	T	E	T	T	N	Q	K	T	E	L	Q	A	2.79E-04	0.13	R
	Cw*0304	Q	A	Q	P	D	K	S	E	S	E	L	V	N	Q	I	I	E	Q	L	I	K	5.83E-04	0.17	
Nef	Cw*0302	A	D	C	A	W	L	Q	A	Q	E	E	E	E	E	E	G	V	G	F	P	V	8.47E-04	0.19	R
	Cw*0702	R	P	Q	V	P	L	R	P	M	T	Y	K	A	A	F	D	L	S	F	F	L	7.01E-04	0.18	
	Cw*0701	G	G	L	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	W	V	Y	7.85E-07	0.00	R
	Cw*08	E	G	L	I	Y	S	K	K	R	Q	E	I	L	D	L	W	V	Y	H	T	Q	8.42E-04	0.15	
	Cw*0802	V	P	V	D	P	R	E	V	E	E	A	N	K	G	E	N	N	C	L	L	H	4.05E-04	0.13	
	Cw*0804	G	M	E	D	E	E	R	E	V	L	K	W	K	F	D	S	S	L	A	R	R	5.82E-04	0.16	
	Cw*0602	E	D	E	E	R	E	V	L	K	W	K	F	D	S	S	L	A	R	R	H	L	2.92E-04	0.08	
	Cw*1601	E	D	E	E	R	E	V	L	K	W	K	F	D	S	S	L	A	R	R	H	L	7.70E-07	0.00	
	Cw*1601	L	A	R	R	H	L	A	R	E	L	H	P	E	Y	Y	K	D	C				3.74E-04	0.09	R

Table 3.6: Statistical associations between HLA-Cw alleles and HIV-1 polymorphisms in Gag, Pol and Nef (q<0.2). Black bar shows site of polymorphism (listed in order of amino acid position), with ten flanking amino acids (Durban consensus sequence). Known epitopes restricted by the relevant allele are highlighted in grey. R=reversion identified at this site. Where two HLA alleles are identified as operating at the same position and are transmitted in linkage disequilibrium (Appendix II), the HLA-B and HLA-C associations are listed on the same row in this table (separated by comma). If one of the linkage pair is known to restrict an epitope in the region of the polymorphism, the association is attributed to this allele and the linked molecule is listed in square brackets.

3.3.3 Known epitopes containing HLA-associated polymorphisms are predominantly restricted by HLA-B

Of the 205 known epitopes identified in association with HLA-polymorphisms, the majority (58%) were HLA-B restricted. Fig. 3.7 summarises the data grouped by protein and by HLA-A, HLA-B and HLA-C alleles. Reverting mutations accounted for ≥50% of all polymorphisms in each protein, irrespective of the q value threshold.

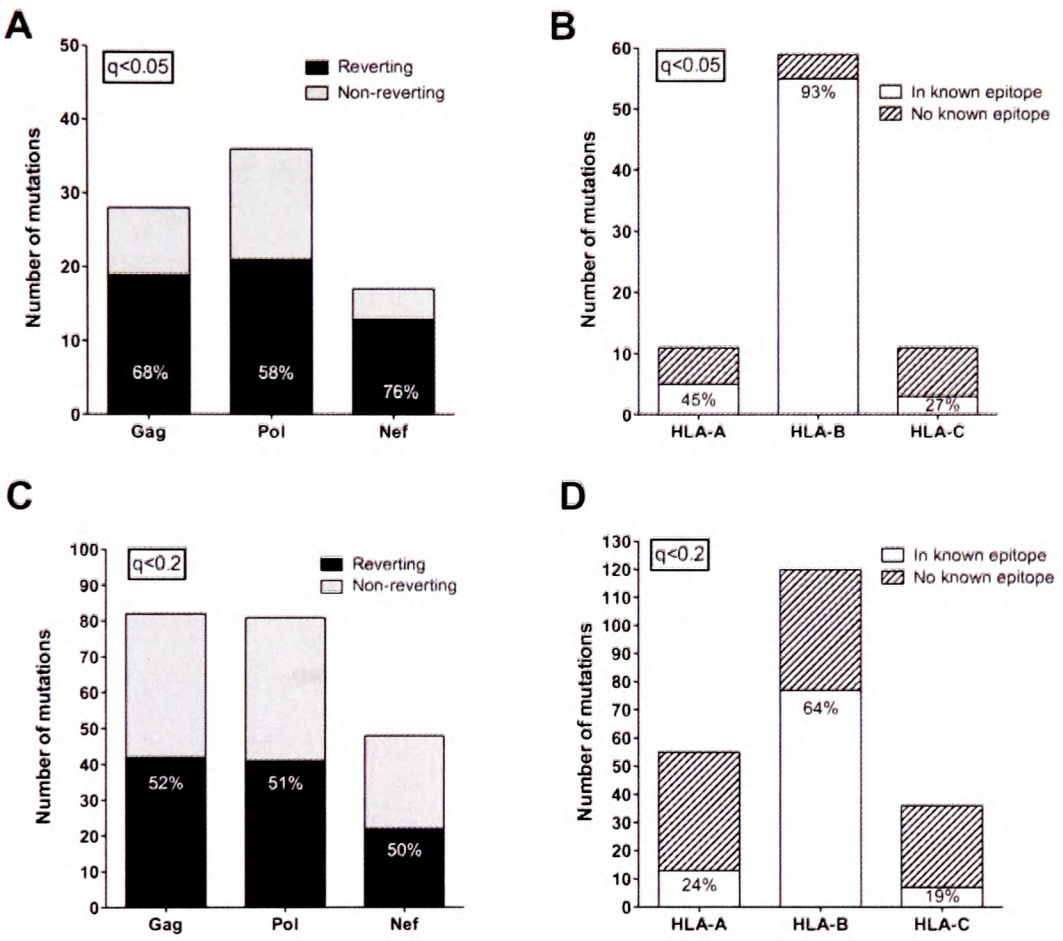


Figure 3.7: Proportion of HIV polymorphisms associated with host HLA according to the presence/absence of reversion (A, C) and the presence/absence of a known epitope of the relevant restriction (B, D). Associations are shown by protein and by HLA type, with q value (FDR) <0.05 (A, B) and q value <0. 2 (C,D).

The proportion of polymorphisms occurring in, or flanking, known epitopes was greatest for HLA-B (93% at q<0.05), but this may reflect – at least in part - a bias towards identification and publication of epitopes restricted by HLA-B.

3.3.4 An increased number of HLA-associated polymorphisms correlates with lower median viral load

A correlation was next sought between the number of HLA-associated polymorphisms and median VL for each of the different alleles operating in Gag, Pol and Nef. As anticipated, the number of HLA-B-associated polymorphisms in Gag was strongly associated with reducing VL ($q<0.05$: $r=-0.56$, $p=0.003$; Fig. 3.8A). The proportion of polymorphisms identified within known epitopes restricted by the relevant HLA allele was reduced by increasing the number of false-positives included to 20% ($q<0.2$; Fig 3.7D), but the correlation remained unaffected ($r=-0.51$, $p=0.0096$; Fig. 3.8B).

No such association was identified between the number of HLA-B polymorphisms in Pol or Nef and median VL for each allele (Table 3.7).

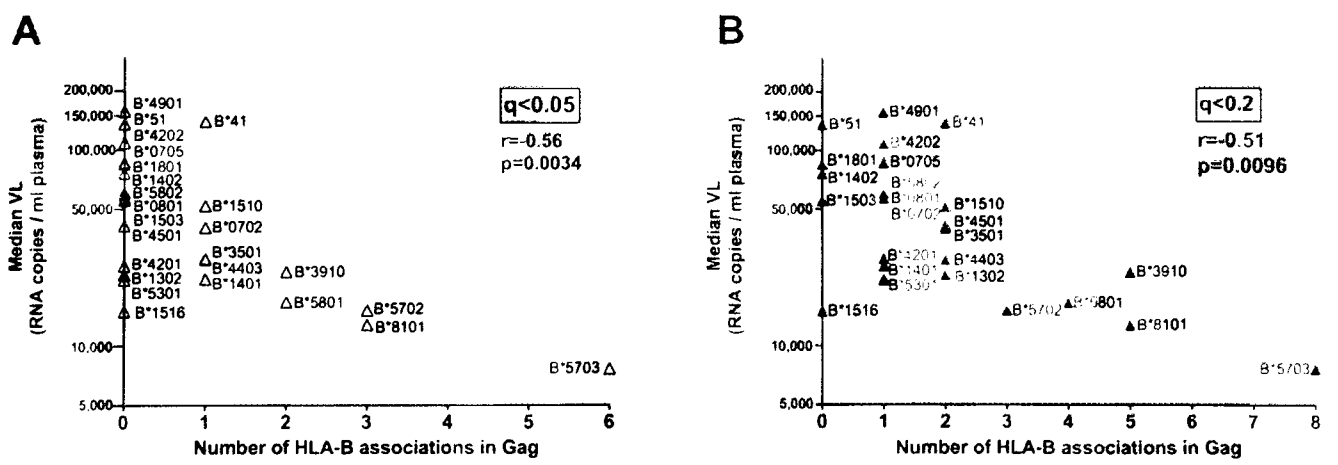


Figure 3.8: Relationship between number of HLA-B-associated polymorphisms in Gag and median viral load (VL) for subjects with each allele. A: $q<0.05$, B: $q<0.2$.

Protein	HLA Class I Locus	All HLA associations r, (p)	Reverting associations only r, (p)	Non-reverting associations only r, (p)
Gag	HLA-A	-0.47, (0.03)	-0.35, (0.098)	-0.29, (0.18)
	HLA-B	-0.56, (0.0034)	-0.62, (0.0009)	-0.14, (0.51)
	HLA-Cw	0.087, (0.75)	0.33, (0.21)	-0.31, (0.25)
Pol	HLA-A	0, (1)	0.23, (0.3)	-0.16, (0.46)
	HLA-B	-0.29, (0.17)	-0.30, (0.14)	-0.05, (0.83)
	HLA-Cw	-0.67, (0.0047)	-0.57, (0.02)	-0.32 (0.22)
Nef	HLA-A	-0.25, (0.24)	-0.19, (0.4)	-0.16, (0.46)
	HLA-B	-0.15, (0.48)	-0.07, (0.74)	-0.20, (0.34)
	HLA-Cw	0.21, (0.45)	-0.03, (0.92)	0.31, (0.25)

Table 3.7: Relationship between number of polymorphisms in Gag, Pol and Nef selected by HLA-A, B and C alleles and median viral load per allele (q<0.05). r and p values calculated by linear regression. Associations reaching statistical significance highlighted in red, before correction for multiple comparisons.

The dominant role of HLA-B*57/5801 alleles in mediating this effect is shown by the reduction in statistical significance when these alleles were removed from analysis (for HLA-B associations in Gag at $q<0.05$, $r=-0.37$ and $p=0.09$).

Unexpectedly, a strong correlation was also observed between the number of polymorphisms within Pol selected by HLA-C alleles and median VL for each allele, that was again largely unaffected by choice of q value ($q<0.05$: $r=-0.67$, $p=0.005$; $q<0.2$: $r=-0.54$, $p=0.03$; Fig. 3.9).

Although reaching statistical significance, potential correlations between polymorphisms selected by HLA-A alleles and median VL for the allele were more difficult to assess, since these were so few at low q values (Fig. 3.7B, Fig. 3.9C).

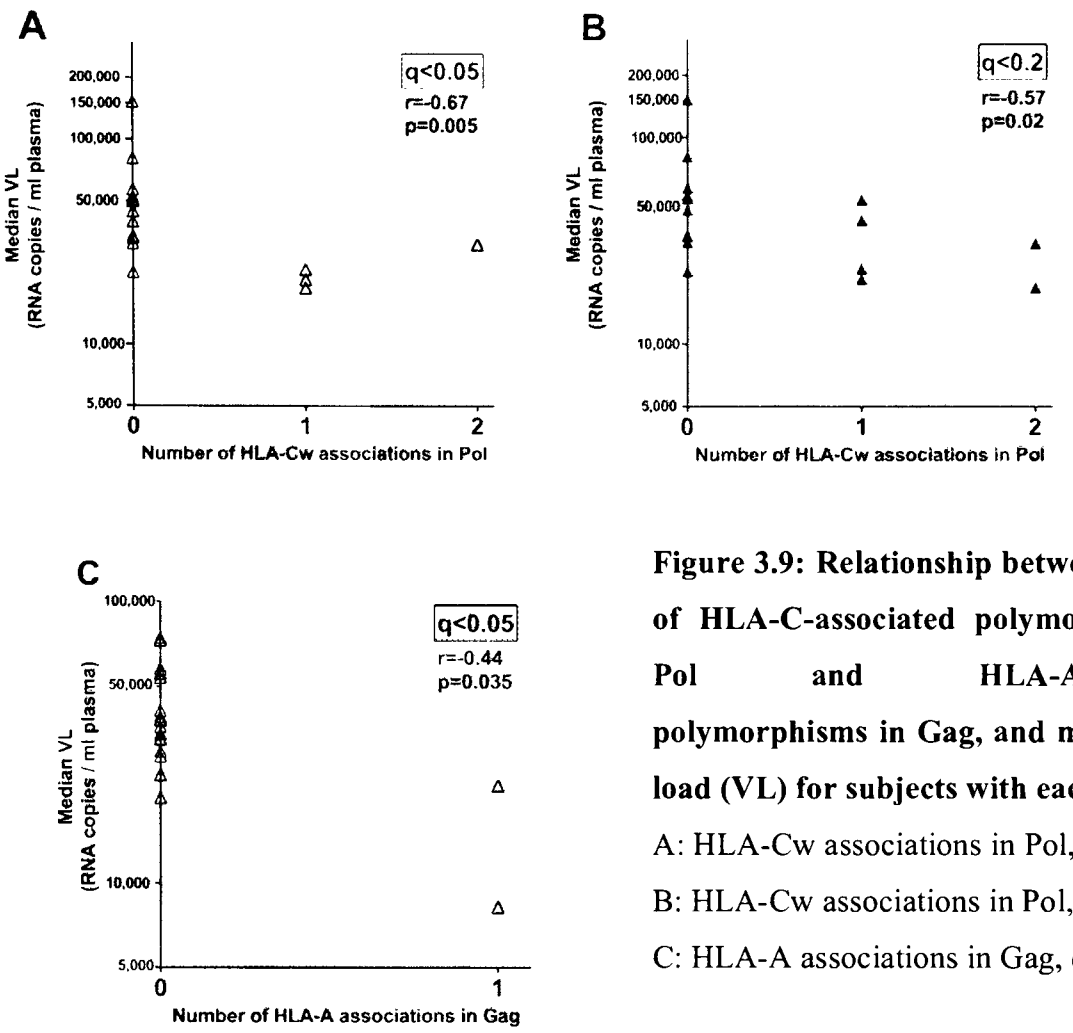


Figure 3.9: Relationship between number of HLA-C-associated polymorphisms in Pol and HLA-A-associated polymorphisms in Gag, and median viral load (VL) for subjects with each allele.

A: HLA-Cw associations in Pol, q<0.05;

B: HLA-Cw associations in Pol, q<0.2;

C: HLA-A associations in Gag, q<0.05.

3.3.5 Correlation between HLA-B polymorphisms in Gag and viral load persists with reduced number of Gag sequences

In order to establish whether the associations observed in Gag were, in part, a consequence of inclusion of more sequences for Gag than for other proteins, the pool of Gag sequences was reduced to 430 and lineage-corrected analysis was repeated. The significant association between number of HLA-B responses and median viral set point was maintained (q<0.2: r=-0.59, p=0.002; q<0.05: r=-0.63, p=0.0007, Fig. 3.10), showing a consistent overall effect in Gag irrespective of reduction in sequence numbers.

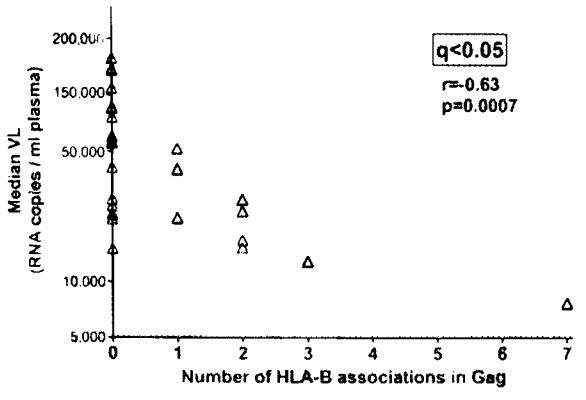


Figure 3.10: Relationship between number of polymorphisms in Gag and median viral load (VL) in a reduced cohort of 430 subjects.

3.3.6 Number of polymorphisms that revert following transmission correlates with reducing viral load

Previous studies have suggested that the selection of escape mutations that reduce viral replicative capacity, although allowing escape from the particular $CD8^+$ T cell specificity, may contribute to subsequent control through the remaining immune responses (Kelleher *et al.* 2001; Friedrich *et al.* 2004; Leslie *et al.* 2004; Brockman *et al.* 2007; Crawford *et al.* 2007; Schneidewind *et al.* 2007). Therefore, the relationship between selection of reverting polymorphisms and VL was examined.

There was a strong correlation between the median VL for different HLA-B alleles and the number of reverting HLA-B-associated polymorphisms in Gag (for associations at $q<0.05$, $r=-0.62$, $p=0.0009$, Fig. 3.11A) but not with non-reverting HLA-B polymorphisms in Gag ($r=-0.14$, $p=0.51$, Fig. 3.11C). This relationship retained borderline statistical significance even in the absence of HLA-B*57/5801 alleles ($r=-0.44$, $p=0.04$). The same relationship was also confirmed for associations with $q<0.2$ (Figs 3.11B, D). Within Pol, there was a difference between reverting and non-reverting HLA-B-associated polymorphisms (for associations at $q<0.05$, $r=-0.30$, $p=0.14$ and $r=0.05$, $p=0.83$, respectively) and between reverting and non-reverting

HLA-C-restricted polymorphisms (r=-0.57, p=0.020 and r=-0.32, p=0.22, respectively; Table 3.7).

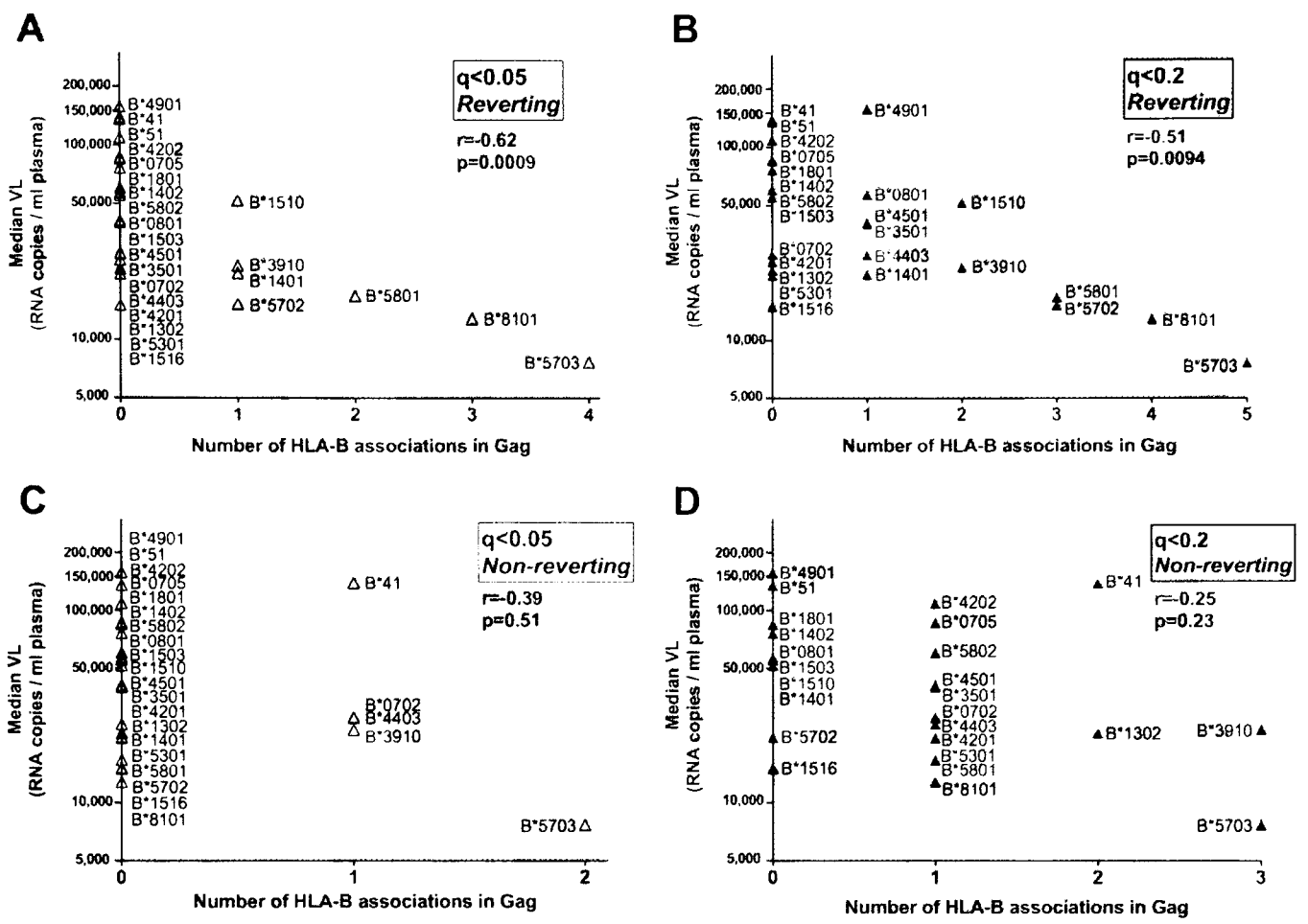


Figure 3.11: Relationship between the number of HLA-B-associated polymorphisms and the median viral load (VL) for each allele, according to the presence or absence of reversion. The number of sites of HIV polymorphism associated with HLA-B alleles is plotted against the median VL for patients with that allele, at sites where reversion was identified (A, B) and where no reversion was identified (C, D), for associations with $q<0.05$ (A, C) and $q<0.2$ (B, D).

For HLA-A-associated polymorphisms, the numbers of polymorphisms were too few to justify extensive analysis, but even here there is an indication that reverting mutations are contributing to HLA-associated immune control (Table 3.7). These data support the hypothesis that the escape mutations that reduce viral fitness are those most strongly associated with control of viraemia.

3.3.7 An increased number of HLA-B associated polymorphisms in Gag correlates with higher median CD4$^+$ T cell count

The analysis was repeated using median CD4$^+$ T cell count rather than VL for each allele. Again, there was a significant correlation between number of HLA-B associated polymorphisms (at $q<0.2$ and $q<0.05$) and median CD4 count in Gag (Fig. 3.12A) that was more strongly significant for reverting polymorphisms ($r=0.65$, $p=0.0005$ at $q<0.2$, Fig. 3.12B), but not for non-reverting associations ($r=0.23$, $p=0.28$). As for VL, there was also an association between HLA-C associated polymorphisms in Pol and median CD4 count, although this only reached significance at $q<0.05$ ($r=0.65$, $p=0.006$, Fig. 3.12C).

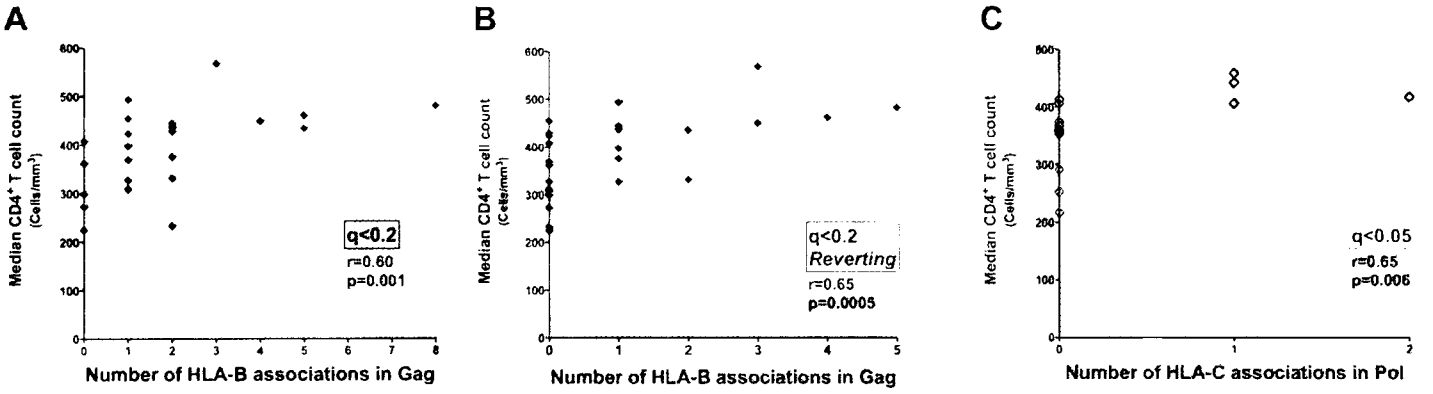


Figure 3.12: Relationship between the number of HLA-B-associated polymorphisms and the median CD4$^+$ T cell count for each allele. The number of sites of HIV polymorphism associated with HLA-B alleles is plotted against the median CD4 count for patients with that allele. A: All HLA-B associations in Gag ($q<0.2$). B: HLA-B reverting associations in Gag ($q<0.2$). C: All HLA-C associations in Pol ($q<0.05$).

3.3.8 Identification of 45 new HLA-associations with HIV peptide responses using IFN-γ ELISpot assays

In order to address the question of whether the association of particular HLA class I alleles with characteristic levels of viraemia relates to the number of Gag-specific

epitopes that are targeted in chronic infection, the HIV-specific $CD8^-$ T cell responses were characterised by testing recognition in ELISpot assays. Previously, 160 dominant $CD8^+$ T cell responses and HLA restrictions were defined from a study cohort of 578 subjects (Kiepiela *et al.* 2007). This previous analysis was repeated for the extended cohort of 681 subjects described here. HLA restriction of the ELISpot response was determined by statistical association, using a Bonferroni correction. Using this approach, an additional 45 significant (in each case, $p<0.00075$) new associations between HLA class I molecules and peptide recognition, were identified (Table 3.8).

Protein	HLA	OLP sequence	Optimal	p value
p17 Gag	B*0801	EKIRLRPGGKKHYMLKHL	GGKKKYKL	9.00E-05
p24 Gag	B*0702	AAEWDRLHPVHAGPIA	HPVHAGPIA	2.30E-04
	A*6802	SDIAGTTSTLQEQIAWM	not defined	8.00E-05
	Cw*0303	GPKEPFRDYVDRFFKTLR	YVDRFFKTL	3.00E-04
	Cw*0801	TILRALGPGASLEEMMTA	not defined	2.00E-05
	A*0205	ACQGVGGPSHKARVLAEA	VLAEAMSQV	1.40E-04
p15 Gag	B*4201	PKQEPKDREPLTSLK	not defined	5.10E-04
Protease	B*5301	FFREDLAFPQGEAREF	not defined	3.70E-05
RT	B*4202	PSINNETPGIRYQYNVL	not defined	2.00E-04
	B*1402	PSINNETPGIRYQYNVL	IRYQYNVL	5.00E-04
	B*3910	QGWKGSPAIFQSSMTKIL	LPQGWKGSPAI	6.00E-04
	B*8101	QGWKGSPAIFQSSMTKIL	SPAIFQSSM	7.40E-04
	B*0702	QGWKGSPAIFQSSMTKIL	SPAIFQSSM	2.00E-06
	B*1801	VKQLTEAVQKIAMESIVI	not defined	1.10E-04
	B*3910	QKIAMESIVIWGKTPKFR	not defined	1.00E-04
	B*5301	FRLPIQKETWETWWTDYW	not defined	2.00E-06
	A*3201	FRLPIQKETWETWWTDYW	PIQKETWETW	1.00E-04
	A*2402	FRLPIQKETWETWWTDYW	not defined	2.70E-08
	B*3501	QLEKEPIAGAETFYVDGA	EPIVGAETF	1.00E-06
	B*4501	QLEKEPIAGAETFYVDGA	GAETFYVDGA	4.00E-05
	B*5702	ETKIGKAGYVTDRGRQKI	not defined	6.00E-04
	B*4403	LALQDSGSEVNIVTDSQY	not defined	3.10E-09
Integrase	B*3501	GYIEAEVIPAETGQETAY	not defined	1.40E-08
	Cw*0802	DIIATDIQTKELQKQIIK	not defined	9.60E-08
Vif	A*0202	GLADQLIHMHYFDCF	DLADQLIHLY	2.00E-06
	A*6801	SLQYLALTALIKPKKIK	not defined	8.00E-05
Vpr	B*4202	ELKQEAVRHFPRPWLHGL	FPRPWLHGL	8.00E-05
	A*6801	GLGQYIYETYGDTWTGV	DTWAGVEAIR	2.00E-04
Tat	A*6601	MEPVDPNLEPWNHPGSQPK	not defined	1.00E-05
Rev	B*3501	DEALLQAVRIIKILY	not defined	1.00E-05
	B*1801	DEALLQAVRIIKILY	not defined	1.00E-04
	B*3910	DEALLQAVRIIKILY	not defined	3.00E-04
gp120	Cw*0401	KVSFDPIPIHYCAPAGYA	not defined	7.10E-04
	B*3910	KPVVSTQLLLNGSLA	not defined	7.10E-04
	B*4201	VCTRPNNNTRKSIRI	RPNNNTRKSI	3.00E-04
	Cw*0602	AGNITCKSNITGLLLTR	not defined	5.00E-04
gp41	B*5802	AASITLTVQARQLLSGIV	not defined	4.00E-04
	A*0201	ELKKSAVSLLDTIAIAVA	SLLNATDIAV	5.00E-05
Nef	B*4501	RRTEPAAEGVGAASQDL	not defined	3.70E-04
	B*4201	GALTSSNTDTTNADCAWL	not defined	3.10E-04
	B*4501	WLRAQEEEEEVGFPVRPQV	GEVGFPVRPQV	2.00E-04
	A*6802	QVPLRPMTFKGAFDLSFF	not defined	3.00E-06
	A*3402	KKRQEILDLWVYHTQGYF	not defined	3.00E-04
	B*5801	LWVYHTQGYFPDWQNY	HTQGYFPDW	3.90E-04
	B*4202	NYTPGPGVRYPLTFGWCF	TPGPGVRYPL	3.00E-04

Table 3.8: HLA associations with ELISpot responses based on OLP screening of 681 South African subjects.

As described previously in the n=578 subset of the cohort, the most immunogenic proteins identified here were Gag (42/205, 20% of epitopes targeted), Pol (83/205, 41%), and Nef (22/205, 11%); 58% of responses were HLA-B-restricted, 27% HLA-A-restricted and 13% HLA-C-restricted.

3.3.9 Number of HLA-B-restricted Gag epitopes targeted is correlated with reducing viral load.

Since HLA-B allele expression has the strongest HLA-mediated influence on viral setpoint (Kiepiela *et al.* 2004), and the breadth of only Gag-specific $CD8^+$ T cell responses is associated with decreasing VL (Kiepiela *et al.* 2007), a correlation was sought between median VL associated with each HLA-B allele and the number of statistically significant ELISpot responses made by subjects with that HLA-B allele.

A significant association was observed between Gag ELISpot responses and median VL for HLA-B alleles (r=-0.49, p=0.013, Fig. 3.13A, Table 3.9). Even excluding the alleles most strongly associated with control of viraemia in this cohort, HLA-B*57 and B*5801, the median VL for each HLA-B allele remained correlated significantly with the number of Gag epitopes presented by that HLA-B allele (r=-0.50, p=0.018, Fig. 3.13B). In contrast, no significant associations were observed for the number of Pol or Nef epitopes presented by HLA-B (Fig. 3.13C, D), or for epitopes of any protein presented by HLA-A or C alleles (Table 3.9).

These data support earlier findings that, of the HLA class I alleles expressed, HLA-B alleles have the strongest impact on VL (Kiepiela *et al.* 2004), and that Gag-specific responses in chronic infection are those most strongly associated with immune control of HIV (Kiepiela *et al.* 2007). However, the stronger correlations described above in relation to the number of reverting Gag and Pol polymorphisms associated with different HLA molecules suggest that non-HLA-B alleles and Pol-specific responses can also make important contributions to immune control of HIV.

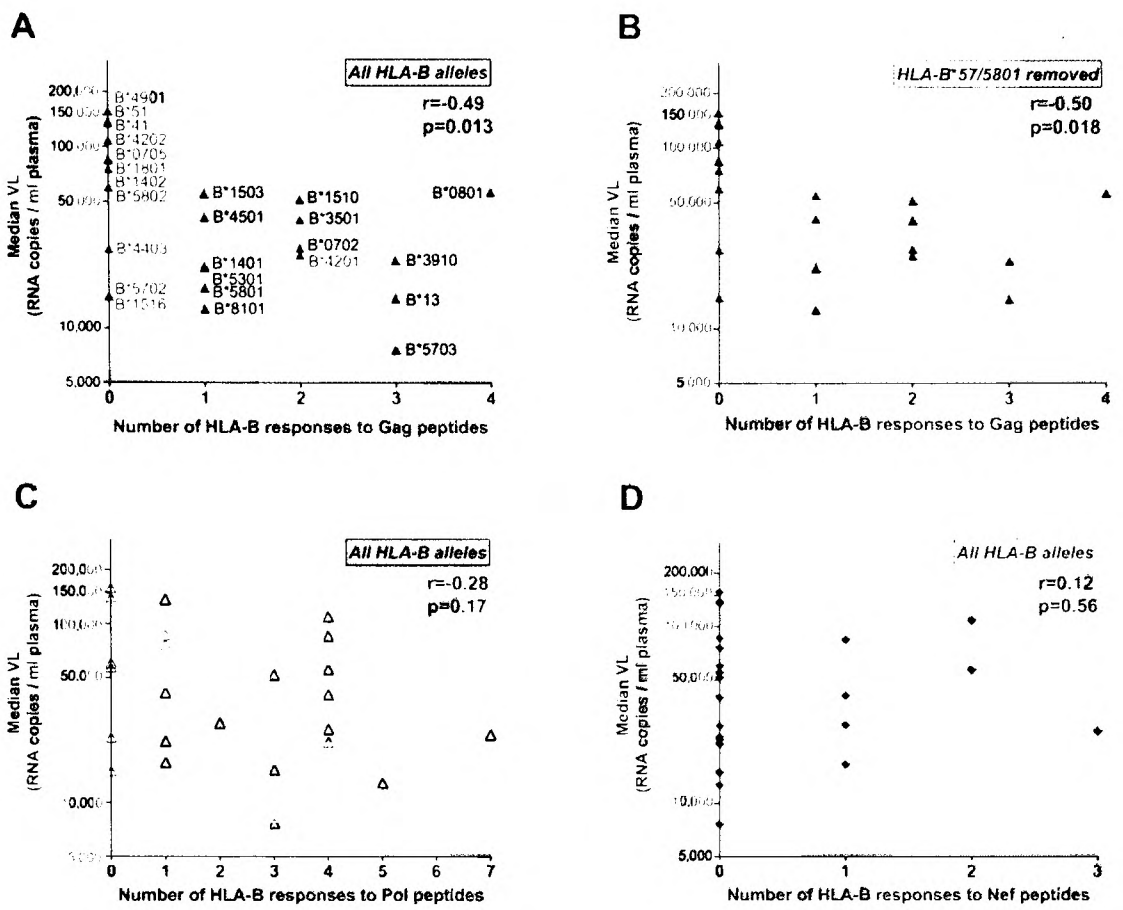


Figure 3.13: Relationship between number of CD8$^+$ T cell ELISpot responses and median viral load (VL) by allele for 681 HIV-infected individuals. Number of ELISpot responses per allele (significant response defined as $p<0.00075$) is plotted against median VL for patients expressing that allele. A: HLA-B-associated responses to peptides in Gag. B: HLA-B analysis in Gag repeated following removal of favourable alleles HLA-B*5702, HLA-B*5703 and HLA-B*5801. C: HLA-B-associated responses to peptides in Pol. D: HLA-B-associated responses to peptides in Nef.

	HLA-A	HLA-B	HLA-C
	r, (p)	r, (p)	r, (p)
Gag	0.12, (0.58)	-0.49, (0.013)	0.05, (0.84)
Pol	0.16 (0.47)	-0.28, (0.17)	0.13, (0.62)
Nef	0.01, (0.94)	0.12, (0.56)	-0.24, (0.36)

Table 3.9: Relationship between number of CD8$^+$ T cell ELISpot responses and median viral load by allele for HLA-A, B and C in Gag, Pol and Nef. r and p values by Spearman rank correlation.

3.3.10 Sites of HLA selection are associated with increased variability

In order to explore the relationship between HLA selection and amino acid variability, the Shannon entropy score was calculated for each residue in Gag, Pol and Nef. Sites at which escape only was detected were statistically the most variable, while sites of no HLA selection were the most conserved (Fig. 3.14).

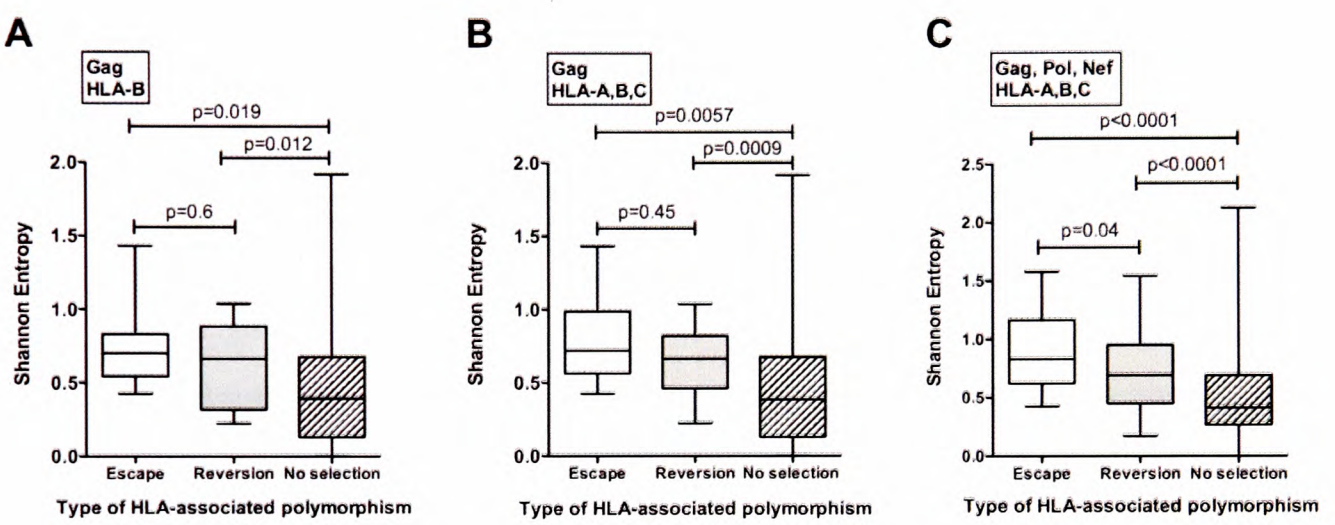


Figure 3.14: Relationship between HLA selection and amino acid variability. Analysis for associations with $q<0.05$. Boxes show median + 25th and 75th percentiles; whiskers show minimum-maximum. A: Sites of HLA-B-selection in Gag (p17+p24+p15). B: Sites of HLA-A, B and C selection in Gag (p17+p24+p15). C: Sites of HLA-A, B and C selection in Gag, Pol and Nef. p values by Mann-Whitney test.

Sites that revert following transmission, due to the pressure for conservation of wild-type, are still more variable than sites of no selection but tend to be less variable than sites at which reversion does not occur. Additional analysis was undertaken to assess the impact of different numbers of sequences used to compute entropy (Appendix III), and the overall relationship between sites of HLA-selection and entropy scores is further addressed in Chapter 4 (see Table 4.2).

Overall, these entropy data point to a possible role of HLA-selection in driving HIV sequence variability, underlining the evolutionary consequences of immunological selection pressure across populations. This is consistent with findings in previously published work, in which sites of HLA selection correlate with highly variable sites (Allen *et al.* 2005). However, conclusions about HLA as the cause of such variation should be drawn with caution, as sites that are the most variable also have the greatest statistical power for the detection of HLA-associations, and may have the least fitness cost associated with mutations.

3.4 DISCUSSION

3.4.1 Relationship between selection of reverting mutations and control of viraemia

Overall, these data demonstrate that the $CD8^+$ T cell responses driving immune control are those that select escape mutations that impose a fitness cost upon the virus, and that therefore revert post-transmission. Specifically, HLA class I associations with viral setpoint that have previously been described are here shown to hinge on the number of reverting mutations selected in Gag and Pol by HLA-B and HLA-C alleles.

Impact of HLA-B-mediated selection on viraemic control

The correlation between lowered viraemia and increased selection of reverting polymorphisms is most clear-cut for HLA-B alleles. This association also holds true when using $CD4^+$ T cell count as a marker of disease progression rather than VL. Broad Gag-specific $CD8^+$ T cell responses are associated with good immune control (Kiepiela *et al.* 2007) and HLA-B effects have previously been shown to dominate HLA-mediated influence on disease outcome in HIV infection (Kiepiela *et al.* 2004).

The data presented here support both of these findings, but also further develop our understanding of the mechanism of these effects by highlighting the central role of mutations that impose a fitness cost. The HLA-B associations with viral setpoint are significantly correlated with the number of Gag epitopes targeted, even if B*57 and B*5801 are excluded, showing that this finding is not only mediated by the two alleles known to have the strongest effect on VL.

Impact of HLA-A and HLA-Cw-mediated selection on viraemic control

In contrast to previous studies, these data also highlight a potential role for mutations selected by HLA-C in Pol. Again, reverting mutations drive the statistical association, highlighting the importance of mutations that impose a cost to viral fitness. The role of HLA-C has historically been liable to being underestimated, partly because the limitations of serological HLA typing methods affected HLA-C disproportionately, and partly because of the many strong linkage disequilibrium effects between HLA-B and HLA-C. This analysis has the advantage of high resolution HLA typing and robust correction for LD, and is consistent with the analysis of another large cohort that suggested HLA-C-mediated effects (Fellay *et al.* 2007). However, it is not possible to eliminate altogether the potential for linkage-mediated effects, such that associations with HLA-C alleles are in fact driven through linked HLA-B molecules.

The observation that there may be a role for HLA-C-restricted targeting of Pol is unexpected, given the previous lack of evidence that HLA-C-restricted responses were associated with improved immune control of HIV, even when Gag was targeted (Kiepiela *et al.* 2007). It is possible that, given the small number of reverting HLA-C-associated mutations, the beneficial effect of these responses may be outweighed by

the larger number of HLA-C-restricted responses that tend not to drive selection pressure on the virus. In contrast, the proportion of mutations associated with HLA-B in Gag, Pol and Nef (58 of 76, or 76%) is somewhat higher than the proportion of $CD8^+$ T cell (ELISpot) responses in these proteins that are associated with HLA-B (84 of 150, or 63%; p=0.05, Fisher's Exact test).

Even more striking is the observation of relatively few HLA-A-associated mutations (8 of 76, or 11%) for the number of HLA-A-restricted $CD8^+$ T cell responses to Gag, Pol and Nef (41 of 150, or 27%; p=0.004, Fisher's Exact test). These comparisons suggest that HLA-B-restricted $CD8^+$ T-cell responses may be more likely to drive selection pressure on HIV than HLA-A or HLA-C, as previously described (Kiepiela *et al.* 2004). However, in the minority of cases where HLA-A and HLA-C-restricted responses do drive reverting mutations, these are also of benefit in terms of immune control.

Impact of reverting and non-reverting mutations on viraemic control

The strong correlation between the number of reverting mutations and decreasing VL may appear somewhat counterintuitive, since escape mutations are selected only if they are of benefit to the virus. However, the benefit may only be significant when balanced against the risk of being eliminated by the $CD8^+$ T cell response, and several studies now demonstrate the consequences, for both donor and recipient, of escape mutations that reduce viral fitness (Crawford *et al.* 2007; Schneidewind *et al.* 2007; Chopera *et al.* 2008; Goepfert *et al.* 2008).

In contrast to the effect mediated by reverting sites, these data show that 'non-reverting' mutations do not contribute substantially to reduction in VL. These are likely to be escape mutations that do not confer a fitness cost to the virus, or where a compensatory mutation alleviates fitness constraints (Crawford *et al.* 2007; Schneidewind *et al.* 2007). Mutations that do not revert can accumulate in populations and may eventually replace population wildtype (Leslie *et al.* 2005). The implication of these accumulating mutations is that, first, epitopes are potentially lost over time and no longer contribute to CTL-mediated control and, second, that the escape occurring at these sites does not confer the advantage of reduced viral fitness. Epitope stability may therefore be an important consideration in selection of potential immunogens for a T cell vaccine.

3.4.2 Relationship of this work to other recent studies

In vitro studies of Gag epitopes restricted by HLA-B*57 and B*27, the class I molecules most consistently associated with control of viraemia, have demonstrated that the escape mutations selected substantially reduce viral fitness (Migueles *et al.* 2003; Martinez-Picado *et al.* 2006). Another study based on published epitopes hypothesised that 'effective' $CD8^+$ T cell responses may be defined by their ability to select escape mutations that impose a fitness cost on the virus (Frater *et al.* 2007). Studies of early HIV infection support the findings described here, indicating that transmission of viruses carrying escape mutations within HLA-B-restricted Gag epitopes is of benefit to the recipient, and is associated with lower viral setpoints (Chopera *et al.* 2008; Goepfert *et al.* 2008). Corresponding studies of SIV infection in vaccinated Burmese macaques indicated similarly that durable control is achieved through early escape within a Gag epitope, resulting in a fitness cost to the virus, in

combination with multiple additional Gag-specific $CD8^+$ T cell responses (Matano *et al.* 2004; Kawada *et al.* 2006).

3.4.3 Advantages of lineage-corrected statistical approach

The methodology used here was designed to maximise the sensitivity and specificity with which HLA-selected polymorphisms were detected in a large cross-sectional cohort. Sensitivity for the detection of HLA-selected polymorphisms, particularly for common alleles, was optimised by maximizing patient numbers as far as possible. The output is validated by the finding of many HLA associations that are already well substantiated, or occur in previously defined epitopes of the appropriate restriction.

Particularly in light of recent data highlighting the potential role of HLA-C (Fellay *et al.* 2007), it was essential to use an approach that did not focus on previously characterised epitopes. Thus, associations were sought between any polymorphism with any HLA class I allele. Critically, the method depended on appropriate correction for multiple comparisons, for sequence relatedness (Bhattacharya *et al.* 2007), and for linkage disequilibrium effects.

3.4.4 Limitations of lineage-corrected statistical approach

Factors other than HLA genotype at a single locus influence disease control

This study simplifies the control of viraemia to median VL for a single allele at a given Class I locus. However, within the HLA genotype, diversity of haplotypes and combinations of HLA alleles may also account for differences in viraemic control. Although viral setpoints tend to be broadly characteristic of individual HLA alleles,

there is still unexplained variance in control of viraemia, and clearly factors other than HLA genotype also have a significant impact.

Sensitivity of detection of escape and reversion

The ability of the algorithm to detect escape and reversion depends on the rate of selection. Escape that happens early, (e.g. Thr to Asn at position 242 in the presence of HLA-B*5703), can be more reliably detected than escape that is selected later in the course of infection.

The methodology employed here relies upon statistical determination of reversion. However, the occurrence of reversion may be underestimated if reversion occurs late in the course of disease or following the selection of a rare mutation. Likewise mutations that are associated with a fitness cost may not be detected as reverting due to the coexistence of compensatory mutations. Furthermore, in order to detect reversion, this model makes the universal assumption that escape mutations are transmitted from donor to recipient. This is an oversimplification, as the 'bottleneck' operating at transmission across mucus membranes reduces transmission to a median of one virion (Keele *et al.* 2008; Salazar-Gonzalez *et al.* 2008); with a stochastic component to this process, it is possible that even a widely-selected escape variant is not transmitted.

An alternative approach to identifying sites of positive selection is calculation of a dN/dS ratio (non-synonymous / synonymous mutation rate), where a ratio >1 is indicative of positive selection occurring above the rate of genetic drift (de Oliveira *et al.* 2004; Delport *et al.* 2008). However, this strategy is only a reliable indicator of

diversifying selection in the absence of reversion (Delport *et al.* 2008), and is also dependent upon the frequency of the selecting allele.

False positive and false negative results

False positives are accepted by the model according to the q-value cut-off that is selected; in the data reported here, a value of 0.2 (20% FDR) or 0.05 (5% FDR) has been used. False positive associations may also arise as a consequence of linkage disequilibria between alleles, particularly when an allele at one locus is linked to two alleles at a second locus that are both involved in selection. In this instance, the strength of the association of the former will outweigh either of the two latter, although the true association is driven by the latter.

False negatives may arise as a consequence of insufficient statistical power caused by low sequence numbers or by association with rare alleles. The model is based on the assumption that phylogenetic clusters occur as a result of founder effect, and that shared mutations within these clusters are as a result of descent from a common ancestor rather than arising as a consequence of HLA-mediated selection. However, in certain instances it is possible that shared HLA-footprints may themselves cause phylogenetic clustering; a true HLA-mediated effect may therefore be missed. This hypothesis is explored in further detail in Chapter 4.

3.4.5 Mechanisms for HLA-mediated control of viraemia

The mechanisms by which HLA-B and HLA-C specifically mediate their effects through mutations in Gag and Pol respectively are unknown. The marked genetic diversity of HLA-B may account for a broad presentation of HIV epitopes, and HLA-

B restricted epitopes are particularly enriched within Gag. Furthermore, in the SIV model, Gag has been shown *in vitro* to be the first protein presented on the cell surface following acute infection (Sacha, Chung, Rakasz *et al.* 2007). Early targeting of Gag may enable killing of target cells before viral progeny are produced, therefore reducing VL.

It has been suggested that high expression of HLA-C might be associated with successful control of HIV (Fellay *et al.* 2007). This prompts the hypothesis that HLA-C alleles selecting reverting escape mutants in Pol might be those expressed at high levels. The emergence of Pol as a potentially important $CD8^+$ T cell target is also consistent with recent kinetic studies that showed both Gag- and Pol-specific $CD8^+$ T cells may be activated by virus-infected cells prior to Nef-mediated MHC class I down-regulation (Sacha, Chung, Reed *et al.* 2007). Other host genetic factors may also contribute to control of viraemia. These may be artificially attributed to HLA-mediated effects if they are in genetic linkage with the MHC locus.

3.4.6 Relevance of these studies to HIV vaccine design

These data may be relevant in the context of the failed Merck vaccine trial (Steinbrook 2007; Moore *et al.* 2008), although the reasons for vaccine failure are incompletely understood. One possibility is that the breadth of induced Gag, Pol and Nef $CD8^+$ T cell responses in vaccinees (a median of one response per protein) was not sufficient to protect against transmission or to influence viral setpoint in vaccinees who subsequently became infected. The implications from the data presented here underline again that a successful vaccine is likely to need to induce broad Gag-specific $CD8^+$ T cell responses restricted by HLA-B alleles. However, these data also

show that targeting of Pol epitopes by HLA-C-restricted $CD8^+$ T cells may contribute to improved control, via the same mechanism of selecting escape mutations that inflict a fitness cost on the virus.

Although escape by the virus from a single epitope is almost inevitable, the induction of several Gag and Pol responses by a vaccine would ensure that multiple mutations would be needed to bring about escape. The data presented here suggest that these mutations in themselves may contribute to immune control via an effect on viral replicative capacity.

CHAPTER 4: The impact of HLA selection on HIV phylogeny

The majority of the work presented in this chapter has been published in *J. Virol.* 2009 83(9):4605-15. HLA footprints on HIV-1 are associated with interclade polymorphisms and intraclade phylogenetic clustering. Matthews PC, Leslie AJ, Katzourakis A, Crawford H, Payne R, Prendergast A, Power K, Kelleher AD, Klenerman P, Carlson J, Heckerman D, Ndung'u T, Walker BD, Allen TM, Pybus OG, Goulder PJR.

4.1 INTRODUCTION

4.1.1 Selection of HLA footprints and relationship to HIV clades

HLA selection pressure generates viral escape mutations that reduce epitope presentation or recognition, seen in viral sequences as a characteristic HLA 'footprint' (McMichael and Klenerman 2002) – a predictable array of mutations arising to escape from the responses restricted by a particular HLA allele (Moore *et al.* 2002; Martinez-Picado *et al.* 2006; Brockman *et al.* 2007; Schneidewind *et al.* 2007; Matthews *et al.* 2008). These HLA footprints are particularly strongly associated with alleles associated with effective immune control of HIV, such as HLA-B*57 and HLA-B*27 (Kaslow *et al.* 1996; O'Brien *et al.* 2001; Kiepiela *et al.* 2004).

The starting point for this work is the observation that patterns of escape mutations selected in one clade may represent consensus in another clade. Specifically, from previous studies of B- and C-clade infected cohorts (Leslie *et al.* 2004; Martinez-Picado *et al.* 2006), escape mutations selected in B-clade infected individuals expressing HLA-B*57 frequently represent the consensus in C-clade sequences. Another example is the CRF_AE strains that dominate the epidemic in Thailand, that bear many of the signature mutations of the HLA-B*57 footprint.

4.1.2 Immune control of HIV in association with HLA-B*57 and HLA-B*27

HLA-B*57 immune control and escape

HLA-B*57 is strongly associated with immune control of HIV and is universally enriched in groups of LTNPs (Migueles *et al.* 2000; Migueles *et al.* 2003; Emu *et al.* 2008; Miura *et al.* 2009). Although not all subjects with HLA-B*57 maintain immunologic control indefinitely, HLA-B*57 remains strikingly over-represented even in the most stringently defined cohorts of elite controllers (VL<50 RNA copies/ml plasma and normal CD4 count after >10 years of infection in the absence of HAART) (Gea-Banacloche *et al.* 2000). HLA-B*5703 is most strongly associated with the selection of 5 escape mutations in Gag (Draenert *et al.* 2004; Leslie *et al.* 2004; Crawford *et al.* 2007), (Fig. 4.1A) and 2 in Nef (Fig. 4.1B) (Leslie *et al.* 2005, Matthews *et al.* 2008).

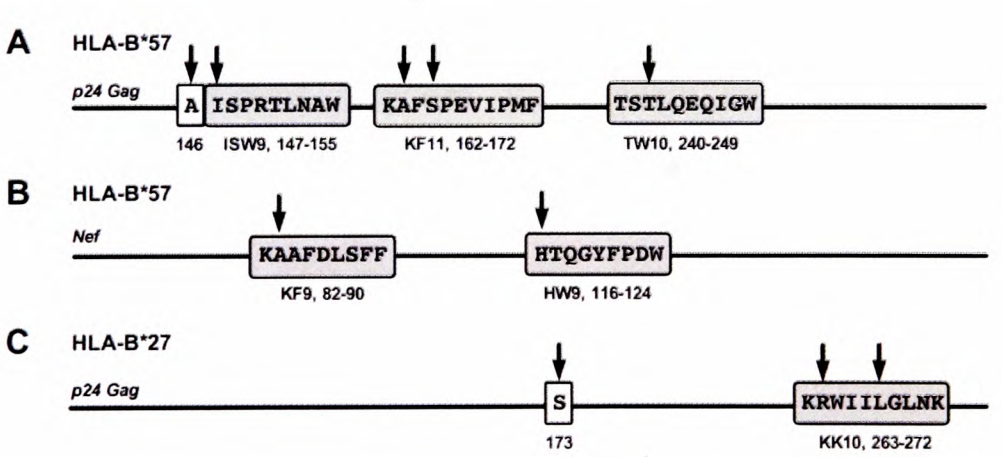


Figure 4.1: HLA footprint arising in subjects with HLA-B*57 (A, B) and HLA-B*27 (C) in Gag (A, C) and Nef (B). Epitopes are marked in grey boxes, and sites of associated compensatory mutations in white boxes. (Leslie *et al.* 2004; Matthews *et al.* 2008; Boutwell *et al.* 2009; Schneidewind *et al.* 2007; Schneidewind *et al.* 2008). Locations of common escape variants are shown with arrows. Amino acids listed are subtype B consensus. Adapted from figure in Payne *et al.* (Payne *et al.* 2009).

Early in the course of HIV infection with clade B or C, subjects with HLA-B*57 select for Thr to Asn mutation at Gag position 242 (T242N) in the epitope TSTLQEQIAW (TW10, Gag 240-249) (Leslie *et al.* 2004). Selection also occurs at position 248, where the wild-type amino acid differs between HIV subtypes (Gly in B-clade; Ala in C-clade). G248X mutation is strongly associated with HLA-B*57 in subtype B infection, while the association in subtype C is also present, but weaker (Leslie *et al.* 2004). Following transmission to an HLA-B*57 negative recipient, these mutations revert to wild-type (Leslie *et al.* 2004). In the case of the HLA-B*57 selected Gag mutation T242N, viral replicative capacity and infectivity have been shown to be diminished compared to wild type *in vitro* (Martinez-Picado *et al.* 2006). Thus, although escape from the dominant $CD8^+$ T cell response arises early, the resulting viruses are partially crippled by the presence of the mutation.

HLA-B*27 immune control and escape

HLA-B*2705, an allele occurring almost exclusively in Caucasoid populations, is also associated with favourable control of viraemia in chronic infection. Like HLA-B*57, the immunodominant response is directed at a Gag epitope, KK10 (KRWIILGLNK, Gag 263-272). Characteristic escape mutations are selected by HLA-B*27 at positions 173, 264 and 268 (Schneidewind *et al.* 2007; Schneidewind *et al.* 2008), (Fig. 4.1C). However, in contrast to the early mutations selected by HLA-B*57, the escape variants selected by HLA-B*27 arise late in the course of infection and are associated with rapid progression to AIDS (Goulder, Phillips *et al.* 1997).

4.1.3 Hypothesis and aims

The initial hypothesis was that amino acid differences between clades are commonly also those that represent escape variants associated with particular HLA alleles. A second, related, hypothesis was that selection pressure imposed by HLA alleles might have an impact on viral evolution within a clade. This would arise if genetically unrelated sequences acquired the same characteristic combinations of escape mutations as a consequence of selection by the same HLA Class I allele, causing them to cluster phylogenetically.

The aim of this work was therefore to investigate the extent to which HLA-footprints impact on HIV sequence diversity and phylogeny, first by considering their relationship to sites of difference between clades, and second by quantifying their influence on clustering of sequences within a clade.

4.2 METHODS

4.2.1 Impact of HLA selection on sites of amino acid variation between clades

In order to investigate whether HLA footprints might account for some of the amino acid differences occurring between HIV-1 Clades, sequences for Gag, Pol and Nef for clades A1, A2, AE, B and C, were downloaded from the Los Alamos HIV database. These subtypes were chosen as subtypes A, B and C account for the majority of infections worldwide, and accordingly most sequence data are available for these groups. In addition, the AE CRF strain that dominates the Thai epidemic (Lau *et al.*

2007) was included to explore the observation that these sequences show evidence of an ancestral HLA footprint.

The most recent (2004) consensus for each of these clades was used to identify residues at which there are amino acid differences between clade consensus sequences. Sites of HLA-mediated polymorphism were identified according to the analysis of C-clade Gag, Pol and Nef sequences generated from Durban sequences as described in Chapter 3 (Matthews *et al.* 2008). Fisher's Exact Test was used to test whether sites at which there is inter-clade variability are also statistically likely to be sites of HLA-associated polymorphism. In order to determine the extent to which the results of this analysis are applicable despite longitudinal changes that occur in HIV sequence, ancestral HIV sequences for clades A1, B and C were also compared to current (2004) consensus sequences for these three clades.

4.2.2 Evaluating the impact of clade-specific amino acid polymorphisms on phylogeny

To investigate the effect of the sites of inter-clade difference on the phylogenetic distinction between clades, Gag p24 population sequences were downloaded from Los Alamos HIV database. To determine whether these variable sites are fundamental to defining clades, 20 taxa were selected at random from clades A1, A2, AE, B and C, and all 12 taxa available for clade A2 were used. Neighbour joining (NJ) phylogenetic trees were constructed from nucleotides in PAUP (Hasegawa *et al.* 1985), in the presence and absence of 27 codons at which there were amino acid differences between clades.

Next, to investigate whether amino acid polymorphisms characteristic of one clade alter phylogenetic clustering when superimposed onto sequences from another clade, 20 sequences were selected at random from clades A1, B and C. Sequences were then modified at sites of variability between clades, by substituting the consensus codon for one clade onto taxa selected from another clade. NJ phylogenetic trees were constructed using the original sequences for each clade, plus the altered sequences bearing the characteristic codons of a different clade.

4.2.3 Relationship between Shannon entropy and sites of HLA selection

Population sequences from Durban Gag, Pol and Nef were used to investigate the relationship between Shannon entropy and HLA-footprint sites. Sequences with gaps of >5 consecutive amino acids were removed in order to avoid false over-estimation of entropy due to missing data. Total sequence numbers for this analysis were p17 Gag: 584, p24 Gag: 646, p15 Gag: 421, Protease: 402, Reverse Transcriptase: 254, Integrase: 244, Nef: 424. Sites of HLA-associated amino acid polymorphism were taken from the previous analysis of Durban sequences (Chapter 3) (Matthews *et al.* 2008). A conservative approach was adopted by excluding sites at which there is no amino acid variation (entropy score = 0), as, by definition, HLA-selection cannot operate at these sites. Entropy at sites of HLA-A, -B and –Cw selection was compared to entropy at sites where no HLA-selection pressure had been identified. Significant differences in entropy scores between these two groups were sought using Mann-Whitney test.

4.2.4 Investigating phylogenetic clustering among sequences bearing HLA footprints

*HLA-B*5703 footprint*

This footprint occurs in C-clade infected-cohorts; taxa were selected from the Durban cohort from a total pool of 566 Gag sequences (p17+p24, 1080 nucleotides) and 443 Nef sequences (621 nucleotides). HLA-B*5703 was present in 35 subjects with Gag sequences (6.2%) and 16 subjects with Nef sequences (3.6%).

*HLA-B*27 footprint*

This footprint occurs predominantly in the context of B-clade infection. Sequence data were derived from 149 subjects with acute HIV-1 infection recruited from Boston, USA and Sydney, Australia (Allen *et al.* 2005; Li B 2007), and 234 subjects with chronic infection (cohort described by Schneidewind *et al* (Schneidewind *et al.* 2007); sequences available at Los Alamos HIV database). 6 subjects from the acute cohort and 19 subjects with chronic infection had HLA-B*27. To contribute to a pool of 'HLA-B*27 negative' subjects, sequences from Los Alamos database expressing the wild-type amino acid at HLA-B*27 footprint sites were selected. Due to limited sequence availability in the chronic cohort (Schneidewind *et al.* 2007), phylogenetic trees were restricted to taxa of 330 nucleotides length.

Quantification of phylogenetic clustering

In order to quantify phylogenetic clustering, a maximum parsimony approach was adopted. This method calculates the minimum number of mutations required to produce an evolutionary history consistent with the specified amino acid pattern. All amino acid changes at the specified HLA-footprint sites on an NJ tree were mapped

using the parsimony algorithm implemented in the MacClade program (Maddison and Maddison 2000) (Fig. 4.2).

The minimum number of mutations for each of the footprint sites was then summed, giving a total minimum number of evolutionary changes. A smaller number of changes reflects fewer inferred escape mutations and a greater degree of phylogenetic clustering of the sequences bearing the HLA-footprint.

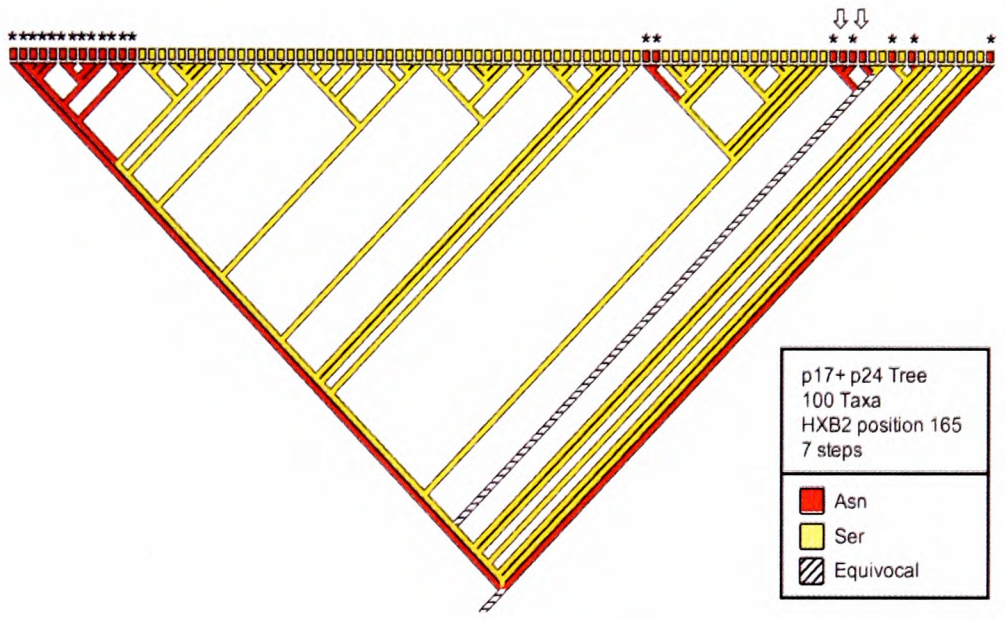


Figure 4.2: **Neighbour joining tree of 100 HLA-B*5703-negative taxa (Gag p17 + p24) analysed for clustering using the maximum parsimony algorithm in MacClade.** Taxa were selected at random from the Durban cohort. Five footprint mutations selected by HLA-B*5703 were artificially superimposed on the 20 taxa marked with stars. The position shown corresponds to HXB2 position 165, within the HLA-B*57 restricted epitope KF11 (Gag 162-172; KAFSPEVIPMF). Arrows indicate taxa that naturally bear a mutation at this position (n=2). The parsimony algorithm calculates that a minimum of 7 mutations is needed to explain the phylogenetic distribution of taxa at this position.

*Clustering of sequences bearing a naturally selected HLA-B*5703 footprint*

Phylogenetic clustering of sequences bearing a true (naturally selected) HLA-B*5703 footprint was assessed by generating a dataset containing 100 C-clade Gag sequences composed of 20 HLA-B*5703 positive and 80 HLA-B*5703 negative individuals

selected at random from the Durban cohort. This selection process was repeated 100 times over, constructing a single NJ tree for each dataset, and quantifying phylogenetic clustering in the presence and absence of the HLA-B*5703 footprint sites. For a comparator, the analysis was repeated using phylogenetic trees constructed with the same data but from which the five HLA-B*5703 footprint sites had been stripped. The same process was undertaken for Nef, using 50 randomised datasets, each containing 100 C-clade taxa. Each dataset contained the same 16 sequences from HLA-B*5703 positive patients (due to limited data, these were not randomised), and 84 selected at random from HLA-B*5703 negative patients.

In order to account for the statistical variance arising from phylogeny estimation using NJ trees, one set of 100 Gag taxa was used to generate 100 bootstrap trees, and clustering in the presence and absence of the footprint sites was assessed as above.

Clustering of sequences bearing a simulated HLA-B*5703 footprint

To further explore the phylogenetic impact of the HLA-B*5703 footprint on viral phylogeny, a model was developed to quantify the phylogenetic impact of between 1 and 5 footprint mutations on between 5 and 20% of sequences. Using 100 Gag sequences selected at random from HLA-B*5703-negative subjects, a characteristic, conserved, HLA-B*5703 footprint was superimposed on a varying proportion of sequences (5%, 10%, or 20%) from this pool of 100, starting with the mutation site with the strongest HLA-B*5703 association (defined by Fisher's exact test). The full footprint of five mutations was added, one site at a time, to each of the 20 selected taxa. This process of random sequence selection and addition of a conserved footprint was repeated 20 times.

Clustering of sequences bearing an HLA-B*27 footprint

To assess the impact of a naturally selected HLA-B*27 footprint, taxa from B-clade cohorts were used. The HLA-B*27 footprint is selected late in infection (Goulder, Phillips *et al.* 1997; Schneidewind *et al.* 2007), so the footprint was assessed first in the context of acute infection, and then in a chronically infected cohort. For acutely infected subjects, an NJ tree was generated from all 6 subjects with HLA-B*27, and 93 subjects selected at random from the pool of HLA-B*27 negative subjects; in the chronic cohort, from all 19 subjects with HLA-B*27, and 81 HLA-B*27 negative subjects selected at random.

The impact of an HLA-B*27 footprint on viral phylogeny was first investigated using taxa from the Durban cohort (in which HLA-B*27 is not represented). As for the simulated HLA-B*57 footprint, 100 Gag sequences were selected at random, and the characteristic three-site HLA-B*27 footprint was superimposed on twenty of these; this process was repeated 20 times.

4.3 RESULTS

4.3.1 HLA footprint sites overlap with sites of inter-clade HIV amino acid polymorphism

The potential impact of HLA footprints on HIV evolution was first assessed by investigating whether sites of HLA selection correspond to sites of variation in amino acid residues between clades in Gag, Pol and Nef.

In Gag (p17 + p24 + p15) sequences downloaded from Los Alamos HIV database from clades A1, A2, AE, B and C (n=1230), only 3.2% of sites (16 of 500 residues)

were invariant. Sites of inter-clade variability were defined based on consensus sequences; in Gag, Pol and Nef, 103, 136 and 62 residues, respectively, were identified that vary between clades A1, A2, AE, B and C (p24 Gag shown in Fig. 4.3). In Gag, Pol and Nef these sites were compared with locations of HLA-mediated selection pressure that were identified in the Durban analysis described in Chapter 3 (Matthews *et al.* 2008). A strong association was seen between sites of inter-clade amino acid variability and sites of selection pressure mediated by HLA-A, B and C ($p=4.1x10^{-7}$, $p=9.8x10^{-19}$, $p=1.3x10^{-5}$ respectively; Table 4.1).

		HLA-A	HLA-B	HLA-C
Gag	p17	0.15	$8.68x10^{-3}$	1
	p24	$4.27x10^{-3}$	$5.04x10^{-10}$	0.62
	p15	1	0.47	1
Pol	Protease	0.076	$1.92x10^{-3}$	0.19
	RT	$2.22x10^{-3}$	$2.21x10^{-6}$	$2.63x10^{-7}$
	Integrase	0.14	$1.19x10^{-3}$	1
Nef		0.032	0.018	0.24
Gag+Pol+Nef		$4.13x10^{-7}$	$9.77x10^{-19}$	$1.34x10^{-5}$

Table 4.1: Correlation between sites of HLA-mediated immune selection pressure (q<0.2) and sites of inter-clade variability (defined from clades A1, A2, AE, B, C). p values by Fisher's Exact Test. Statistically significant associations are shown in red ($p<0.001$, corrected for multiple comparisons by the Bonferroni method).

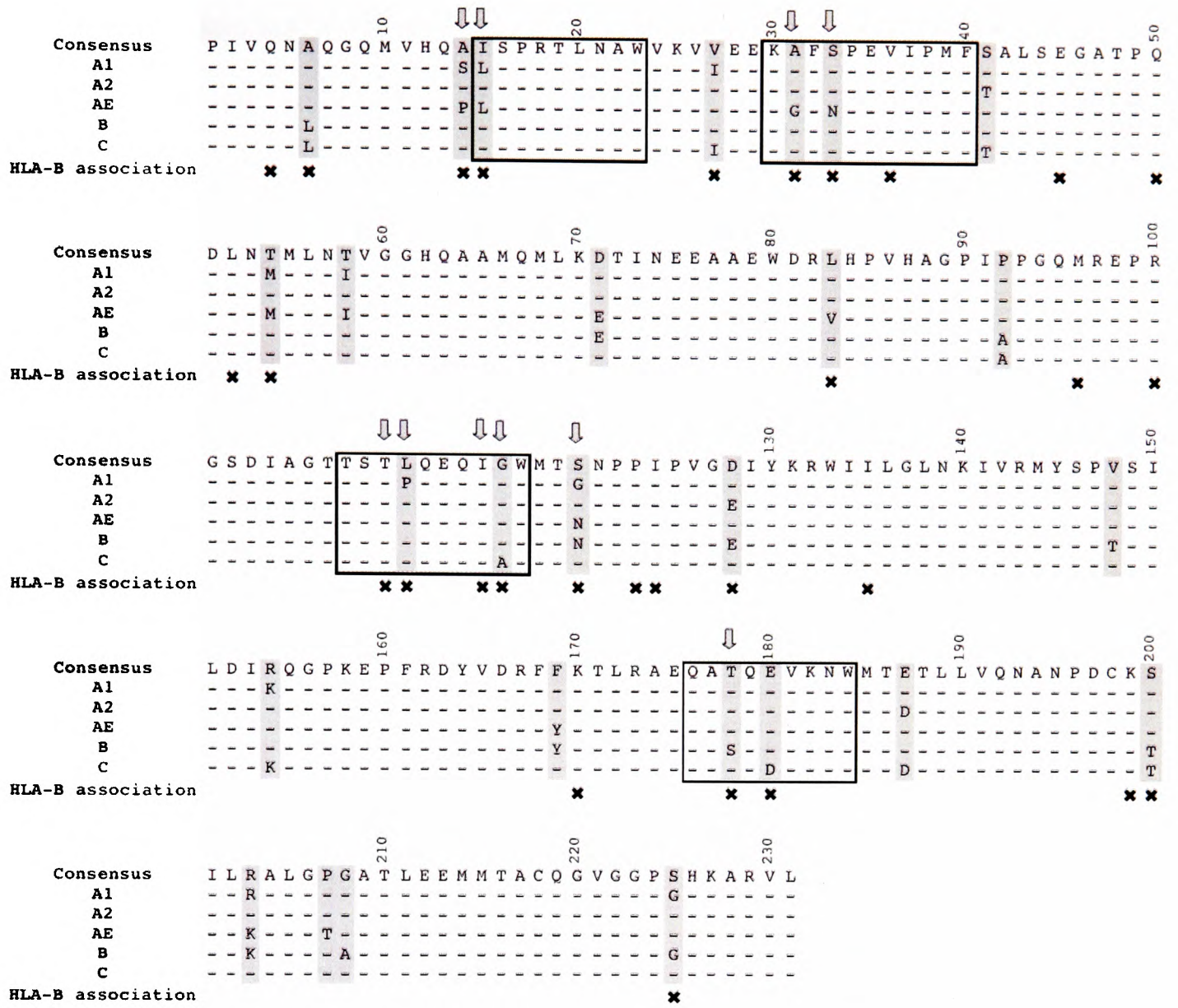


Figure 4.3: Sites of inter-clade variability and HIV-associated polymorphisms in p24 Gag. Consensus sequences (2004) for HIV clades A1, A2, AE, B and C are shown. Sites of inter-clade variability are marked with grey bars, and '×' indicates the sites of HLA-B mediated selection pressure identified from analysis of subjects with C-clade infection in Durban (Matthews *et al.* 2008). The four HLA-B*57/5801 epitopes in p24 Gag are enclosed in open boxes, with arrows indicating sites of HLA-B*57/5801-associated escape mutation that have been described in previous studies of B and C clade infection (Martinez-Picado *et al.* 2006; Brumme, Tao *et al.* 2008; Matthews *et al.* 2008). There is correlation between sites of inter-clade variability and sites of HLA-B mediated selection; $p<10^{-6}$ (Fisher's exact test).

Consistent with previous studies (Kiepiela *et al.* 2004), the strongest association with inter-clade amino acid differences was for sites of selection pressure mediated by HLA-B in p24 Gag ($p=5.0x10^{-10}$) (Table 4.1). This remained significant even when HLA-B*57/5801 was excluded from analysis ($p=3.6x10^{-8}$), demonstrating that the association is not attributable just to the effect of these immunodominant alleles.

Overall, 59% of sites of inter-clade p24 Gag variability were also identified as sites of HLA-B-driven selection pressure (Fig. 4.4A). For the Pol and Nef proteins, 18% of variable residues were also sites of HLA-B-driven escape mutation (example in Fig. 4.4B).

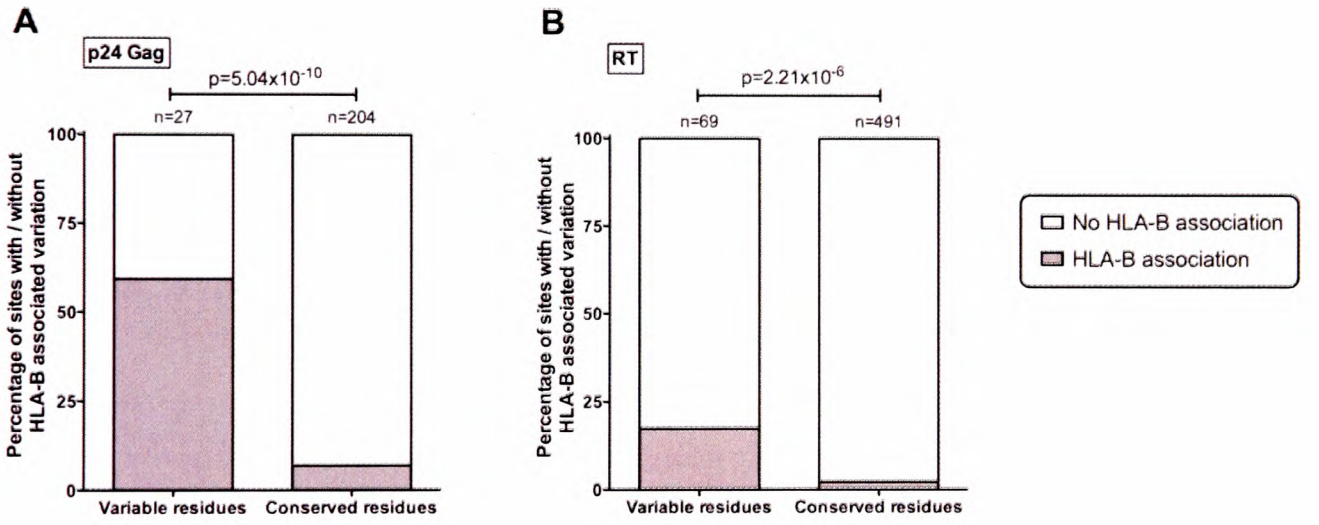


Figure 4.4: Proportion of amino acid residues in p24 Gag (A) and RT (B) at which HLA-B-associated polymorphisms are detected according to the presence or absence of inter-clade variability. All amino acids in each protein are represented, divided into 'variable residues' at which there are inter-clade differences in amino acids or 'conserved residues' that are identical between consensus sequences for clades. Proportion of each of these sites at which HLA-B selection has been previously identified (Matthews *et al.* 2008) is shown in grey. p values by Fisher's exact test.

In comparing predicted ancestral sequences with contemporary (2004) consensus sequences, differences were present in only 5.1% of all residues across Gag, Pol and Nef, suggesting little overall longitudinal change in amino acids that characterise individual clades. The amino acids that differed between ancestral and consensus sequences were largely the same as amino acids that differed between modern consensus sequences for clades A1, A2, AE, B and C (in 90%, 100% and 96% of cases in Gag, Pol and Nef, respectively).

4.3.2 Shannon entropy scores are higher at sites of HLA selection

C-clade sequences were used to compare Shannon entropy for sites at which there is no known HLA selection, to sites at which there is HLA-associated polymorphism (as identified previously (Matthews *et al.* 2008)). As expected, entropy scores were significantly higher at sites of HLA-mediated selection (Fig. 4.5); this relationship was most consistent for HLA-B (Table 4.2).

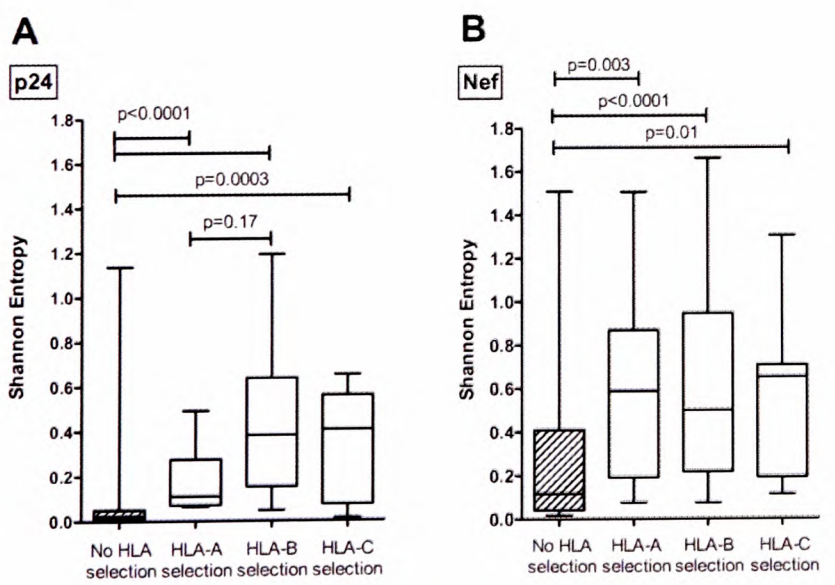


Figure 4.5: Shannon entropy at sites of HLA-A, -B and -C selection compared to sites at which no HLA-selection is detected in C-clade sequences.

A: p24 Gag. B: Nef. P values by Mann-Whitney test.

		HLA-A	HLA-B	HLA-C
Gag	p17	0.0152	0.0005	N/A
	p24	<0.0001	<0.0001	0.0003
	p15	N/A	N/A	N/A
Pol	Protease	N/A	0.0024	0.0703
	RT	<0.0001	<0.0001	<0.0001
	Integrase	0.0009	<0.0001	N/A
Nef		<0.0003	<0.0001	0.0123
Gag+Pol+Nef		<0.0001	<0.0001	<0.0001

Table 4.2: Relationship between Shannon entropy at sites of HLA-associated amino acid polymorphism compared to sites at which no HLA-selection was detected. Data from analysis of C-clade (Durban) sequences. Invariant sites (Shannon entropy score=0) were excluded from analysis. Statistically significant associations (defined as Mann Whitney test, $p<0.001$, corrected for multiple comparisons by the Bonferroni method) are shown in red. N/A (not applicable) applies where too few associations were detected to allow statistical analysis.

4.3.3 Synonymous and non-synonymous nucleotide substitutions contribute to clade specific phylogeny

To investigate the extent to which sites of amino acid variability impact on phylogenetic distinction of HIV taxa into clades, phylogenetic trees were constructed in the presence and absence of the 27 codons that confer amino acid variability between p24 Gag clade consensus sequences. The majority, but not all, of these sites are also residues at which HLA-selection operates (Figs 4.3, 4.4). Not unexpectedly, the phylogenetic distinction between these clades persists in the absence of these variable sites (Fig. 4.6), showing that the distinction between clades exists as a result of synonymous nucleotide substitutions, as well as because of amino acid differences.

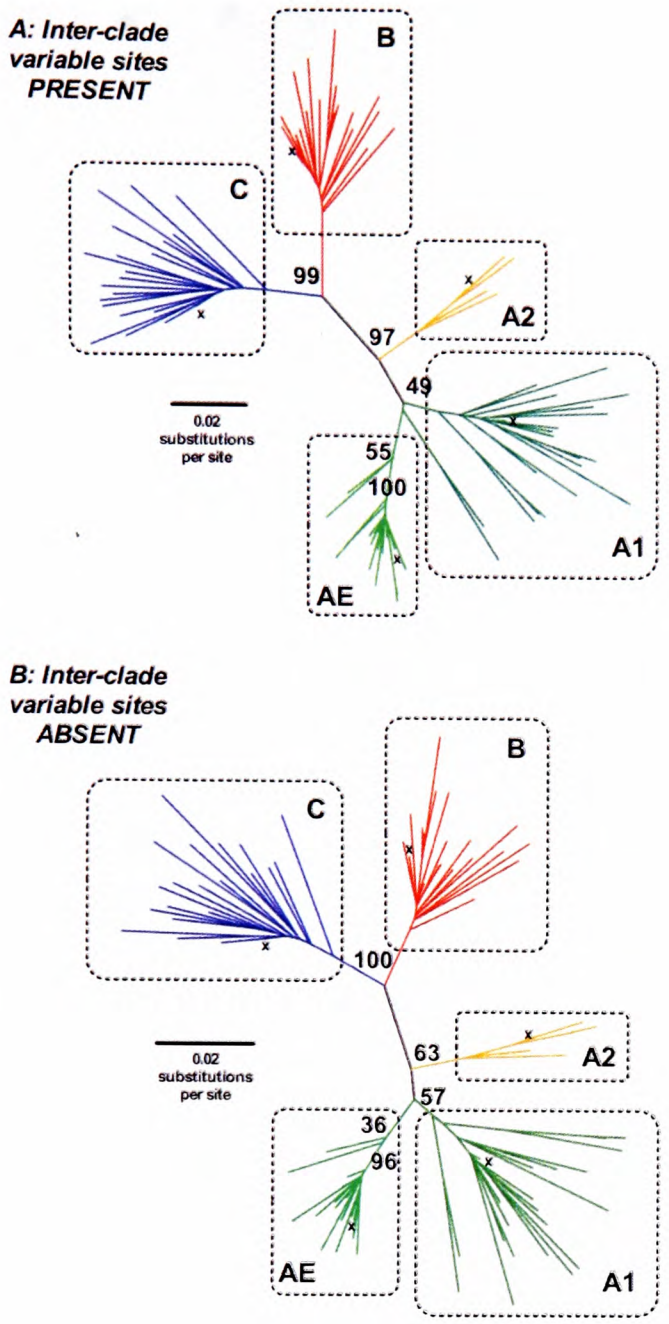


Figure 4.6: Phylogenetic trees illustrating the preservation of clade phylogeny in the presence and absence of sites of clade-specific difference.

Sequences were selected at random from www.hiv.lanl.gov, from clades A1, A2, AE, B and C, and ML phylogenetic trees were constructed (midpoint rooted, bootstrap values based on 100 replicates).

Each clade is enclosed within a dashed box to show clustering in the presence and absence of sites that vary between clades. The 2004 consensus sequence for each clade is marked 'x'.

A: Tree constructed from complete p24 sequences (693 nucleotides).

B: Tree constructed from p24 sequences in the absence of 27 codons at which there is amino acid difference between consensus sequences (612 nucleotides).

To investigate the extent to which the presence of amino acid polymorphisms can affect phylogeny, codons characteristic of one clade were next substituted onto the sequences for a second clade. Not unexpectedly, changes at these sites can alter clade-specific phylogenetic clustering (examples shown in Fig. 4.7).

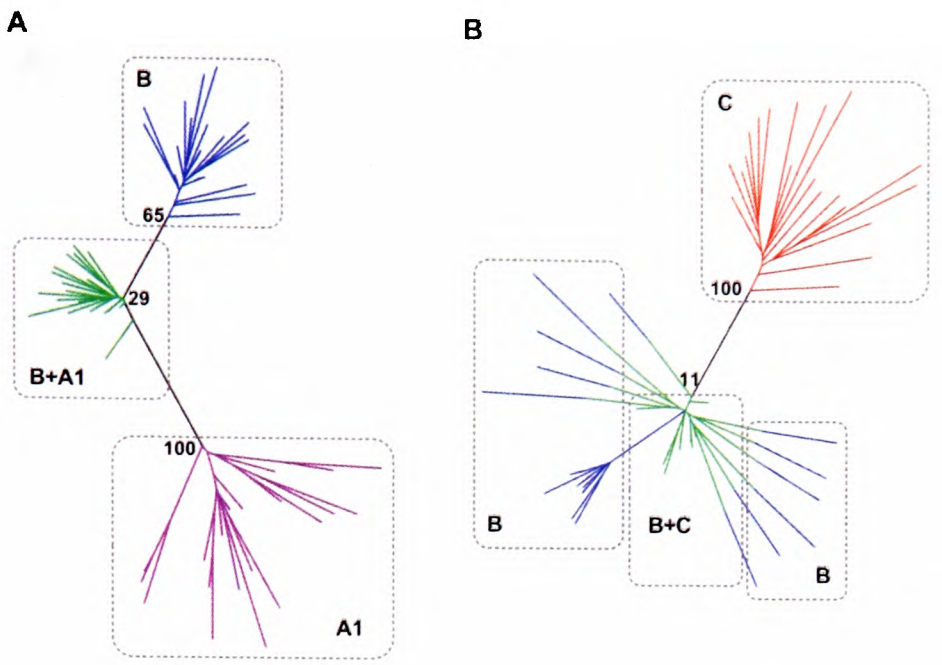


Figure 4.7: Phylogenetic trees illustrating altered distribution of taxa when codons determining clade-specific differences are swapped between clades. Twenty sequences were selected at random from the clades indicated within dashed boxes. ML trees were constructed from nucleotides (midpoint rooted, bootstrap values based on 100 replicates).

A: Codons defining amino acids characteristic of clade A1 (selected from clade A1 consensus) were superimposed on 20 sequences from clade B. The clade B sequences are shown twice, once without alteration (marked 'B') and once with A1-clade amino acids superimposed (marked 'B+A1').

B: Codons defining amino acids characteristic of clade C (selected from clade C consensus) were superimposed on the same 20 sequences from clade B. As before, the clade B sequences are shown twice, unchanged ('B') and bearing the characteristic C-clade codons ('B+C').

4.3.4 CRF subtype AE bears an HLA-B*57 footprint

The AE subtype dominates the HIV epidemic in Thailand (Fig. 4.8A). Despite the low frequency of HLA-B*57 in the Thai population (1.8% (Chandanayingyong *et al.* 1997)), it is apparent that AE consensus sequence closely resembles an HLA-B*5703 escape mutant, incorporating many components of the Gag HLA-B*5703 footprint (that is, A146P, I147L, A163G and S165N; Fig. 4.1). The only omission is T242N, the escape mutation that reverts to wild-type following transmission (Leslie *et al.* 2004; Martinez-Picado *et al.* 2006; Crawford *et al.* 2007) and hence does not persist at a population level. The AE subtype thus exemplifies the hypothesis that HLA selection may contribute significantly to founder strains of HIV (Fig. 4.8B-D).

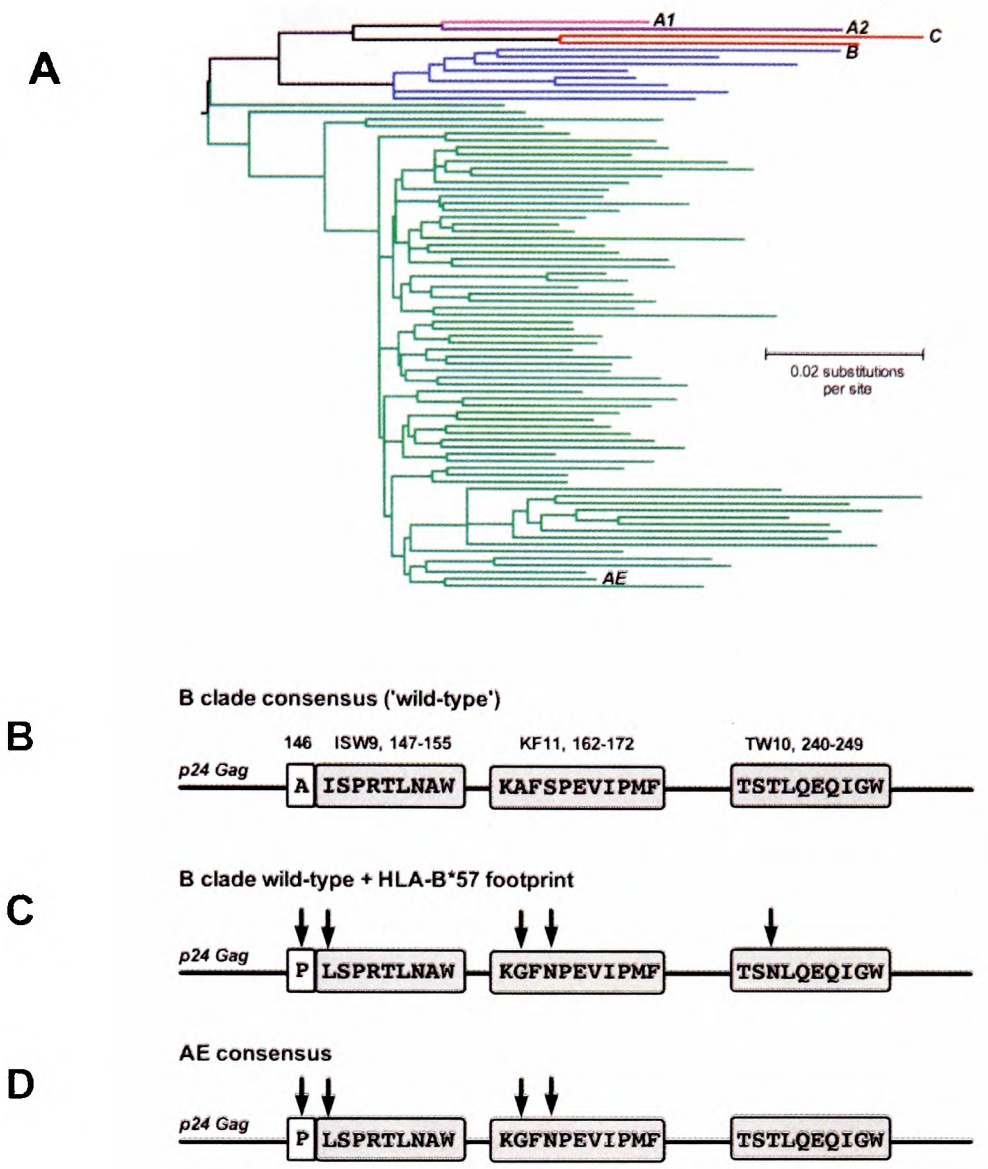


Figure 4.8: CRF_AE strains in Thailand and relationship to HLA-B*57 footprint. A: Maximum likelihood tree constructed from p17+p24+p15 Gag in GARLI from amino acids, using sequence data from 76 Thai subjects (sequences provided by T. Allen, *pers. comm.*). Consensus sequences for subtypes A1, A2, B C and AE are marked. Tree demonstrates that the majority of sequences cluster with CRF_AE subtype. B-D: Potential route of evolution of AE from a founder strain. B: 'Wild-type' strain of HIV. C: Wild-type strain altered by acquisition of HLA-B*57 escape mutations. C: 2004 consensus sequence for CRF_AE strains, bearing the majority of HLA-B*57 footprint.

4.3.5 Phylogenetic clustering can be mediated by an HLA-B*5703 footprint in Gag

*Quantification of HLA-B*5703 footprint in a C-clade cohort*

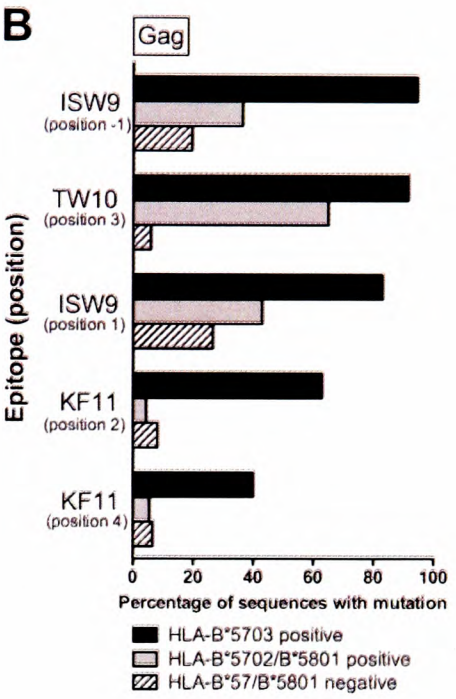
Previous analysis of C-clade sequences identified the HLA allele with the greatest number of associated HIV sequence polymorphisms as HLA-B*5703, with five strong associations in Gag alone ($q<0.05$, $p<10^{-6}$, Fig. 4.1, 4.9) (Matthews *et al.*

2008). For this reason, this allele was initially chosen to investigate the potential impact of a single allelic footprint on viral evolution within a clade.

A

Epitope (position)	Sequence	p value
ISW9 (Gag 147-155)	A I S P R T L N A W	$3.19x10^{-10}$
	A I S P R T L N A W	$1.26x10^{-9}$
KF11 (Gag 162-171)	K A F S P E V I P M F	$8.72x10^{-15}$
	K A F S P E V I P M F	$9.17x10^{-8}$
TW10 (Gag 240-249)	T S T L Q E Q I A W	$1.04x10^{-20}$

B



C

Epitope (position)	Sequence	p value
KF9 (Nef 82-90)	K A A F D L G F F	$3.41x10^{-5}$
HW9 (Nef 116-124)	H T Q G Y F P D W	$6.39x10^{-4}$

D

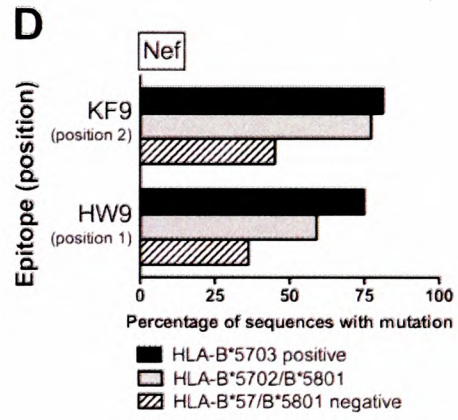


Figure 4.9: Frequency of HLA-B*5703 footprint mutations in Durban cohort. C-clade HLA-B*5703-footprint in (A, B) Gag (n=566 sequences) and (C, D) Nef (n=443 sequences). Epitopes, mutation sites and strength of association between HLA-B*5703, and the related alleles HLA-B*5702 and B*5801, and mutation at each position are determined by Fisher's Exact Test.

*Phylogenetic clustering can be mediated by an HLA-B*5703 footprint in Gag*

To determine whether HLA-B*5703 mediated selection pressure has a potential impact on viral evolution, phylogenetic clustering of C-clade sequences was assessed. Phylogenetic clustering among 100 Gag sequences, of which 20 were taken from subjects expressing HLA-B*5703, was quantified in the presence and absence of the

five HLA-B*5703-footprint sites. In the absence of these five codons, phylogenetic clustering of the sequences was reduced; repeating this analysis on 100 datasets, the clustering effect mediated by the HLA-B*5703 footprint sites was highly statistically significant ($p<0.0001$; paired t-test; Fig. 4.10A).

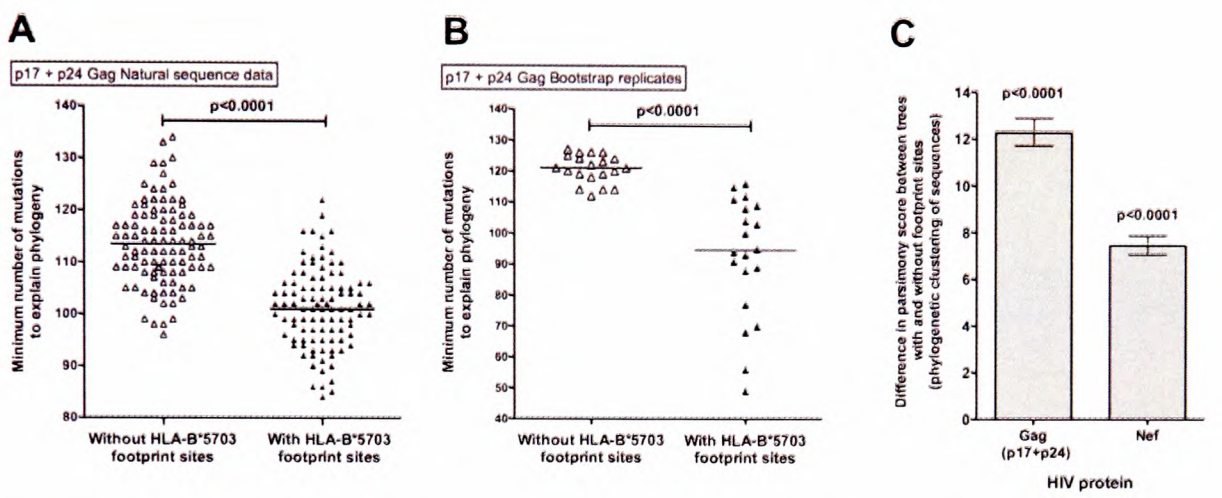


Figure 4.10: Phylogenetic clustering in sequences from individuals with HLA-B*5703 analysed in the presence and absence of footprint sites. p values by paired t-test. A: Phylogenetic clustering mediated by an HLA-B*5703 footprint was assessed in 100 NJ trees for Gag (composed of 100 taxa, including 20 from subjects with HLA-B*5703). B: Phylogenetic clustering in 20 bootstrap replicates of a single tree composed of 100 taxa, including 20 from subjects with HLA-B*5703. C: Difference in clustering (mean ± SEM) accounted for by the HLA-B*5703 footprint in Gag (as shown in panel A) and Nef (50 replicates of 100 taxa with 16 from subjects with HLA-B*5703).

In bootstrap replicates generated from a single dataset of 100 taxa, this clustering effect remained equally significant ($p<0.0001$; paired t-test; Fig. 4.10B). Such clustering occurs even though some HLA-B*5703-negative patients may also possess HLA-B*5703 footprint mutations, as a result of transmission (Bansal *et al.* 2007) or selection by non-HLA-B*5703 alleles (e.g. B*5801, Fig. 4.9) (Honeyborne *et al.* 2007).

Phylogenetic clustering can be mediated by an HLA-B*5703 footprint in Nef

In order to examine the effect of a smaller HLA-B*5703-footprint in a more variable protein than Gag, the same analysis was undertaken in Nef, which contains two polymorphisms associated with HLA-B*5703 (Fig. 4.1, 4.9) (Leslie *et al.* 2005; Matthews *et al.* 2008). Significant phylogenetic clustering of Nef sequences from B*5703-positive individuals was again seen as a consequence of shared footprint mutations ($p<0.0001$; paired t-test; Fig. 4.10C).

4.3.6 Phylogenetic clustering is brought about by a footprint of linked mutations

To address the possibility that excluding *any* group of polymorphic codons might itself significantly affect phylogenetic clustering, the analysis of Gag sequences was repeated using the same methods, but excluding five randomly chosen polymorphic sites in place of the five footprint sites characteristic of HLA-B*5703.

For each of five datasets, five randomly selected polymorphic sites were chosen. Shannon entropy scores of these randomly selected sites did not differ significantly from those of the HLA-B*5703 footprint sites ($p=0.2$, Mann-Whitney test). Comparison of phylogenetic clustering in the presence or absence of these randomly selected 'footprints' had no significant effect on clustering ($p=0.5$; paired t-test; Fig. 4.11).

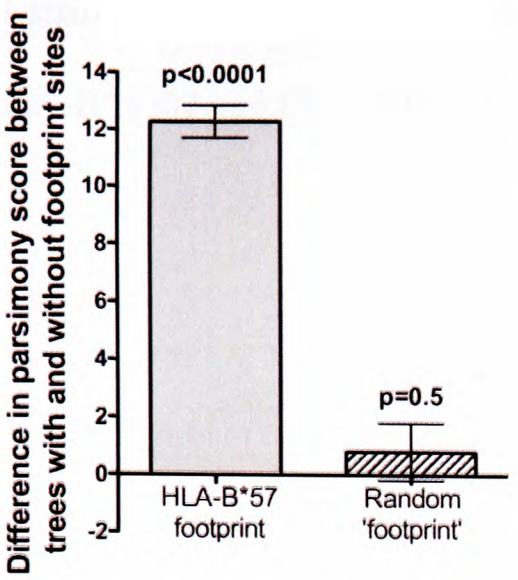


Figure 4.11: Comparison between effect of an HLA-B*57 footprint and a randomly selected 'footprint' on phylogenetic clustering of Gag (p17+p24) taxa.

The HLA-B*57 footprint consists of linked mutations that are shared between taxa, bringing about clustering among these sequences. Conversely, polymorphisms at randomly selected sites occur independently, rather than arising in a distinct subset of taxa as a result of shared selection pressure, and therefore have no significant impact on phylogenetic clustering.

4.3.7 Phylogenetic clustering of Gag sequences increases with addition of sequential footprint mutations or increase in the proportion of footprint-bearing sequences

In order to quantify more specifically the impact of accumulating HLA-selected mutations, a simulation of HLA-B*5703 footprinting was performed. Representative phylogenetic trees constructed from these sequences are shown (Fig. 4.12), demonstrating progressive clustering between sequences bearing the footprint.

Quantifying this clustering demonstrated that progressively fewer mutations were required to explain the phylogeny as more footprint polymorphisms were added, reflecting increasing clustering of footprint-bearing taxa (Fig. 4.13, 4.14A). Thus, significant phylogenetic clustering can arise as a consequence of imposing even a partial HLA-B*5703 footprint on randomly chosen Gag sequences. Using the same model, when footprint sites were excluded from analysis, phylogenetic clustering was reduced (p=0.0004, Fig. 4.13); this reduction in clustering may relate to the presence

of shared mutations at these sites selected by other closely related alleles (such as HLA-B*5702 and HLA-B*5801).

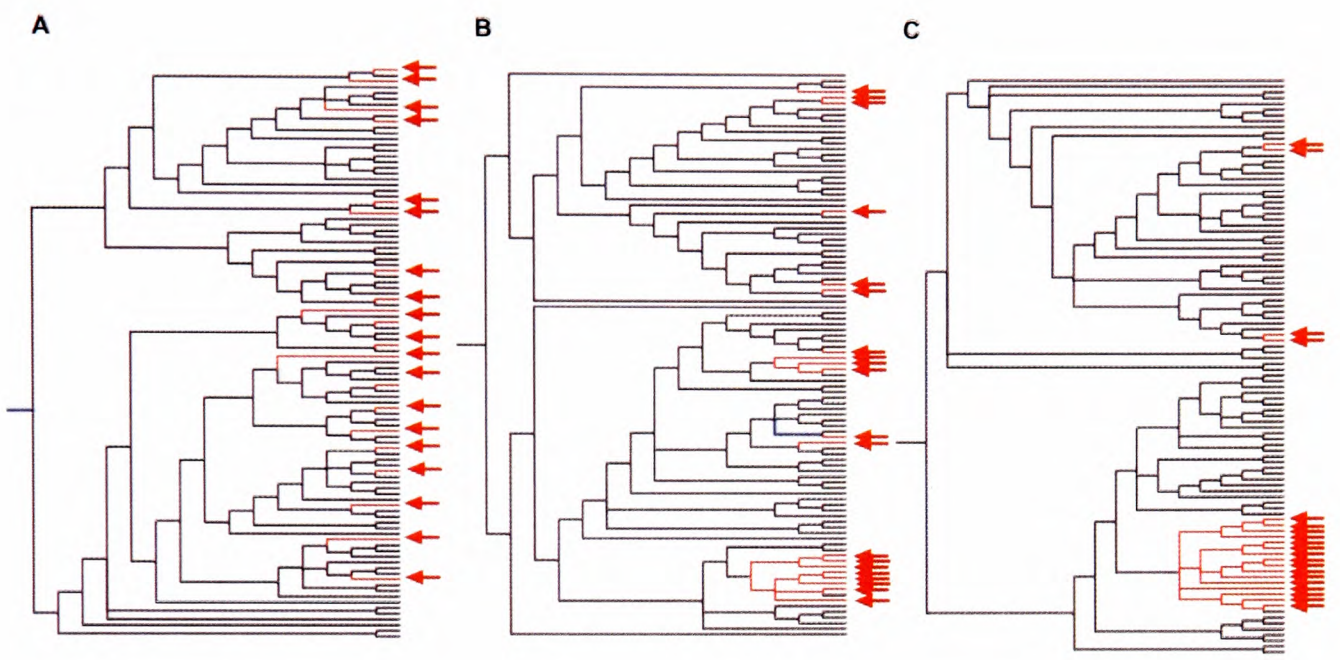


Figure 4.12: Phylogenetic clustering as a consequence of artificial imposition of HLA-B*5703 mutations on HLA-B*5703 negative Gag sequences. ML phylogenetic trees constructed from 100 Gag sequences (selected at random from a pool of HLA-B*5703-negative individuals). The same 100 sequences are represented in each tree, and the same twenty are marked with arrows. A: Twenty taxa were selected at random, marked with arrows. In this panel, no footprint mutations have been added. B: The same sequences after addition of three HLA-B*5703 footprint mutations to the arrowed sequences. C: The same sequences after addition of five footprint mutations.

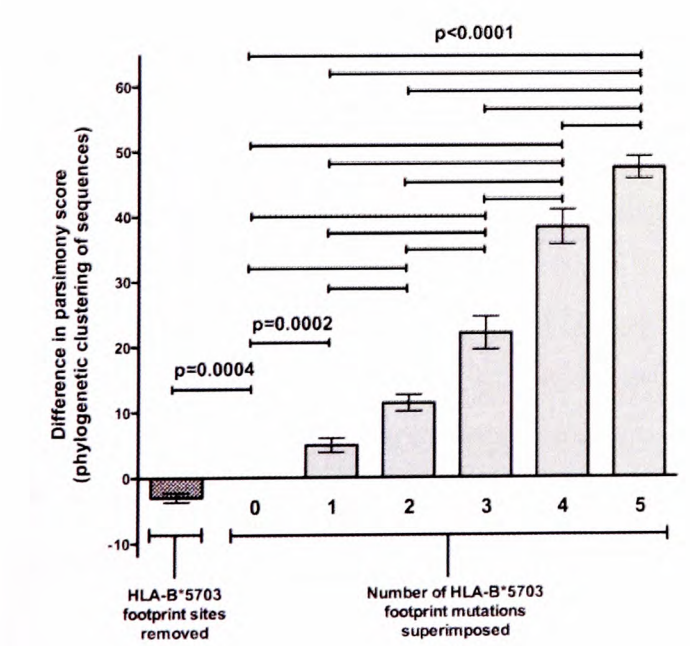


Figure 4.13: Quantification of progressive phylogenetic clustering of sequences bearing a superimposed HLA-B*57 footprint.

Quantifying phylogenetic clustering using the total available sequence pool (n=526), with a variable proportion bearing the footprint, demonstrated that the degree of clustering increases as the proportion of footprint-bearing sequences is increased (Fig. 4.14B). Even with only 5% of sequences bearing the footprint (comparable to the true population phenotypic frequency of HLA-B*5703 in Durban), there is still a significant increase in clustering as the footprint mutations accumulate (r^2=0.58, p<0.0001; linear regression).

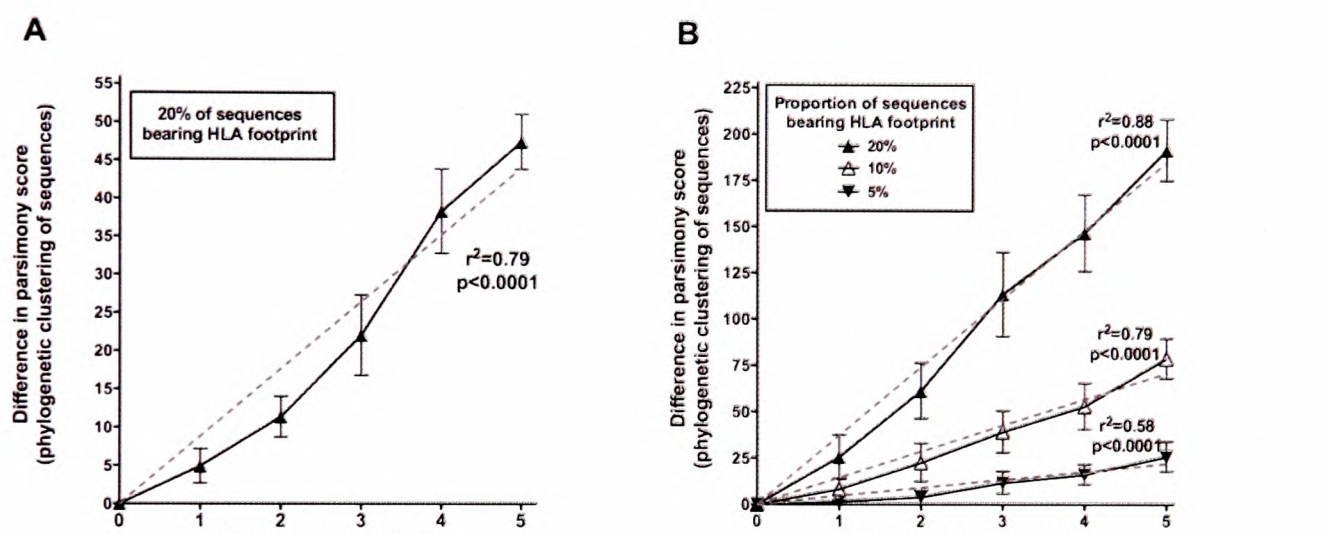


Figure 4.14: Difference in phylogenetic clustering as five HLA-B*5703 footprint mutations were superimposed on Gag sequences generated from simulated data.

Mean difference in parsimony score between trees with no mutations and trees built from the same taxa with sequential HLA-B*5703 footprint mutations superimposed is shown. Addition of mutations increases the difference in parsimony scores, reflecting progressive phylogenetic clustering. r^2 by linear regression with the y-intercept set to go through the origin. Error bars show SEM based on 10 repetitions of each simulation.

A: Clustering modelled on 100 Gag sequences selected at random from HLA-B*5703-negative individuals and used to construct an NJ phylogenetic tree. Five characteristic HLA-B*5703 footprint mutations were added, one at a time, to twenty sequences in each tree. Twenty repetitions are shown.

B: Clustering modelled on all 526 taxa from HLA-B*5703-negative subjects. The footprint mutation was added to a varying proportion of sequences.

4.3.8 Phylogenetic clustering can be mediated by an HLA-B*27 Gag footprint in chronic infection

Like HLA-B*57, HLA-B*27 is also associated with immune control of HIV, but selects a smaller footprint of three mutations in Gag, arising late in the course of infection (Goulder, Phillips *et al.* 1997; Schneidewind *et al.* 2007). Analysis of the impact of the HLA-B*27 Gag footprint arising as a result of natural selection showed, as expected, no clear evidence of phylogenetic clustering among HLA-B*2705 subjects in acute infection, since the escape mutations characteristically arise late (Goulder, Phillips *et al.* 1997) (Fig. 4.15A).

In contrast, in an analysis of sequence data from subjects with chronic infection (Schneidewind *et al.* 2007), there is clustering of taxa from HLA-B*2705 positive subjects (Fig. 4.15B), indicating that the selection pressure imposed by this allele can drive convergent evolution over time. In simulated data (an artificially superimposed HLA-B*27 footprint of three mutations), significant phylogenetic clustering was again observed in association with the footprint ($p<0.0001$, paired t-test, Fig 4.15C,D).

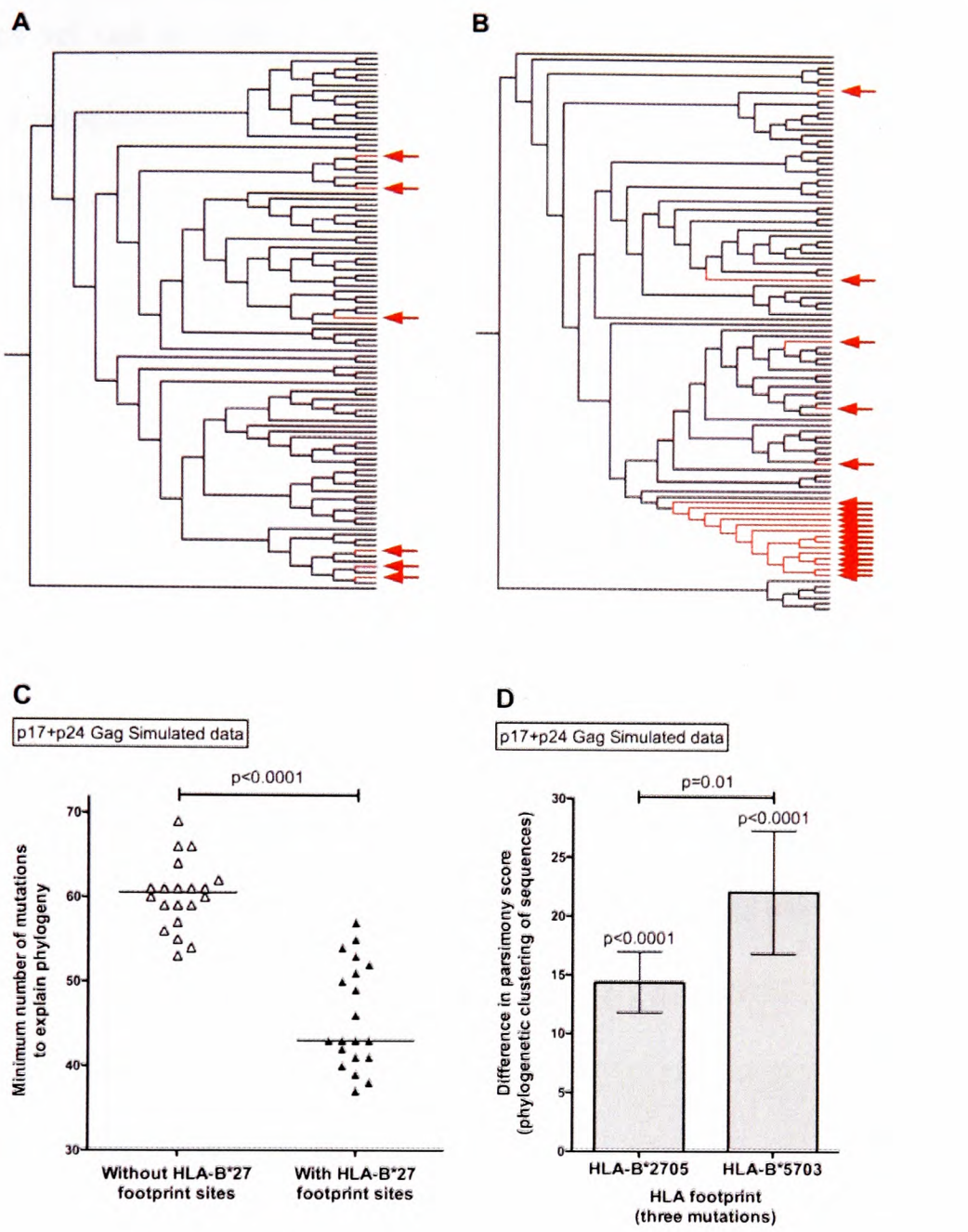


Figure 4.15: Phylogenetic clustering among taxa bearing an HLA-B*27 footprint.

A, B: Phylogenetic clustering among taxa bearing true HLA-B*27 footprint. Sequences from individuals with HLA-B*27 are marked with red arrows. A: Midpoint rooted ML phylogenetic tree of 100 subjects with acute B-clade infection. B: Midpoint rooted ML phylogenetic tree of 100 subjects with chronic B-clade infection.

C, D: Phylogenetic clustering among taxa bearing simulated HLA-B*27 footprint. C: Clustering among 20 sequences in the presence and absence of the B*27 footprint of three mutations. D: Comparison between clustering associated with HLA-B*57 and HLA-B*27 footprints of three mutations, each of which is associated with significant clustering, $p<0.0001$, paired t-test. Greater clustering is associated with HLA-B*57, $p=0.01$, Mann-Whitney test.

4.4 DISCUSSION

These studies set out to address the impact of HLA-selected mutations on HIV-1 evolution at a population level. In summary, these data demonstrate the potential role of HLA in driving HIV evolution. Evidence is shown for HLA in establishing clade-specific differences (exemplified by the AE epidemic in Thailand) and in contributing to diversification via the selection of escape polymorphisms. The phylogenetic impact of HLA in driving convergent evolution among otherwise unrelated taxa is demonstrated for HLA-B*57 and HLA-B*27.

4.4.1 Reasons why sites of difference between clades coincide with sites of HLA selection

A strong association was observed between sites of HLA selection and amino acid variability between and within clades, demonstrating that amino acid substitutions between clades can alter phylogenetic clustering and that the footprint of even a single HLA allele can cause clustering within a clade.

The observation that HIV amino acid differences between clades tend to be those that are also selected as $CD8^+$ T-cell escape mutations has two possible explanations:

i. these sites vary because of HLA selection; or

ii. sites of HLA escape are less constrained by a fitness cost to the virus.

Both explanations may contribute to the observed findings. Analysis of the CRF_AE subtype that predominates in Thailand suggests that the former explanation may operate at least in some circumstances. The Gag mutations A163G and S165N are selected by the HLA-B*5703 $CD8^+$ T cell response to the KF11 epitope (Gag 162-

172) in C-clade-infection (Crawford *et al.* 2007). When the A163G mutant arises alone, it significantly reduces viral replicative capacity *in vitro* and reverts rapidly to wild-type *in vivo* in the absence of HLA-B*5703 (Crawford *et al.* 2007). More commonly, A163G is found in association with the compensatory mutation S165N (Crawford *et al.* 2007). The observation that T242N, the one mutation with an uncompensated fitness cost, does revert suggests that the other polymorphisms (presumably with compensatory mutations) do not revert because there is no fitness cost rather than as a consequence of ongoing selection pressure. Therefore, it can be hypothesised that the founder strain was transmitted by an individual with HLA-B*5703; that is, that the AE subtype of HIV came to incorporate A163G/S165N as a consequence of HLA-B*5703 being the original driving force.

The contribution of HLA-selection to amino acid sequence variation may be substantial, with 59% of variation between Gag clade consensus sequences associated with sites of HLA-B selection. Moreover, these results are likely to be an underestimate of the true effect, as the HLA escape mutations considered here are limited to those identified in the Durban cohort (Matthews *et al.* 2008).

4.4.2 Alternative approaches to identifying positive selection

Sites that are under positive selection may also be identified by calculating the *dN/dS* ratio (de Oliveira *et al.* 2004; Delport *et al.* 2008), and this approach could be used to generate additional evidence that differences between clades relate to HLA selection. However, the result of *dN/dS* calculations vary as a function of the frequency of the selecting allele, are affected by the rate of reversion, and are difficult to apply to large

populations; for these reasons, this approach is outside the scope of this current analysis.

4.4.3 Genetic determinants of clade specificity

This analysis demonstrates that sites of amino acid difference between clades are not required to distinguish clade specificity, suggesting that clades were originally defined by nucleotide differences between founder sequences. Stripping sites of only non-synonymous nucleotide substitutions is somewhat artificial since non-synonymous and synonymous changes may arise in the same codon. However, the finding that clade clustering is preserved in the absence of sites bearing non-synonymous changes is consistent with the previous observation that all HIV clades may be traced to ancestral sequences from the same region of Africa (Vidal *et al.* 2000; Worobey *et al.* 2008), rather than occurring subsequently as a consequence of HLA selection. Nevertheless, this analysis also shows that exchange of polymorphisms between clades does have the potential to affect phylogeny, highlighting a role for HLA-selection in shaping the future evolution of the epidemic.

4.4.4 HLA as a mediator of convergent evolution

Unrelated HIV sequences within a clade may cluster together phylogenetically as a consequence of selection pressure imposed by even a single HLA allele, both in the true sequence data shown here, and in a model of serial mutations. This simulated approach is robust because the underlying sequences are altered only at footprint sites, and the mutations themselves are inferred from genuine sequence data. The model shows more phylogenetic clustering than that seen among the true sequences as a consequence of complete conservation of the mutations applied in the simulation

compared to variations in the true sequences (number of mutations, sites of mutations, and nucleotide substitutions are all conserved in the model, but may vary in real sequences). Irrespective of this, clustering mediated by the HLA-B*5703 footprint sites remains highly statistically significant in true biological sequences. The significance of this clustering effect is all the more striking when considered in the context of the enormous diversity of HIV and the potential for multiple HLA footprints to co-exist.

4.4.5 Practical applications of these results

Relevance to future statistical analysis

These observations are of potential utility when considering the use of lineage-corrected approaches in the detection of HLA-mediated selection pressure on viruses. Statistical approaches to identify associations between HIV sequence polymorphisms and HLA alleles (Moore *et al.* 2002; Kiepiela *et al.* 2004) have been refined (Bhattacharya *et al.* 2007); lineage-based correction accounts for similarities among taxa generated by their common ancestry, thus distinguishing *bona fide* HLA-escape mutations from artefactual associations mediated by founder effect (Ridley 1983; Holmes *et al.* 1992).

However, these data show that viral amino acid polymorphisms arising independently in individuals who share an HLA allele (McMichael and Klenerman 2002; Martinez-Picado *et al.* 2006; Schneidewind *et al.* 2007) may not be identified as independent mutations (homoplasies) in phylogenetic reconstruction, but instead can mistakenly appear as shared, ancestral mutations (synapomorphies). This may result in sequences that share common escape mutations being artificially grouped together during

phylogenetic reconstruction. This phylogenetic bias has previously been addressed in the setting of HIV adaptation to neutralising antibodies (Holmes *et al.* 1992) and drug therapy (Lemey *et al.* 2005), but not in the context of adaptation to $CD8^+$ T cell responses.

This confounding effect can be minimised by excluding the nucleotides under analysis (footprint sites) when constructing the phylogenetic tree. However, this requires *a priori* knowledge of the sites of escape mutation. In the absence of this information, phylogenetic clustering is likely to be reduced by maximizing the length of sequence analysed. This is relevant to many studies that carry out phylogenetic analysis using short protein fragments. Artefactual clustering is likely to be a particular problem for within-clade HIV data sets, which are characterised by many phylogenetically uninformative singleton polymorphisms.

Relevance to vaccine design

The substantial overlap between sites of amino acid and sites of HLA-driven escape mutation is of relevance to T-cell based vaccines. These sites of amino acid variability are not simply 'toggle' sites (Frahm *et al.* 2007) of little significance to T-cell recognition. On the contrary, toggling between amino acid variants may be a function of positive selection (Delport *et al.* 2008), and these data suggest that vaccine constructs may need to be matched to the clade of virus prevailing in the target population. $CD8^+$ T-cell vaccines may therefore need to be geared to the clade of virus affecting the target population, and modified over time to keep pace with evolutionary changes in the virus driven by HLA.

CHAPTER 5: Global Adaptation of HIV to selection pressure imposed by HLA.

The majority of the work presented in this chapter has been published in *Nature*. 2009;458:641-5. Adaptation of HIV-1 to HLA I. Kawashima Y*, Pfafferott K*, Frater J*, Matthews P*, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, Kuse N, Oka S, Duda A, Prendergast A, Crawford H, Leslie A, Brumme Z, Brumme C, Allen T, Brander C, Kaslow R, Tang J, Hunter E, Allen S, Mulenga J, Branch S, Roach T, John M, Mallal S, Ogwu A, Shapiro R, Prado J, Fidler S, Weber J, Pybus O, Klenerman P, Ndung'u T, Phillips R, Heckerman D, Harrigan P, Walker B, Takiguchi M, Goulder P.
(joint 1st author)*

5.1 INTRODUCTION

5.1.1 Impact of HLA selection on HIV evolution

The potential impact of HLA Class-I mediated selection on HIV evolution is demonstrated by the analyses described in Chapters 3 and 4 (Matthews *et al.* 2008; Matthews *et al.* 2009). Other authors have also discussed the evolutionary impact of accumulating HIV escape mutations on sequence changes and amino acid diversity (Goulder *et al.* 2001; Moore *et al.* 2002; Allen *et al.* 2005; Leslie *et al.* 2005). In particular, longitudinal analysis of sequence changes in a small number of acutely infected subjects identified >2/3 of all changes as being HLA-associated, highlighting the dominant role of HLA in driving HIV evolution (Allen *et al.* 2005). In this chapter, the evidence for HIV adaptation to HLA is examined in more detail, focussing on evolutionary trends across whole populations rather than in individual subjects.

Determining the population dynamics of an individual HLA-associated polymorphism can be complex, as the effect can be altered or obscured by other factors, including

the rate of accumulation of the specific HLA-mediated selection, the selection of compensatory mutations, the presence of overlapping $CD8^+$ T cell epitopes, and the reversion or persistence of mutations after transmission to HLA mismatched recipients. Founder effect also has an important impact on HIV diversity, and may confound analysis of HLA-mediated selection (Bhattacharya *et al.* 2007; Matthews *et al.* 2009).

5.1.2 Impact of HLA-selected mutations on disease control

In transmission studies, non-reverting escape mutations that persist can be associated with different disease outcomes in the recipient. In a study of HLA-B*27, maternal transmission of the Gag KK10 escape variant led to failure of viraemic control and rapid disease progression in the HLA-B*27-positive infant (Goulder *et al.* 2001). Likewise, horizontal transmission of HLA-B*57 escape variants has been associated with rapid disease progression in HLA-matched recipients, due to loss of crucial $CD8^+$ T cell epitopes (Crawford *et al.* 2009). Conversely, transmission of $CD8^+$ T cell escape mutants that reduce viral fitness *in vitro* has been shown to be associated with improved control of viraemia in HLA-mismatched recipients (Chopera *et al.* 2008; Goepfert *et al.* 2008). Overall, therefore, the long term significance of the accumulation of transmitted viral sequence changes is a balance between the potential loss of $CD8^+$ T cell mediated immune responses (favouring the virus) and the fitness costs to the virus imposed by the selection of such mutations (favouring the host) (Goulder *et al.* 2001; Crawford *et al.* 2009; Schneidewind *et al.* 2009).

At a population level, HLA-adaptation may have the consequence that alleles currently associated with favourable control of disease lose their impact on viraemic

control over time. Furthermore, there are implications for vaccine design, as clade distinctions between populations may be altered or enhanced due to the differential selection pressure imposed by the different phenotypic frequency of HLA Class I alleles in different populations (Matthews *et al.* 2009).

5.1.3 Hypothesis and aims

The studies presented in this chapter are based on the hypothesis that the frequency of a given escape mutant is correlated with the prevalence of the selecting HLA allele in that population.

In order to draw robust conclusions about adaptation of HIV to HLA, sequence numbers were maximised as far as possible, and patient cohorts from diverse locations were considered. The initial focus of the analysis was on non-reverting mutations selected by HLA Class I alleles, as the effect of accumulating polymorphisms was deemed most likely to be statistically demonstrable in the absence of reversion. The study was then expanded to include polymorphisms known to be associated with a cost to viral fitness, and therefore likely to revert following transmission. In addition, longitudinal data were collected to demonstrate the impact of evolution within a single population, and to determine the possible impact of accumulating polymorphisms on disease outcome.

Finally, analysis of another Southern African cohort (Zimbabwean subjects recruited in the UK; unpublished data) was added, to study the effects of an increased phenotypic frequency of the favourable allele HLA-B*5703 on disease outcome and viral evolution.

Together, these data are pertinent to characterising sequence changes that are likely to occur as the epidemic evolves, affecting both natural immune control, and potentially informing the selection of immunogens for vaccine design.

5.2 METHODS

5.2.1 Study cohorts

The locations of nine independent study cohorts are shown (Fig. 5.1), from which adult subjects with HIV-1 infection were recruited. All subjects were HAART-naïve. Population sequences were obtained for *gag* and *pol* genes from pro-viral DNA, and Class I HLA typing was undertaken for each subject.

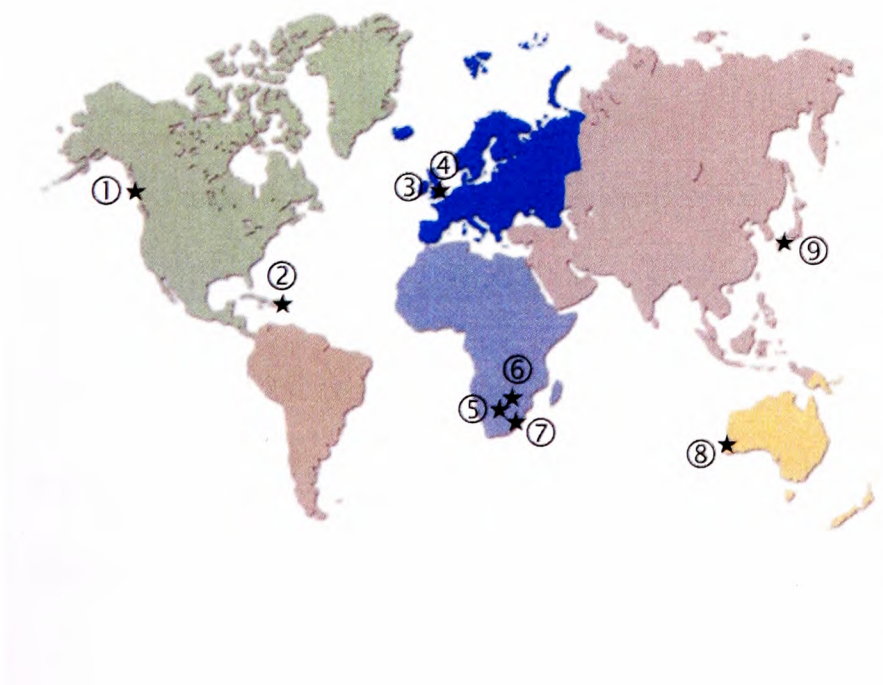


Figure 5.1: Locations of nine study cohorts used for recruitment of HIV-infected individuals. The number of subjects for whom Gag sequence data was derived is listed in brackets.

① Vancouver, Canada (n=566)
② Bridgetown, Barbados (n=53)
③ Oxford, UK (n=106)
④ London, UK (n=142)
⑤ Gaborone, Botswana (n=297)
⑥ Lusaka, Zambia (n=226)
⑦ Durban, South Africa (n=673)
⑧ Perth, Australia (n=481)
⑨ Kumamoto, Japan (n=277)

All subjects were chronically infected, apart from the London cohort that recruited acutely infected individuals*. This difference justified the inclusion of two UK cohorts (both comprising B-clade infected subjects only).

For longitudinal analysis, additional data were obtained from Japan pertaining to a cohort of 117 haemophiliacs. These subjects were infected with HIV by the transfusion of infected blood products** early in the course of the epidemic (1981-1983), and were enrolled in 1983. An additional group of 54 Japanese subjects presenting with acute infection from 1997 onwards was used for analysis of declining $CD4^+$ T cell counts over time.

Sequence data from all cohorts were determined from time points after 2000, apart from the Vancouver cohort (1996-1999), and 9 subjects in the chronic cohort from Kumamoto.

5.2.2 Selection of polymorphisms for analysis

Analysis was initially focused on HLA-associated polymorphisms predicted to accumulate at a population level; that is, HIV mutations selected early in the course of infection that do not revert post-transmission. To improve the chances of statistical detection, mutations selected by only a single HLA-Class I allele were selected for analysis. The HLA-B*51 selected mutation at the C-terminal position of the TI8 epitope (TAFTIPSI, RT 128–135) meets all of these criteria (Tomiyama *et al.* 1999; Frater *et al.* 2007; Brumme, Tao *et al.* 2008) (Fig. 5.2).

---

* SMART (Strategies for Management of Anti-Retroviral Therapy) study subjects
** Japanese haemophilia cohort likely to have been infected with American Factor VIII

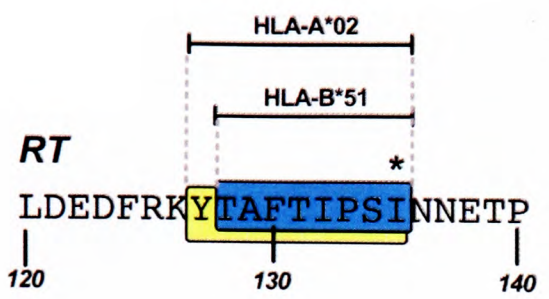


Figure 5.2: Epitope map of RT amino acids 120-140 (HXB2) showing location of HLA-B*51-restricted epitope TI8 (blue box), and associated escape mutation I135X (marked *).

The only overlapping epitope recognised at this position is YI9 (yellow box), a subdominant response restricted by HLA-A*02 (Shankar *et al.* 1998). Polymorphisms at position 135 have not been previously published in association with this response, and were not detected in multi-cohort analysis despite a high phenotypic frequency of HLA-A*02.

A further non-reverting HLA-B*51-associated mutation, Int I31V in the LI9 epitope (LPPIVAKEI, Int 28-36), was added to the analysis, as this has also been hypothesised to increase in parallel with increasing phenotypic frequency of HLA-B*51 (Leslie *et al.* 2005). In addition, data obtained from the analysis described in Chapter 3 (Matthews *et al.* 2008) were used to identify other sites of selection meeting these criteria. Three non-reverting sites in Gag that were strongly associated ($p<10^6$) with HLA-selection pressure were D260X in association with HLA-B*35, D312X in association with HLA-B*44, and S357X in association with HLA-B*07 (Table 5.1).

Following analysis of these polymorphisms that are known to accumulate in populations over time, reverting mutations were also selected for comparison. Well characterised reverting mutations were chosen for analysis on the basis of their selection by alleles that are known to be associated with control of viraemia, namely HLA-B*27 and HLA-B*57/5801. These mutations have all been associated with a fitness cost demonstrated *in vitro* (by fitness assays) or *in vivo* (by reversion) (Leslie *et al.* 2005; Martinez-Picado *et al.* 2006; Crawford *et al.* 2007; Schneidewind *et al.*

2007; Brumme, Brumme *et al.* 2008; Goepfert *et al.* 2008; Matthews *et al.* 2008). All polymorphisms selected for analysis are listed in Table 5.1.

Polymorphism category	HLA restriction & epitope name	Epitope sequence (position)	Variant
Non-reverting (persist after transmission)	HLA-B*35 PY9	PPIPVG**D**IY (Gag 254–262)	D260X
	HLA-B*44 AW11	AEQATQ**D**VKNW (Gag 306-316)	D312X
	HLA-B*07 GL9	GP**S**HKARVL (Gag 355–363)	S357X
	HLA-B*51 TI8	TAFTIPS**I** (RT 128-135)	I135X
	HLA-B*51 LI9	LPP**I**VAKEI (Int 28-36)	I31V
Reverting (return to wild-type after transmission)	HLA-B*57 ISW9	**A**ISPRTLNAW (Gag 147–155)	A146X
	HLA-B*57 ISW9	**I**SPRTLNAW (Gag 147–155)	I147X
	HLA-B*5703 KF11	K**A**FSPEVIPMF (Gag 162–172)	A163X
	HLA-B*5703 KF11	KAF**S**PEVIPMF (Gag 162–172)	S165X
	HLA-B*5703 KF11	KAFSPE**V**IPMF (Gag 162–172)	V168I
	HLA-B*57/5801 TW10	TS**T**LQEQIAW (Gag 240–249)	T242X
	HLA-B*5703 TW10	TSTLQEQ**I**AW (Gag 240–249)	I247X
	HLA-B*27 KK10	K**R**WIILGLNK (Gag 263–272)	R264X
	HLA-B*27 KK10	KRWII**L**GLNK (Gag 263–272)	L268X

Table 5.1 HLA-selected polymorphisms selected for analysis. The epitope name and sequence and position in the HIV genome (HXB2 numbering) are listed. Variants are listed by HXB2 position, with X designating >1 possible amino acid substitutions, and V/I designating substitution to valine/isoleucine only. Sites of mutation are shown in bold and underlined. HLA-B*57-associated polymorphisms Gag A146X and I147X are selected by all HLA-B*57 types, while A163X, S165X, I168V and I247X are associated only with HLA-B*5703 (Crawford *et al.* 2007; Matthews *et al.* 2008). Polymorphism at T242 is associated with all HLA-B*57 subtypes and the related allele B*5801 (Leslie *et al.* 2004; Matthews *et al.* 2008).

Of these 14 mutations selected for analysis, 9 (65%) are known to revert following transmission. This reflects the likely overall ratio of reverting:non-reverting mutations that has been demonstrated in previous population-based studies of B and C-clade infected subjects (Brumme, Tao *et al.* 2008; Matthews *et al.* 2008), including the analysis presented in Chapter 3 in which 68% of Gag polymorphisms (and 65% of all Gag, Pol and Nef polymorphisms) were identified as reverting (Matthews *et al.* 2008).

5.2.3 Statistical analysis

Statistical analysis of the relationship between HLA frequency and the prevalence of an HLA-associated polymorphism was undertaken using two approaches:

i. In order to account for the differing number of subjects in each cohort, confidence limits for the mutation frequencies were calculated (using the adjusted-Wald method at www.measuringusability.com/wald). Logistic regression was then calculated in GLMStat v.6.0.0 using a binomial error distribution.

ii. Linear regression / Spearman rank correlation were calculated in Graphpad Prism (these were unweighted for sample size).

Analysis was undertaken with the two UK cohorts considered separately, but also with the data from Oxford and London pooled in order to reduce any potential bias from using two cohorts from the same region. Initial analysis sought a correlation between the phenotypic frequency of a selecting allele and the frequency of the associated polymorphism in the whole population; this analysis was also modified to consider only subjects not possessing the selecting allele to reduce confounding caused by a potentially high frequency of the allele in the population sampled.

Co-variation analysis was undertaken using a lineage-corrected algorithm as described in Chapter 3 (Carlson *et al.* 2008; Matthews *et al.* 2008), performed by J. Carlson at Microsoft.

5.2.4 Cellular assays

These experiments were performed in Japan by Y. Kawashima. In brief, competition assays were carried out using *in vitro* infection of H9 cells*, with transmission to new cells from culture supernatant every 7 days. Sequencing of virus from supernatant was performed every 2 weeks. Viral suppression assays were undertaken using target cells expressing HLA-B*51 infected with viral variants, differing only at RT position 135. These cells were co-cultured with a TI8-specific $CD8^+$ T cell clone (effector) at differing effector:target (E:T) ratios. The impact of I135X mutations and different E:T ratios was quantified by measuring viral p24 antigen concentration by ELISA.

5.2.5 Analysis of HIV-infected subjects from Zimbabwe

In addition to the nine cohorts analysed and published (Kawashima *et al.* 2009), an additional cohort of Zimbabwean subjects was subsequently studied. This cohort was selected for additional study as it represented the largest subgroup of African patients from within the Thames Valley Cohort, reflecting a cohort that is understudied compared to other Southern African countries. In addition, in this cohort the Zimbabwe population appeared to have a high phenotypic frequency of HLA-B*5703; analysis of this cohort therefore offered opportunity for the further investigation of the way in which increased prevalence of HLA-B*57 might affect disease outcome and HIV sequence evolution.

Sixty-six treatment-naïve adult Zimbabwean subjects were recruited from the Thames Valley Cohort. For analysis of HLA phenotypic frequencies, data from 19 unlinked

* HIV-permissive immortal T lymphocytes originally isolated from a cutaneous T cell lymphoma

Zimbabwean paediatric subjects were also added. In total, high resolution HLA types were available for 79 individuals. Sequence data for adult subjects were obtained from proviral DNA, except in 3 individuals for whom RNA extraction was undertaken from plasma.

5.3 RESULTS

5.3.1 RT-I135X mutation is commonly selected by subjects with HLA-B*51

The previously documented strong association between HLA-B*51 and an escape mutation at the C-terminal position of the RT TI8 epitope (Frater *et al.* 2007; Brumme, Tao *et al.* 2008) was confirmed in the study subjects under analysis. Pooling all nine study cohorts, the I135X mutation occurred in 96.2% of subjects with HLA-B*51, compared to 28.9% of subjects without this allele (Fig. 5.3). The most commonly selected amino acid variant at RT position 135 is threonine; however, other escape mutants also occur at this position. The frequency of each polymorphism among subjects with HLA-B*51 is shown (Fig. 5.4).

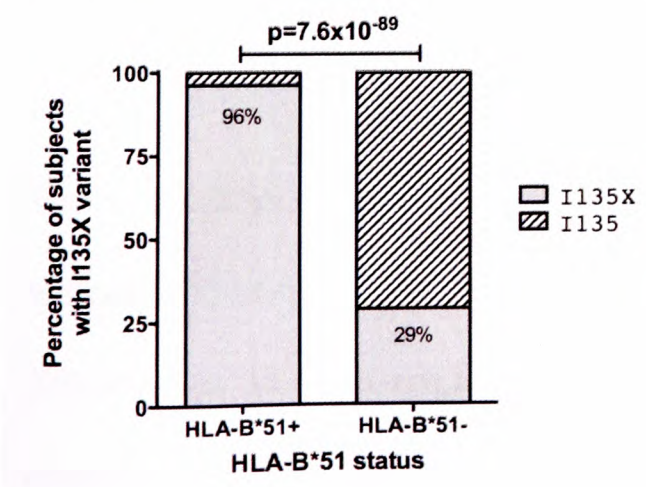


Figure 5.3: Prevalence of RT mutation I135X in HLA-B*51 positive (n=213) and HLA-B*51 negative (n=1994) study subjects from nine cohorts.

The I135X mutation is selected in the presence of HLA-B*51 ($p=7.6x10^{-89}$, Fisher's Exact Test). Mutations at this position in 29% of HLA-B*51 negative subjects reflects lack of reversion post-transmission.

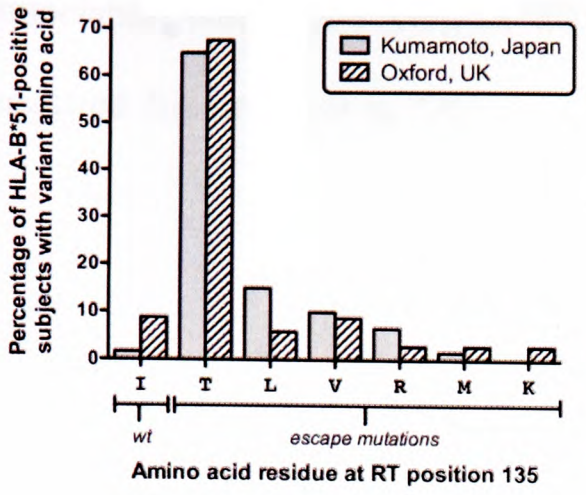


Figure 5.4: Frequency of amino acid substitutions at RT position 135 in association with HLA-B*51 in two B-clade cohorts.
94 HLA-B*51 positive subjects are represented (Kumamoto n=60; Oxford n=34). Overall, 95.7% of HLA-B*51 positive subjects have an escape mutation; wild-type (WT) isoleucine (I) is replaced by threonine (T) in 66% of cases. Other variants are L (12%), V (10%), R (5%), M (2%) and K (1%).

5.3.2 Most RT-I135X mutations do not impose a cost to viral fitness

In vitro competition assays were used to assess the impact of the different amino acid variants on viral replicative capacity (Fig. 5.5).

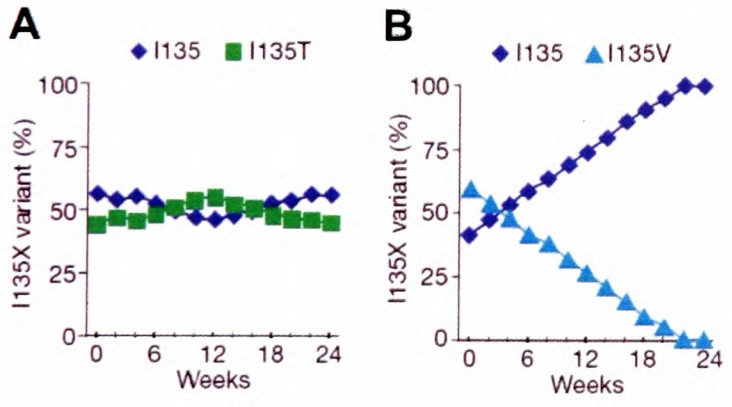


Figure 5.5: *In vitro* competition assays between NL4-3 wild-type (WT, I135) and I135X variants. A: WT competed with I135T, showing persistence of the mutant in equal proportions with WT, confirming no fitness cost in association with I135T substitution. B: WT competed with I135V, showing rapid outgrowth of WT, suggesting a significant fitness cost imposed by I135V.
Experiments undertaken by Y. Kawashima in Kumamoto, Japan.

The observation that I135T does not revert following transmission to an HLA-mismatched recipient (Brumme, Tao *et al.* 2008) was replicated in longitudinal follow-up of 38 B*51-negative subjects, all of whom acquired an I135X mutation at transmission. In keeping with *in vitro* data pointing to a fitness cost only in

association with I135V (Fig. 5.5B), reversion was observed only in subjects who acquired this variant (Fig. 5.6).

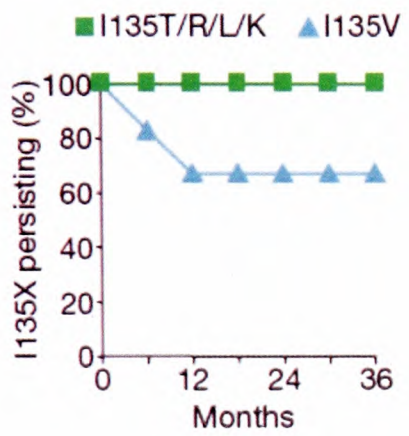


Figure 5.6: Longitudinal follow up of 38 B*51-negative subjects who acquired an RT-I135X mutation.
Japanese subjects who acquired a virus with T, R, L or K at RT-135 showed no change in sequence over 36 months of follow-up, suggesting no significant cost to viral fitness imposed by these polymorphisms. In contrast, subjects infected with a strain bearing 135V reverted to WT in 40% of cases. In parallel with *in vitro* data, these observations point to a fitness cost in association with valine.
Analysis undertaken by M. Takiguchi in Kumamoto, Japan.

5.3.3 RT-I135X mutations are associated with amino acid co-variation

From these data, it appears that valine is selected (in a minority of cases) as an escape mutation at position 135 despite a fitness cost to the virus. In order to identify whether this mutation occurs in conjunction with compensatory mutations that reduce this impact on viral fitness, amino acids that co-vary with position 135 were identified (Table 5.2).

Co-variation observed at these positions may be accounted for by a variety of mechanisms:

i. Independent selection of other HLA-B*51 escape mutations, especially in regions that overlap or flank known epitopes restricted by this allele.
ii. Selection of polymorphisms by HLA alleles transmitted in LD with HLA-B*51.

iii. Compensatory mutations that reduce a fitness cost of the index HLA-B*51-selected mutation. Co-varying residues that lie within 20 angstroms of the index residue in the resolved protein structure are considered most likely to interact structurally with this residue (Carlson *et al.* 2008).

Amino acid residue at 135	Co-varying residue	p	q	Sequence	Distance (angstroms)
L	RT 276	0.0005	0.09	W A S Q I Y P G I K **V** R Q L C K L L R G T	26.5
	RT 386	5.30E-06	0.003	T T E S I V I W G K **T** P K F K L P I Q K E	9.5
R	RT 98	7.45E-05	0.02	W E V Q L G I P H P **A** G L K K K K S V T V	9.9
	RT 123	0.0006	0.09	D A Y F S V P L D E **D** F R K Y T A F T I P	30.3
	RT 223	0.002	0.18	G L T T P D K K H Q **K** E P P F L W M G Y E	25.1
	RT 334	0.002	0.17	D L I A E I Q K Q G **Q** G Q W T Y Q I Y Q E	31.9
T	RT 121	6.56E-05	0.02	V G D A Y F S V P L **D** E D F R K Y T A F T	29.0
	RT 123	9.77E-05	0.03	D A Y F S V P L D E **D** F R K Y T A F T I P	30.3
	RT 214	0.0005	0.09	E L R Q H L L R W G **L** T T P D K K H Q K E	21.1
	RT 215	0.001	0.15	L R Q H L L R W G L **T** T P D K K H Q K E P	25.0
	RT 456	0.001	0.13	A N R E T K L G K A **G** Y V T N R G R Q K V	57.8
V	RT 162	0.0004	0.08	G W K G S P A I F Q **S** S M T K I L E P F R	16.2
	RT 360	0.0002	0.05	K T G K Y A R M R G **A** H T N D V K Q L T E	27.4
	RT 452	0.001	0.15	V D G A A N R E T K **L** G K A G Y V T N R G	57.4
	RT 483	0.001	0.16	T N Q K T E L Q A I **Y** L A L Q D S G L E V	50.5

Table 5.2: Amino acid residues in RT that co-vary with RT-135. Positions of HLA-B*51 epitopes are highlighted (grey bars). The distance between residue 135 and the covarying residue in the folded protein structure is given in angstroms, calculated by J. Carlson (Carlson *et al.* 2008), based on previously published crystal structure ((Rodgers *et al.* 1995) Fig. 5.7). Instances of co-variation with I135X occurring within 20 angstroms are shown in red. Unpublished B-clade data from HOMER cohort / Western Australia provided by J. Carlson and Z. Brumme (pers. comm.).

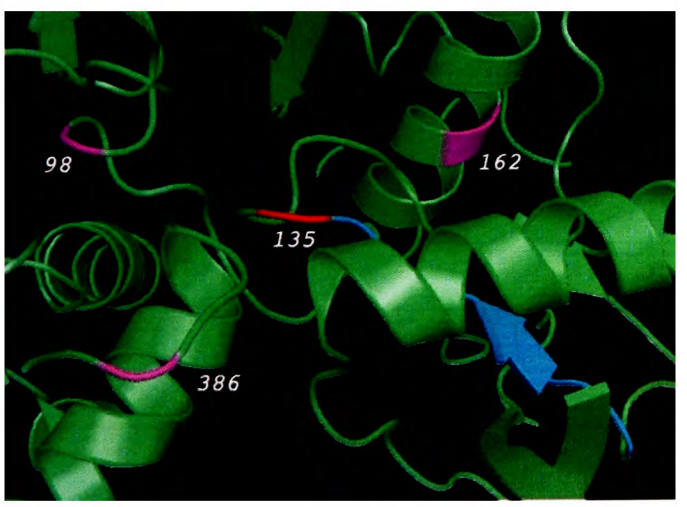


Figure 5.7: Cartoon showing HIV RT TI8 epitope (cyan) with mutation at position I135 (red) in close spatial proximity (≤20 angstroms) with co-varying residues at positions 98, 162 and 386 (magenta). Figure reconstructed in MacPymol from RT crystal structure solved by Rodgers (Rodgers *et al.* 1995).

Substitution at RT positions 98, 162 and 386 occur in association with I135X and are within 20 angstroms, suggesting these as possible compensatory mutations (Fig. 5.7). However, further investigation of these co-varying positions using *in vitro* fitness assays is required to confirm a compensatory role.

5.3.4 RT-135X mutations allow escape from $CD8^+$ T cell recognition

The effect of these HLA-B*51 selected polymorphisms on $CD8^+$ T cell recognition of infected cells was also studied *in vitro*. At high effector:target ratios, viral suppression by the TI8-specific $CD8^+$ T cell line was markedly better in the presence of WT virus compared to in the presence of viruses bearing an I135X mutation (Fig. 5.8). These data point to a reduction in $CD8^+$ mediated immune control in subjects with HLA-B*51 following the selection of escape mutations at this position.

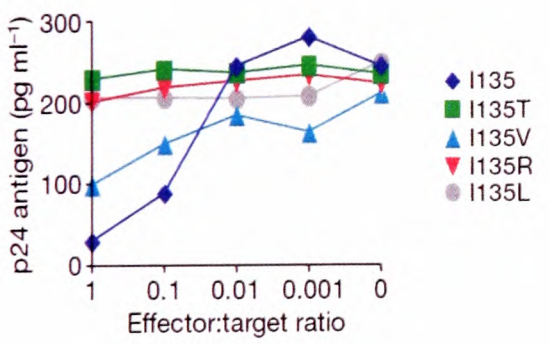
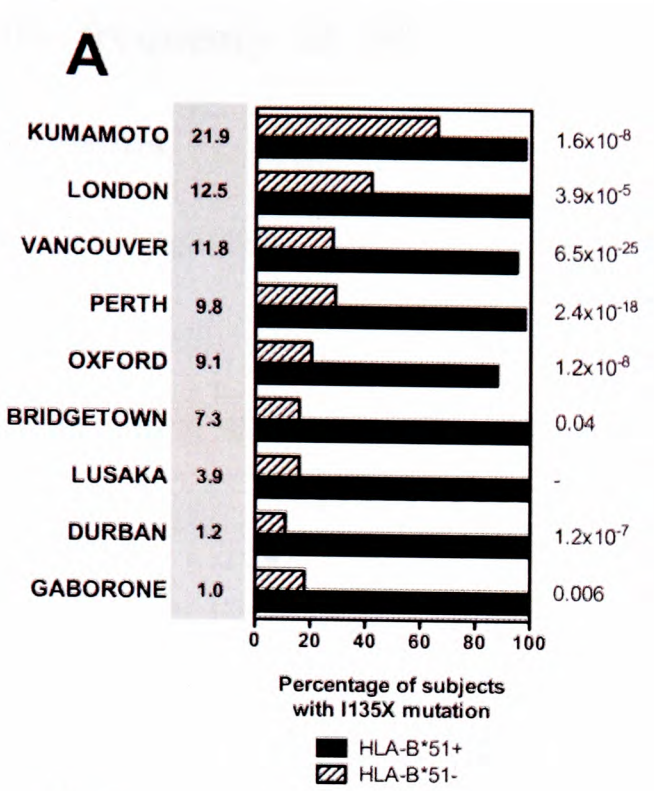


Figure 5.8: Control of viral growth by TI8-specific CTL according to variation at RT position I135 in the infecting virus.
Experiments undertaken by Y. Kawashima in Kumamoto, Japan.

5.3.5 The prevalence of RT-I135X mutations correlates with the phenotypic frequency of HLA-B*51

Having demonstrated an association between HLA-B*51 and selection of the I135X mutation in the RT TI8 epitope, and confirmed that these viral mutations allow *in vitro* escape from TI8-specific $CD8^+$ T cell control, the impact of this HLA Class I selection on circulating viruses in nine different study cohorts was assessed (Fig. 5.9).



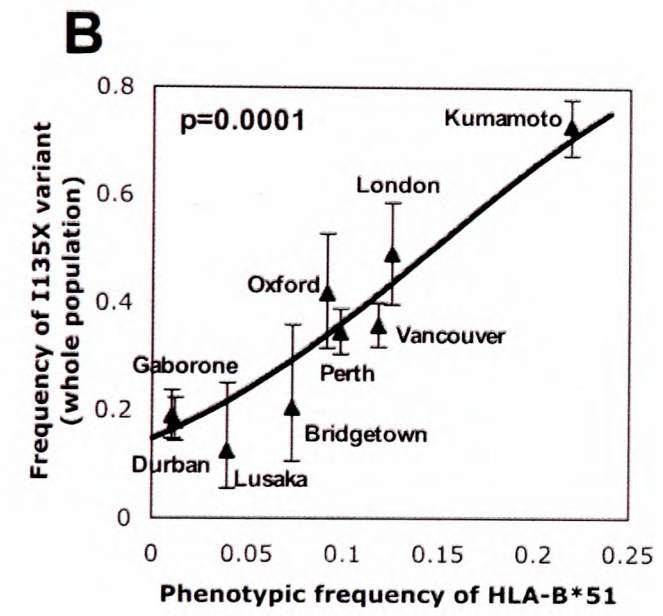
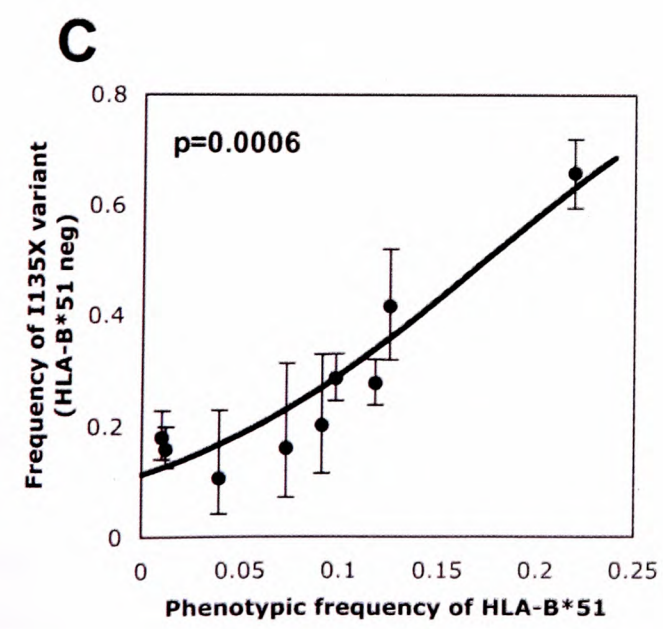


Figure 5.9: Relationship between the phenotypic frequency of HLA-B*51 and the frequency of B*51 associated escape mutation I135X in nine study cohorts.

A. Nine study cohorts are shown in descending order of HLA-B*51 frequency (% phenotypic frequency shown in grey bar). Percentage of subjects with RT mutation I135X in the B*51 restricted epitope TI8 is quantified in the presence and absence of B*51. P values are shown for each cohort (Fisher's Exact Test), apart from Lusaka in which only 1 subject with HLA-B*51 was identified.

B. There is a strong correlation between increased frequency of HLA-B*51 and increased selection of the escape variant. (p=0.0001 logistic regression)



C. The same correlation was also undertaken using the frequency of I135X mutations in HLA-B*51-negative subjects. A significant correlation remained (p=0.0006, logistic regression).

5.3.6 The prevalence of HLA-selected polymorphisms correlates with frequency of the selecting allele for persisting and reverting mutations

The correlation between the population frequency of HLA-B*51 and the presence of HLA-B*51 selected polymorphisms provides initial evidence for the adaptation of HIV to HLA-mediated selection pressure. In order to demonstrate this effect operating more widely, further examples were studied, both in four other non-reverting polymorphisms, and in nine polymorphisms that do revert post-transmission. In the majority of these, a significant relationship was observed, as with HLA-B*51-I135X selection, between the frequency of the selecting allele and the prevalence of the escape mutant (Table 5.3; Fig. 5.10).

HLA restriction and epitope name	Variant	Linear regression model			Logistic regression model	
		r	p	Pooled UK (r,p)	p	Pooled UK (p)
HLA-B*35 PY9	D260X	0.81	0.008	0.84, 0.009	0.016	0.003
HLA-B*44 AW11	D312X	0.3	0.436	0.26, 0.539	0.08	0.103
HLA-B*07 GL9	S357X	0.67	0.051	0.65, 0.082	0.004	0.005
HLA-B*51 TI8	I135X	0.94	0.0002	0.95, 0.0004	0.0001	<0.0001
HLA-B*51 LI9	I31V	0.84	0.017	0.84, 0.017	0.001	0.001
HLA-B*57 ISW9	A146X	0.38	0.317	0.36, 0.378	0.048	0.068
HLA-B*57 ISW9	I147X	0.87	0.002	0.91, 0.002	0.007	0.007
HLA-B*57 KF11	A163X	0.8	0.031	0.80, 0.031	0.11	0.11
HLA-B*57 KF11	S165X	0.46	0.303	0.46, 0.303	0.364	0.364
HLA-B*57 KF11	V168I	0.84	0.019	0.84, 0.019	0.085	0.085
HLA-B*57 TW10	T242X	0.99	<0.0001	0.83, 0.010	<0.0001	0.0003
HLA-B*57 TW10	I247X	0.91	0.004	0.91, 0.004	0.041	0.041
HLA-B*27 KK10	R264X	0.85	0.003	0.92, 0.001	0.003	0.004
HLA-B*27 KK10	L268X	0.93	0.0002	0.94, 0.0005	0.001	0.003
All polymorphisms		0.69	<0.0001	0.69, <0.0001	<0.0001	<0.0001

Table 5.3: Relationship between phenotypic frequency of HLA and the population frequency of 14 different HIV escape variants in nine cohorts. P values by linear regression and logistic regression.

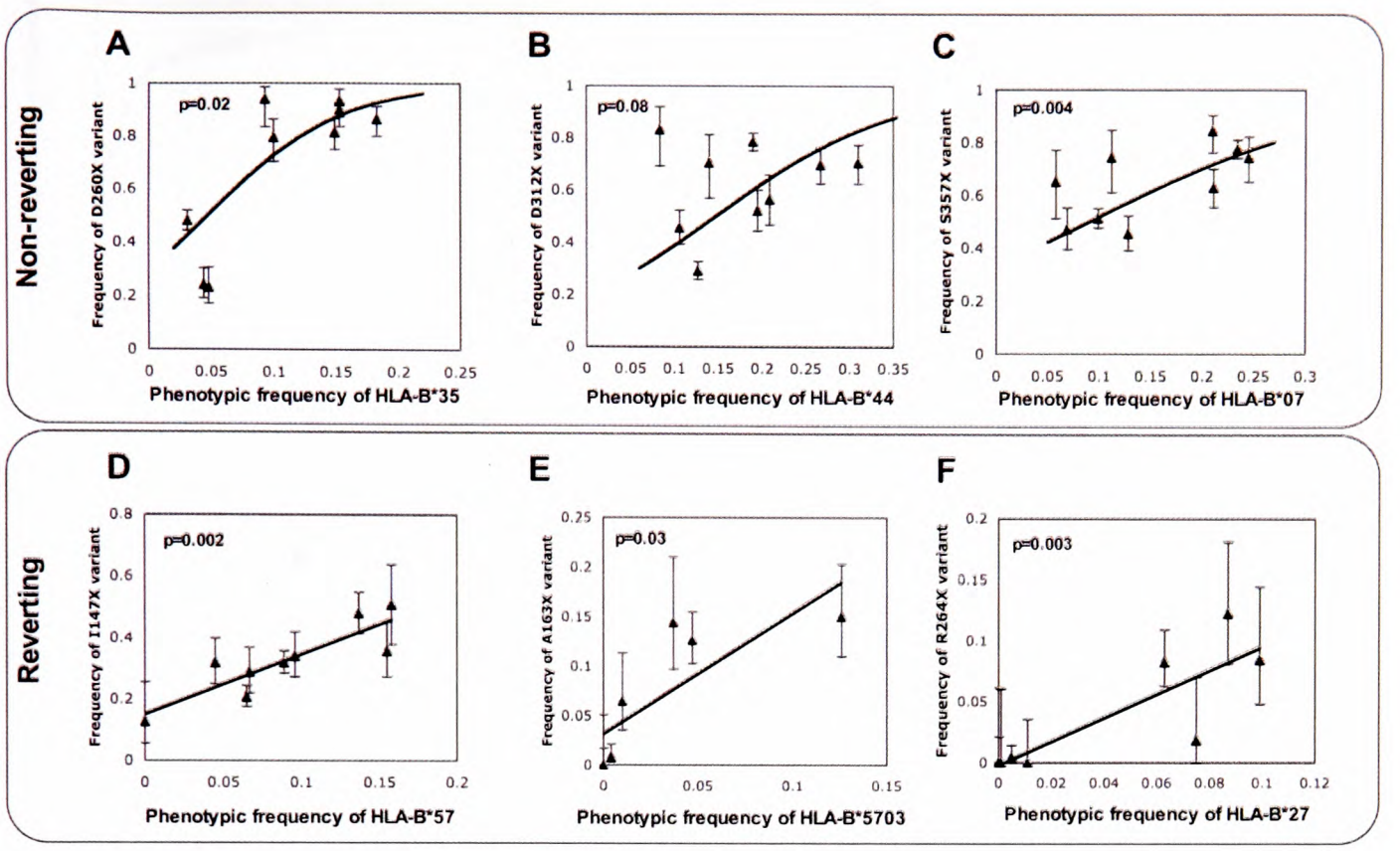


Figure 5.10: Relationship between phenotypic frequency of HLA and the population frequency of HIV escape variants at six positions in Gag. Panels A, B and C: Non-reverting mutations (p values by logistic regression); panels C, D and E: Reverting mutations (p values by linear regression). A: HLA-B*35 prevalence and the frequency of D260X escape variant; B: HLA-B*44 prevalence and the frequency of D312X escape variant; C: HLA-B*07 prevalence and the frequency of S357X escape variant; D: HLA-B*57 prevalence and the frequency of I147X escape variant; E: HLA-B*5703 prevalence and the frequency of A163X escape variant; F: HLA-B*27 prevalence and the frequency of R264X escape variant.

5.3.7 Models of HIV adaptation to HLA

As shown above (Table 5.3) logistic regression is the better model for four of five non-reverting polymorphisms (D312X, S357X, I135X, I31V), while the linear regression model better fits the curves for the majority of reverting polymorphisms (the only exception being A146X). This section explores the relationship between these mathematical models and the pattern of HLA-mediated selection (Fig. 5.11).

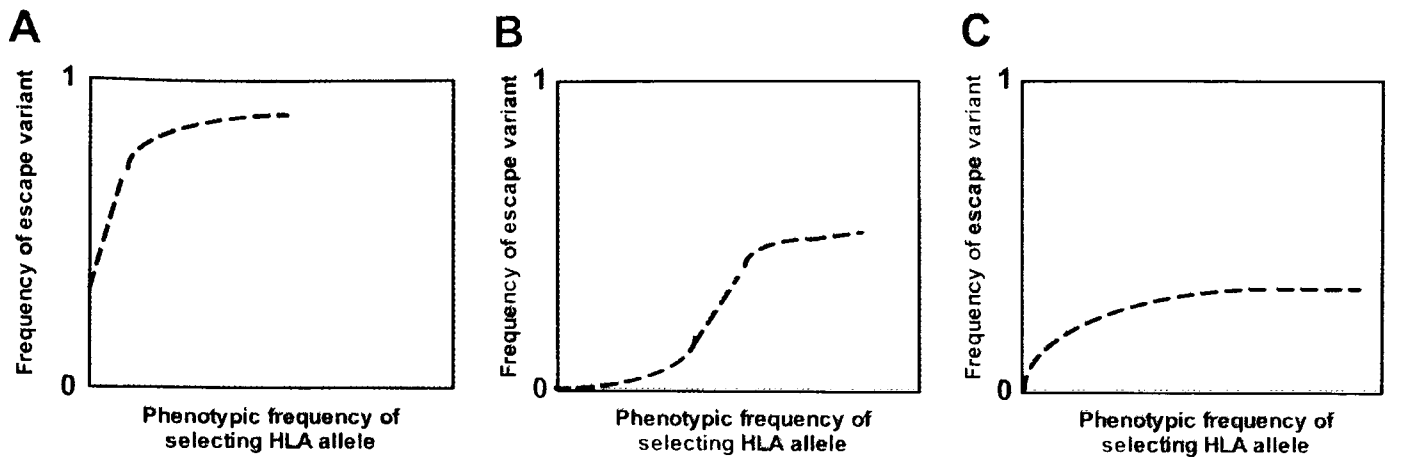


Figure 5.11: Models of different patterns of the accumulation of HLA escape variants according to the phenotypic frequency of the selecting HLA allele.

A: The amino acid variant circulates at high frequency even when the selecting allele is absent. In the presence of selection, the frequency of the variant increases rapidly before levelling off at a high frequency. This pattern is generally characteristic of non-reverting mutations (e.g. B*35-D260X, Fig. 5.9A), and best modelled by logistic regression.

B: The amino acid variant is not observed in the absence of the selecting allele. As the selecting allele becomes more frequent, the frequency of the variant increases more gradually (e.g. B*27-R264X, in which the mutation is selected late in the course of infection; Fig. 5.9F), and the curve is likely to be better fitted by logistic than linear regression.

C: In the presence of the selecting allele, the variant population increases rapidly, but only to a limited plateau. This is characteristic of reverting mutations, (e.g. B*5703-A163X, Fig. 5.9E), and the curves are better fitted by linear than logistic regression.

Not all HLA-selected polymorphisms fit perfectly into any of these models, as many inter-related factors can affect the dynamics of viral adaptation, including:

i. Rate of selection of escape and reversion
ii. Selection of co-varying compensatory mutations
iii. Presence of overlapping HLA epitopes
iv. Characteristics of study population (e.g. sample size, stage of disease)
v. Other selection forces on the virus (e.g. HAART)

However, these models can help to explain the differences between adaptation observed in the examples studied here, and can also account for different results in statistical tests of significance according to the two methods used.

5.3.8 The frequency of all escape mutations combined is strongly correlated with the population frequency of the selecting allele

Pooling the data for all 5 non-reverting mutations, 9 reverting mutations, and all 14 mutations, a strongly significant correlation was observed between the phenotypic frequency of the selecting allele and the prevalence of the mutant circulating in the population (Fig. 5.12; Table 5.3).

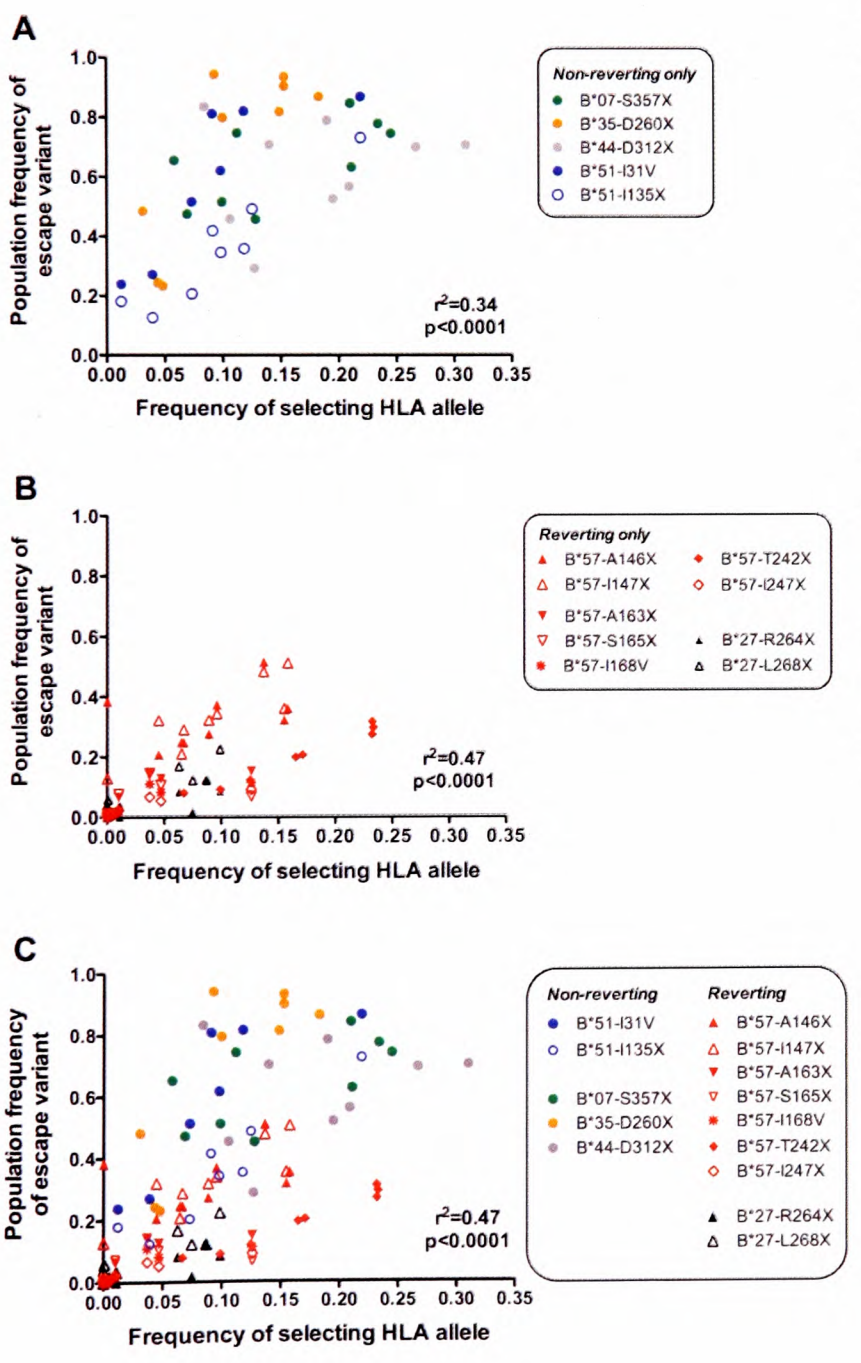


Figure 5.12: Correlation between phenotypic frequency of selecting allele and percentage of escape variant in whole population for 14 HLA-associated polymorphisms.

A: Non-reverting mutations only;
B: reverting mutations only;
C: all mutations.

$p<0.0001$ by linear regression and logistic regression for all three panels.

5.3.9 Founder effect does not explain the relationship between HLA prevalence and HIV amino acid polymorphisms

Founder effect is a potentially confounding effect; the widespread presence of a given polymorphism in a particular population could be related to the ancestral sequence(s) founding that epidemic, rather than as a direct consequence of a particular HLA allele at high frequency in that population (Moore *et al.* 2002; Bhattacharya *et al.* 2007). The influence of this potential bias is here addressed in three ways:

i. The polymorphisms studied have largely been identified as escape mutations in well-characterised epitopes, confirming that the polymorphism is HLA-selected rather than arising from a founder sequence.

ii. The previous statistical analysis used to determine sites of HLA-selection was phylogenetically corrected for viral lineage (Matthews *et al.* 2008).

iii. Sequence data used in this study were subjected to repeat analysis for HLA-associated polymorphisms in the presence and absence of phylogenetic correction (table 5.4); all associations remained significant after correction.

HLA restriction & epitope name	Variant	Uncorrected Fisher's (p)	Phylogenetically corrected (p)
HLA-B*35 PY9	D260X	$3.8x10^{-17}$	$1.8x10^{-9}$
HLA-B*44 AW11	D312X	$4.6x10^{-16}$	$3.1x10^{-12}$
HLA-B*07 GL9	S357X	$6.7x10^{-24}$	$8.7x10^{-19}$
HLA-B*51 TI8	I135X	$1.55x10^{-52}$	$2.9x10^{-45}$
HLA-B*51 LI9	I31V	$1.5x10^{-7}$	$9.2x10^{-3}$
HLA-B*57 ISW9	A146X	$8.2x10^{-20}$	$1.7x10^{-9}$
HLA-B*57 ISW9	I147X	$3.6x10^{-23}$	$7.5x10^{-6}$
HLA-B*57 KF11	A163X	$1.3x10^{-21}$	$1.8x10^{-18}$
HLA-B*57 KF11	S165X	$3.7x10^{-12}$	$5.7x10^{-6}$
HLA-B*57 KF11	V168I	ns	ns
HLA-B*57 TW10	T242X	$3.9x10^{-64}$	$2.1x10^{-45}$
HLA-B*57 TW10	I247X	$5.7x10^{-7}$	$4.9x10^{-6}$
HLA-B*27 KK10	R264X	$4.0x10^{-10}$	$2.8x10^{-9}$
HLA-B*27 KK10	L268X	$7.1x10^{-10}$	$3.4x10^{-10}$

Table 5.4: Fisher's exact test and phylogenetically tested p values for 14 HLA associations with sequence polymorphisms in HIV Gag and Pol. Data shown are for cohorts from Vancouver, Perth, Kumamoto, Durban and Gaborone only.

5.3.10 RT-I135X mutations accumulate longitudinally

The sequence data analysed thus far are all cross-sectional, but lead to the inference that HLA-selected mutations accumulate longitudinally over time in proportion to the frequency of the selecting allele. In order to demonstrate this longitudinal trend directly, data from two Japanese cohorts, separated by 14-25 years, were examined. Comparing a cohort of haemophiliacs infected between 1981 and 1983, with a cohort infected through sexual transmission presenting after 1997, a significant increase in the likelihood of becoming infected with an I135X mutant was demonstrated (Fig. 5.13).

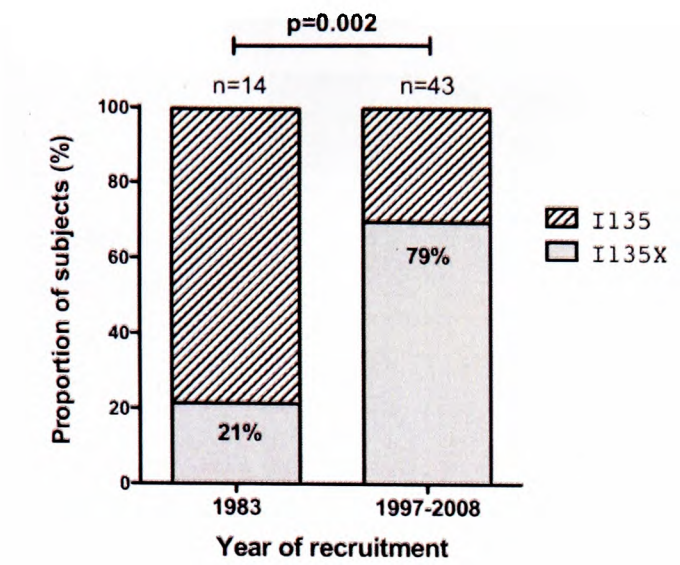


Figure 5.13: Frequency of HLA-B*51-associated I135X mutation in HLA-B*51 negative subjects from a Japanese cohorts diagnosed in 1983, and a subsequent cohort presenting between 1997 and 2008.

In 1983, 21% of subjects had sequences bearing the I135X mutation, compared to 79% presenting from 1997 onwards (p=0.002, Fisher's Exact Test), confirming a longitudinal change in the prevalence of circulating I135X strains in this population.

5.3.11 Class I alleles historically associated with disease control may lose their impact over time

The possible *in vivo* impact of accumulating escape mutations is that $CD8^+$ T cell epitopes that play a crucial role in containment of disease may be lost over time, leading to an alteration in relationships between HLA Class I alleles and control of viraemia. Analysis of sequence data to determine the effect of the I135X mutation on the course of disease in subjects with HLA-B*51 is limited by the extremely high prevalence of escape among these subjects (>95%, Fig. 5.3). In order to assimilate evidence for the changing impact of HLA-B*51 on disease outcome, longitudinal clinical data were examined.

The data in table 5.5 demonstrate that these two cohorts are not directly comparable (due to highly significant difference in median viral load, and borderline significant difference in CD4 count) between the groups. However, the effect of HLA-B*51 on viral load and CD4 count within each cohort can be assessed to determine whether this allele has an impact on disease control (Fig. 5.14).

	1983 cohort			1997-2008 cohort			Difference
	HLA-B*51+	HLA-B*51-	All	HLA-B*51+	HLA-B*51-	All	(p)
n	25	90	115	102	482	584	-
Median VL (RNA copies/ml)	1200	7050	5100	31000	56000	50500	<0.0001
Median CD4 (cells/mm^3)	387	274	276	260	235	242	0.05

Table 5.5: CD4 count and Viral load data from two Japanese cohorts, presenting in 1983 and between 1997-2008. Data are taken from the baseline timepoint for each subject. Statistical difference between cohorts is calculated using data from all patients (irrespective of HLA-B*51 type) using a Mann-Whitney Test.

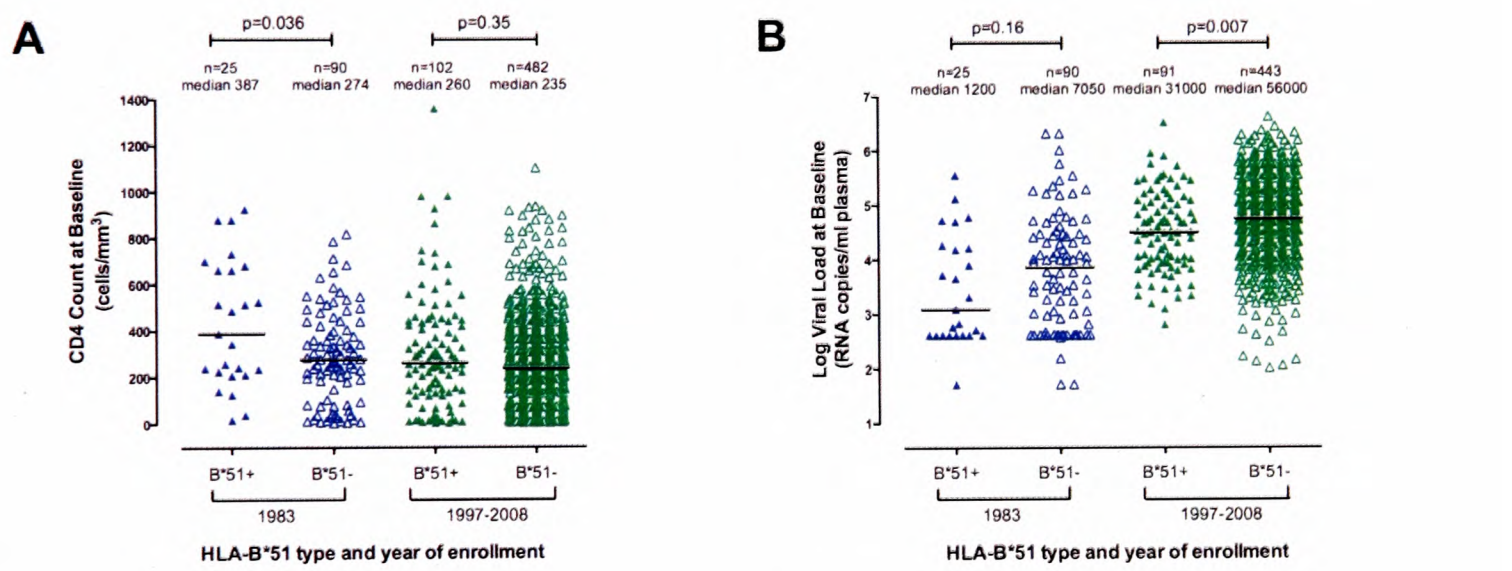


Figure 5.14: Clinical data from Japanese cohort presenting in 1983 compared to cohorts recruited between 1997-2008. A: Baseline CD4 count, B: Baseline viral load.

HLA-B*51 was associated with a preserved $CD4^+$ T cell count in subjects infected early in the course of the epidemic, presenting in 1983 (p=0.036), but this effect appears to have been diminished over time, as subjects presenting after 1997 had no $CD4^+$ T cell count improvement in the presence of HLA-B*51 (Fig. 5.13A). Viral load data show a trend towards lowered set-point in association with HLA-B*51 in the earlier cohort (viral load 0.8log lower in the presence of this allele), although this does not reach statistical significance. Conversely, subjects presenting later did have a

significant benefit in the presence of HLA-B*51, although the difference had lessened to 0.3log lowering of viral load.

Further data to substantiate loss of the benefit of HLA-B*51 comes from Kaplan Meier analysis of the duration of time to starting HAART. In these acutely-infected subjects, there was no benefit of HLA-B*51; in fact, the trend was towards a negative impact of this allele (Fig. 5.15).

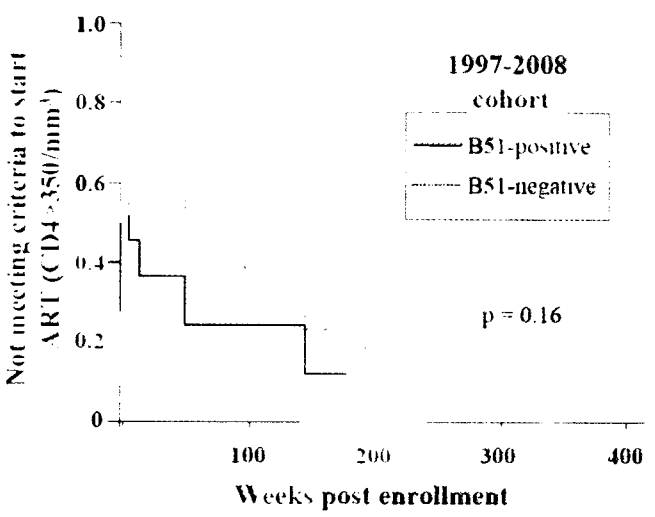


Figure 5.15: Kaplan Meier plot showing duration of time prior to meeting criteria for HAART according to HLA-B*51 status in an acute cohort of 54 subjects presenting after 1997.
In this cohort, there was no significant benefit in preservation of $CD4^+$ T cell count in association with the presence of HLA-B*51 (n=11).

In contrast to the above analysis, HLA-B*51-positive subjects from the 1983 cohort were relatively protected from disease; 4 of 25 subjects with HLA-B*51 remained treatment-naïve for ≥5 years after enrollment in the cohort, compared to 4 of 91 subjects in the absence of HLA-B*51 (p=0.06, Fisher's exact test).

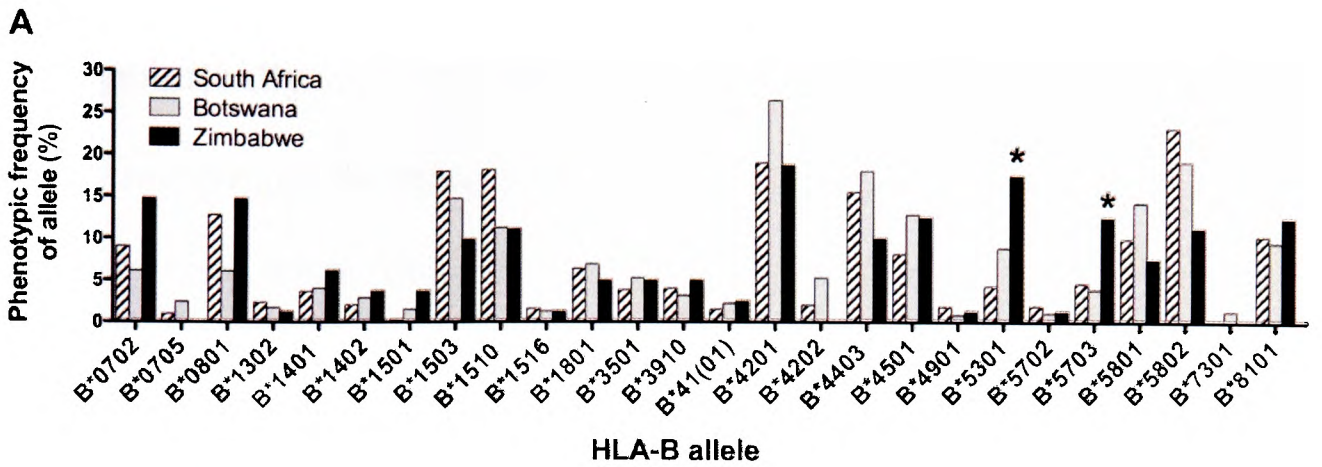
These data are consistent with the accumulation of HLA-B*51-selected mutations and associated loss of HLA-B*51-mediated disease control over time, but do not unequivocally demonstrate it.

5.3.12 HLA-B*5703 is enriched in HIV-infected subjects in Zimbabwe compared to other Southern African cohorts

In the remaining part of this chapter, analysis is made of a new cohort of HIV-infected Zimbabweans, in order to further explore the effect of HLA-B*57 footprints on HIV sequences and disease control.

HLA-B*5301 and B*5703 were statistically enriched in Zimbabwe compared to Botswana and South Africa (Fig. 5.16A). HLA-B*5703 was the most consistently significantly different ($p \leq 0.005$ for comparison with both South Africa and Botswana, Fisher's Exact Test, Fig. 5.16B), although this does not retain statistical significance if subjected to a rigorous Bonferroni correction.

The small size of the Thames Valley Cohort made it difficult to assign HLA allele frequencies with confidence. Comparison was therefore made with allele frequencies in another Zimbabwean population; this showed statistically comparable phenotypic frequencies of HLA-B*5301 and B*5703 (Table 5.6).



B

HLA allele	p value (Fisher's Exact Test) comparing phenotypic frequency to other cohorts	
	South Africa	Botswana
B*0702	0.02	0.11
B*0801	0.02	0.6
B*1501	0.17	0.004
B*5301	0.03	0.00002
B*5703	0.004	0.005

Figure 5.16: Comparison of phenotypic frequency of HLA-B alleles in three Southern African Cohorts. Subjects from Zimbabwe (n=79), from South Africa (Durban; n=1212) and from Botswana (Gaborone ante-natal; n=424).

A: Bar chart showing all alleles occurring at >1% phenotypic frequency in ≥1cohort, and found in ≥2 study subjects. Alleles marked * are statistically more frequent in Zimbabwe than in both South Africa and Botswana ($p<0.05$, Fisher's Exact Test). B: Comparison between Zimbabwe and two other Southern African cohorts, showing all alleles with any p value <0.05, Fisher's exact test.

HLA allele	Thames Valley, Zimbabwe HIV+ (allele frequency, %)	Shona population, Harare HIV- (allele frequency, %)	p value (Fisher's Exact Test)
B*5301	9.9	9.3	0.75
B*5703	6.2	4.4	0.4

Table 5.6: HLA allele frequencies* of two HLA-B alleles occurring at a greater frequency in Zimbabwe than in South Africa or Botswana. Data derived from HIV-infected subjects recruited from Thames Valley Cohort (n=79), and HIV-negative subjects reported by www.allelefrequencies.net (n=230). A greater frequency of HLA-B*5703 in the Thames Valley Cohort (although not statistically significant) may reflect enrichment for protective alleles among HIV-infected subjects presenting as out-patients in the UK.

* Allele frequency rather than phenotypic frequency is used in this table, as www.allelefrequencies.net only lists the former data. Elsewhere in this thesis, phenotypic frequency is routinely used.

5.3.13 HIV taxa from Southern Africa cluster together phylogenetically

In order to determine the extent of phylogenetic diversity within and between subjects from Zimbabwe, South Africa and Botswana, maximum likelihood phylogenetic trees were constructed from nucleotide sequences (Fig. 5.17). For Gag and Nef, some country-specific subclustering was apparent, particularly for subjects from South Africa. Taxa from South Africa and Botswana appeared more phylogenetically distinct, while Zimbabwean sequences were intermingled with the other cohorts. This may reflect genuine differences in the phylogeny of infecting strains, but could also be attributable to the demography of subjects recruited in each cohort; for South Africa and Botswana, recruitment took place in geographically constrained areas, while Zimbabwean subjects are likely to be more epidemiologically diverse. To some extent, these data reflect the results of a previous study that reported little evidence of population-specific clustering in Southern Africa (Bredell *et al.* 2007); however, further investigation with additional sequences from defined cohorts in Zimbabwe is warranted.

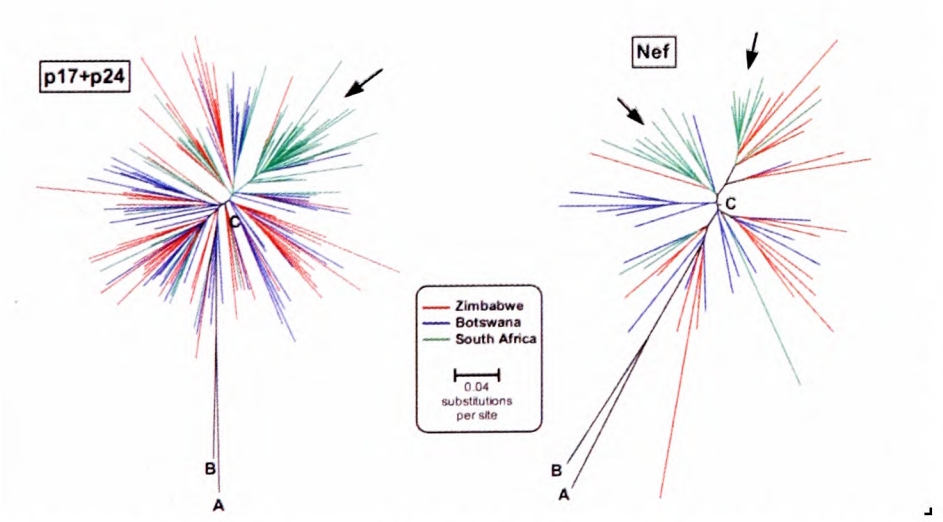


Figure 5.17: Maximum likelihood phylogenetic trees constructed in GARLI using taxa from subjects from South Africa, Botswana and Zimbabwe. Consensus sequences from clades A, B and C are marked in each case.

A: Gag (p17+p24) sequences; 50 selected at random from each cohort. B: Nef sequences; 20 selected at random from each cohort. Distinct South African subclusters indicated with arrows.

5.3.14 Equivocal impact of HLA-B*5703 on viral loads in Zimbabwe

Based on the increased phenotypic frequency of HLA-B*5703 in Zimbabwe compared to other cohorts, the viral loads in Zimbabwean subjects were compared with other Southern African (C-clade infected) individuals from the Thames Valley Cohort. There was no statistical difference in median viral load at baseline (p=0.8, Mann-Whitney test, Fig. 5.18A). UK (B-clade infected) subjects had significantly higher viral loads (p=0.03), possibly reflecting the documented difference between B and C-clade infection (Spira *et al.* 2003; Abraha *et al.* 2009). In Zimbabweans and other Southern African subjects, there was a trend towards lowering of viraemia in association with HLA-B*5703 (Fig. 5.18B).

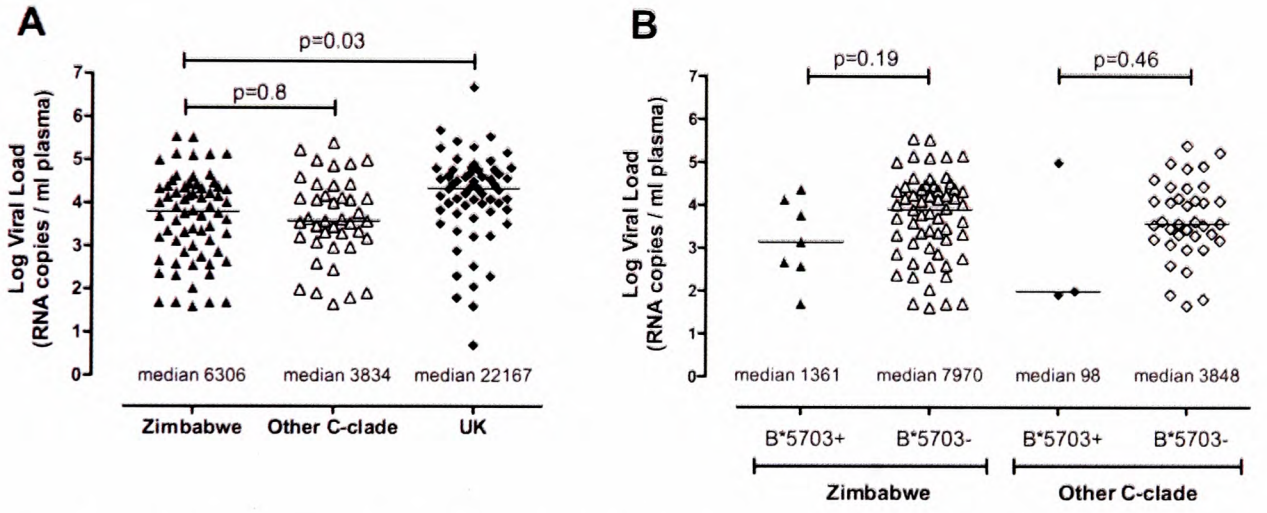


Figure 5.18: Comparison of baseline viral loads in Thames Valley cohort subjects from Zimbabwe with other geographic origins (A), and according to the presence and absence of HLA-B*5703 (B).

The absence of a statistical effect of HLA-B*57 in lowering viral loads in Zimbabwe may relate to a reduction in potency of this allele due to its increased phenotypic frequency, but the study numbers here are too small to draw firm conclusions.

5.3.15 Sites of entropy difference between Zimbabwe and other Southern African cohorts coincide with HLA-B*57 footprint sites

To compare the extent of sequence variability in Zimbabwe with that in South Africa and Botswana, Shannon entropy scores were calculated for each residue in p24 Gag (Fig. 5.19). Sites within HLA-B*57 epitopes contained sites of significant entropy difference; most notably, two sites of HLA-B*57 escape mutations in the KF11 epitope, positions 163 and 165, were more variable in Zimbabwe than South Africa, possibly reflecting the significance of the increased frequency of HLA-B*57 in Zimbabwe in driving sequence evolution.

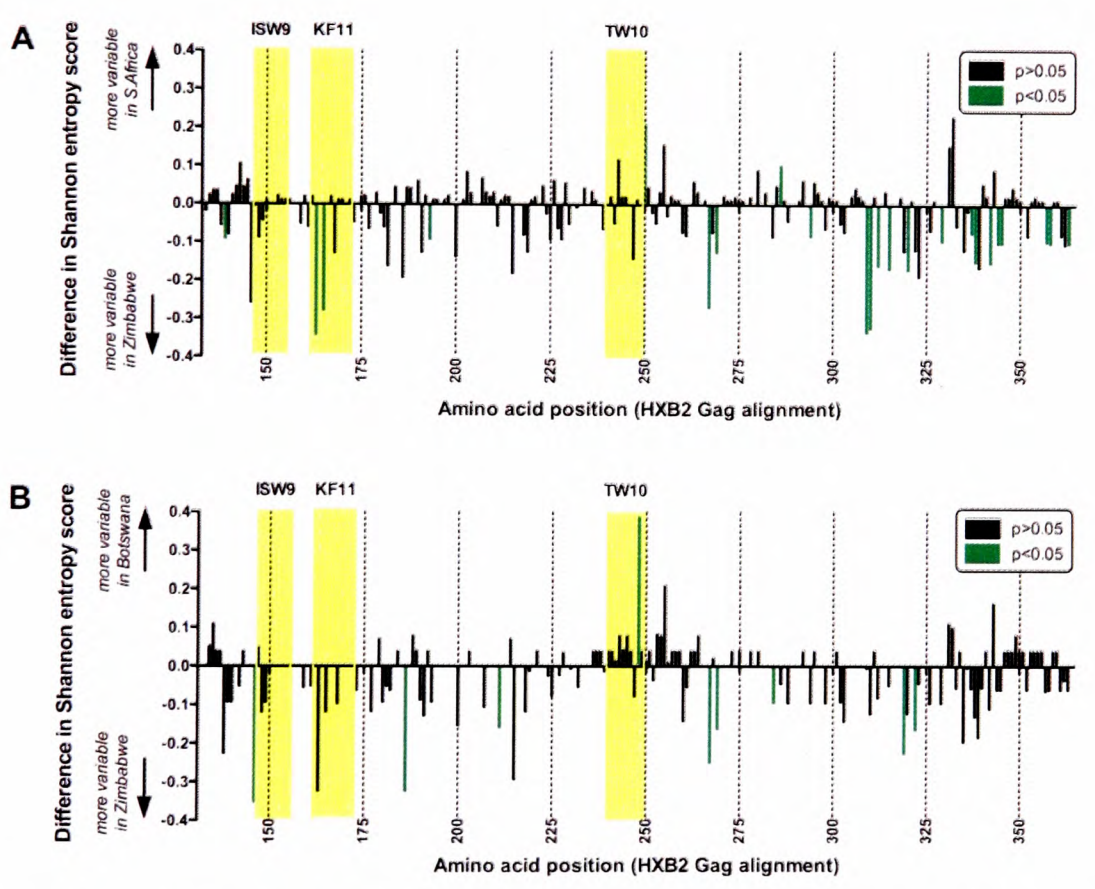


Figure 5.19: Difference in p24 entropy scores between Zimbabwe and South Africa (A) and Botswana (B). Sites that are statistically significantly different are marked in green; HLA-B*57 epitopes are highlighted in yellow.

5.3.16 HLA-B*57 footprints in Zimbabwe occur in proportion to the phenotypic frequency of HLA-B*57

Data for the Zimbabwean cohort were added to the previous analysis of the relationship between the frequency of the selecting allele and the population prevalence of the escape variant. For all seven HLA-B*57-selected polymorphisms in Gag, the Zimbabwean data conformed to the trends seen in the other study cohorts (Fig. 5.20). Linear and logistic regression were calculated for all 12 polymorphisms in Gag. Correlations were universally stronger following the addition of this tenth cohort (Table 5.7).

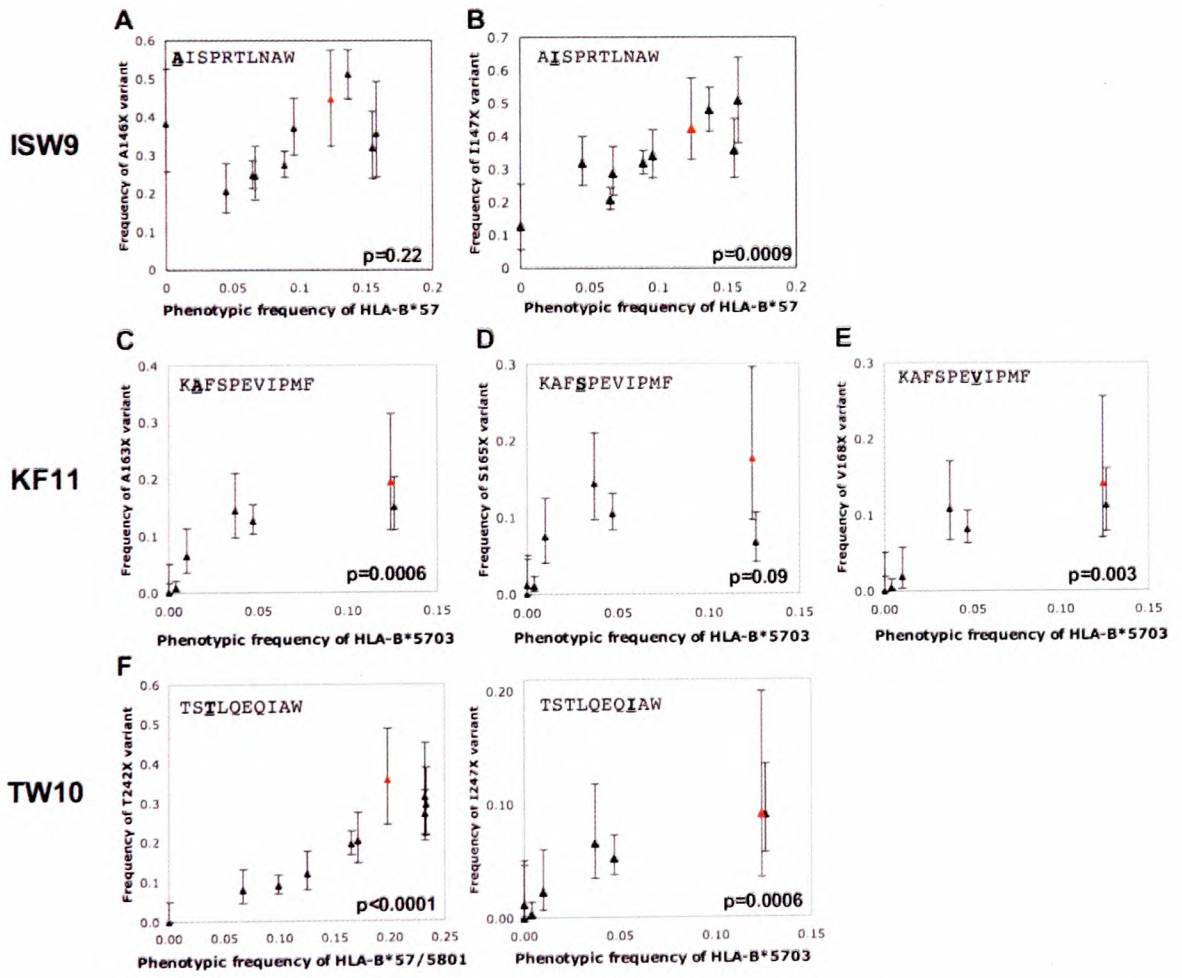


Figure 5.20: Relationship between phenotypic frequency of HLA and the population frequency of HIV escape variants at six positions in Gag for nine previously studied cohorts plus Zimbabwean cohort. Data from nine cohorts, as previously described, shown in black; Zimbabwean data in red, using phenotypic frequencies from Thames Valley Cohort (n=79). p values by linear regression.

HLA restriction and epitope name	Variant	Linear regression model				Logistic regression model	
		Minus Zimbabwe		Plus Zimbabwe		Minus Zimbabwe	Plus Zimbabwe
		r	p	r	p	p	p
HLA-B*35 PY9	D260X	0.81	0.008	0.83	0.003	0.016	0.0007
HLA-B*44 AW11	D312X	0.3	0.436	0.27	0.456	0.08	0.075
HLA-B*07 GL9	S357X	0.67	0.051	0.57	0.08	0.004	0.003
HLA-B*57 ISW9	A146X	0.38	0.317	0.42	0.22	0.048	0.029
HLA-B*57 ISW9	I147X	0.87	0.002	0.88	0.0009	0.007	0.003
HLA-B*57 KF11	A163X	0.86	0.031	0.87	0.006	0.11	0.066
HLA-B*57 KF11	S165X	0.46	0.303	0.65	0.09	0.364	0.239
HLA-B*57 KF11	V168I	0.84	0.019	0.89	0.003	0.085	0.047
HLA-B*57 TW10	T242X	0.99	<0.0001	0.94	<0.0001	<0.0001	<0.0001
HLA-B*57 TW10	I247X	0.91	0.004	0.94	0.0006	0.041	0.021
HLA-B*27 KK10	R264X	0.85	0.003	0.81	0.004	0.003	0.002
HLA-B*27 KK10	L268X	0.93	0.0002	0.91	0.0002	0.001	0.0005

Table 5.7: Relationship between phenotypic frequency of HLA and the population frequency of 12 different HIV escape variants in Gag in original nine cohorts, plus additional Zimbabwean cohort. P values by linear regression and logistic regression.

5.3.17 $CD8^+$ T cell responses to Gag HLA-B*57 epitopes are preserved in Zimbabwe

Based on the increased phenotypic frequency of HLA-B*5703 in Zimbabwe compared to South Africa, and the corresponding increase in frequency of escape mutations selected by this allele, ELISpot data were examined for evidence of differences in $CD8^+$ T cell responses to Gag peptides containing HLA-B*57-restricted epitopes between these two cohorts (Fig. 5.21). There were no significant differences between cohorts on the basis of responses to OLPs, either across the whole cohort, or within HLA-B*57 positive subjects.

These data suggest that HLA-B*57-restricted $CD8^+$ T cell responses in Zimbabwe continue to target Gag epitopes – at the present time - despite the increased frequency of HLA-B*57 escape mutations in this population compared to South Africa. Loss of responses would only be expected in subjects bearing HLA-B*57 escape mutations;

these footprints may not yet have accumulated to the extent of population-wide loss of HLA-B*57-restricted $CD8^+$ T cell responses.

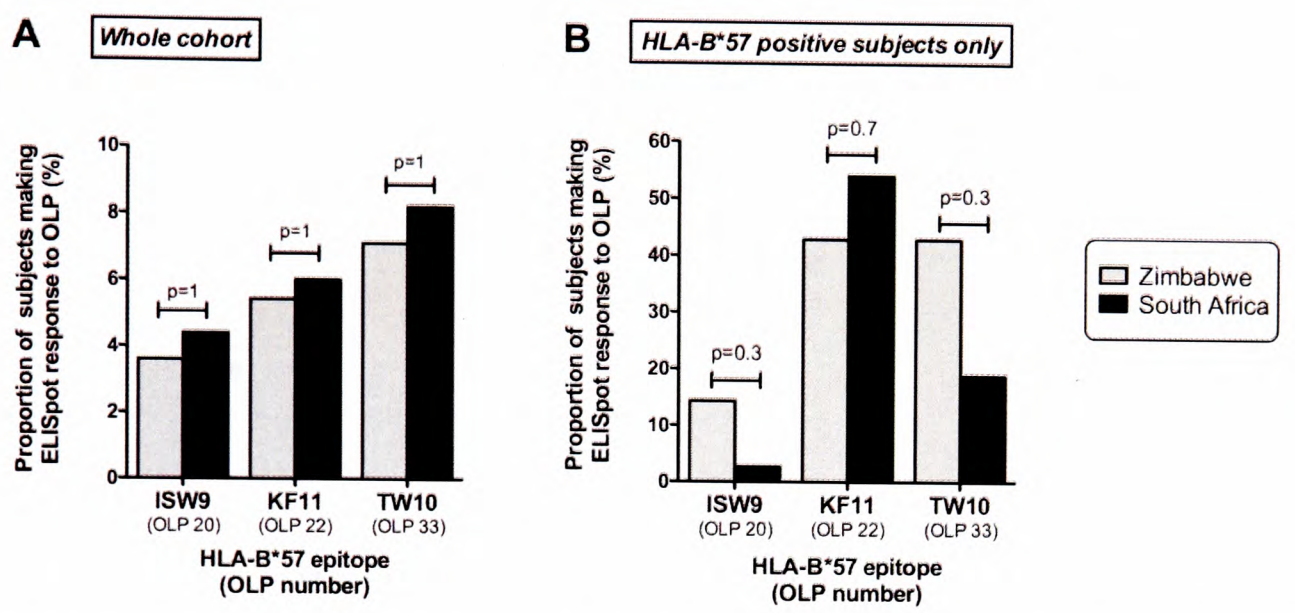


Figure 5.21: Proportion of subjects in cohorts from Zimbabwe (Thames Valley Cohort) and South Africa (Durban cohort) making ELISpot responses to Gag OLPs containing HLA-B*57 epitopes. A: Whole cohorts, B: HLA-B*5703-positive subjects only.

5.4 DISCUSSION

5.4.1 Evidence for HIV adaptation to HLA

The rapid evolution of the HIV epidemic provides a rare opportunity to undertake studies of host-pathogen co-evolution. Overall, the cross-sectional data presented here demonstrate that HLA-selected polymorphisms are present in different populations according to the frequency of the selecting allele, suggesting that – within a given population – adaptation of the virus to prevalent $CD8^+$ T cell immune responses is occurring. These effects are not limited to mutations that persist following transmission, but are also seen in association with reverting polymorphisms.

5.4.2 HLA-B*51 selected mutation RT-I135X contributes to disease control

Initial data presented here confirm the well-recognised association between HLA-B*51 and selection of I135X escape mutations in the RT TI8 epitope (Tomiyama *et al.* 1999; Frater *et al.* 2007; Brumme, Brumme *et al.* 2008). Cellular assays confirm that the B*51-restricted TI8 response contributes to control of viral replication in an *in vitro* system, although the precise mechanism(s) through which this allele mediates disease control *in vivo* remain unknown. As the escape mutation I135X is selected early in the majority of subjects with HLA-B*51, it is likely that disease control hinges on factors other than just the response to this single RT epitope. The mechanism for selection of I135V, that imposes a fitness cost on the virus, is uncertain but covariation at RT-162 is identified as a putative compensatory mutation.

5.4.3 Challenges in identifying factors driving HIV evolution

Differentiating between the selective forces that operate to drive HIV evolution, in addition to accounting appropriately for underlying founder effect, is difficult. The HLA-selected mutations studied here have been clearly defined in previous studies, apply to known epitopes, and persist following lineage-correction, validating the assumption that the polymorphisms are genuinely HLA-selected and are not arising as a result of founder effect.

Epitope clustering that leads to the selection of escape mutations at the same position can potentially confound statistical analyses aiming to correlate the prevalence of a single HLA allele with the frequency of an escape polymorphism. This effect has been described with reference to HLA-B*4801-LI10 (LQGQMVHQAI, Gag 138-

147) that overlaps with the HLA-B57-ISW9 epitope (Gag 147-155) (Kawashima *et al.* 2009). Both of these alleles select for escape mutations at A146, as do other alleles including B*1302, B*1510, B*3910 and B*5801, leading to a reduction in the correlation between any single HLA allele and this polymorphism.

5.4.4 Relationship between reverting mutations and HIV evolution

It is surprising that mutations that are known to impose a cost to viral fitness (and therefore revert following transmission) persist in populations to the extent demonstrated here. Accumulation of these polymorphisms is dependent on the rate of selection exceeding the rate of reversion. This effect has been demonstrated in analysis of longitudinal sequences from the acute London cohort (data not shown), in which the rate of selection of the Gag T242X escape mutant in the presence of HLA-B*57 exceeds the rate of reversion of this mutation in HLA-B*57-negative subjects (Kawashima *et al.* 2009). In the presence of strong selective forces that operate early in the course of infection (Leslie *et al.* 2004; Duda *et al.* 2009; Kearney *et al.* 2009), and reversion that is delayed (possibly by the co-selection of compensatory mutations (Crawford *et al.* 2007; Carlson *et al.* 2008; Schneidewind *et al.* 2008)), even mutations that impose a cost to viral fitness can be seen to accumulate in populations over time.

5.4.5 Impact of viral adaptation to HLA on disease control

The ongoing process of viral adaptation to prevalent immune responses may alter previously characterised relationships between host HLA Class I phenotype and disease outcome. Specifically, over time, the impact of this phenomenon of accumulating escape mutations is likely to be loss of certain $CD8^+$ T cell epitopes,

and therefore potential associated reduction in viraemic control associated with favourable alleles, such as HLA-B*51, HLA-B*27 and HLA-B*57.

Longitudinal data from Japan provide preliminary evidence for this effect, demonstrating that the favourable impact of HLA-B*51 seen early in the epidemic may be diminishing over time in parallel with the increased frequency of B*51-selected escape mutants in circulation. However, as the mechanism of the protective effect in association with HLA-B*51 has not been completely defined (Frater *et al.* 2007), it is impossible to draw precise conclusions as to why this diminution of disease control might occur over time. Future longitudinal analyses would be of great benefit in assessing the relationship between accumulation of polymorphisms and alterations in disease control.

The alterations in disease control that are seen in association with HLA-B*51 are paralleled by studies of transmission pairs focussing on other alleles. For example, the protective effect of HLA-B*57 may be diminished in a new host if infected with a strain bearing B*57 escape mutants (Crawford *et al.* 2009), and a similar effect has been observed in the SIV-macaque model (Seki *et al.* 2008). Evidence for altered HIV disease control in a cohort with increased phenotypic frequency of HLA-B*57 was examined here by analysis of a cohort of infected Zimbabweans. In this setting, despite the increased prevalence of escape mutations, HLA-B*57-restricted T cell responses to Gag were preserved. In this population, therefore, there is no evidence to suggest loss of successful HLA-B*57-restricted responses in the context of an increased frequency of this allele at the current time. Longitudinal studies, and

screening using optimals rather than OLPs, would be informative in demonstrating alteration of this observation over time.

5.4.6 Long term implications of HIV adaptation to HLA

Questions remain as to the long-term implications of these findings as the epidemic evolves. The rate and extent of HIV adaptation to $CD8^+$ T cell responses is difficult to assess, and is likely to be altered particularly by the increasingly widespread introduction of HAART. Transmission and accumulation of compensatory mutations (Crawford *et al.* 2007; Carlson *et al.* 2008) might also influence the dynamics of accumulation of $CD8^+$ T cell escape mutations.

As the precise correlates of immune protection remain unknown, the implications of these data for vaccine immunogen selection are uncertain. However, to optimise the chances of eliciting a broad response, a vaccine construct is likely to bear most chances of success if it is well matched to the circulating strains of virus. As demonstrated here, viral consensus sequences are likely to change over time, differ between populations, and adapt to selective forces. All of these observations suggest that a successful vaccine may need to be tailored to the population in which it is used, and to be modified over time to keep pace with viral evolution.

CHAPTER 6: The effect of HLA-B*35 on disease control in HIV

6.1 INTRODUCTION

6.1.1 Rationale for the study of HLA-B*35 in HIV

HLA-B*35 was selected as a Class I allele of interest for further studies based on initial analyses in which it was found to be associated with the selection of multiple HIV polymorphisms (Chapter 3), and to operate as a potential selecting force for HIV sequence evolution in Gag (Chapter 5). These observations point to a strong immunological selection pressure imposed by HLA-B*35 alleles.

HLA-B*35 alleles are represented in most populations, varying from an estimated phenotypic frequency of 5-10% in most Asian and African populations to 20% in Caucasoid populations (Marsh *et al.* 2000). The majority of published studies of this allele have focused on B-clade infected subjects; the large cohorts of study subjects with C-clade infection in the cohorts available here offered a valuable opportunity for further investigation.

6.1.2 HLA-B*35 subtypes and peptide binding

HLA-B*35 is divided into high resolution subtypes, the most common of which are B*3501, B*3502, and B*3503 in the B and C-clade cohorts studied here. The architecture of the B-pocket (the determinant of binding at position 2 of the epitope) is common to B*35 and B*53 alleles, with a preference for binding small, hydrophobic residues (Dorrell *et al.* 2001). The F-pocket is more spacious, permitting a broader

range of possible C-terminal residues (Dorrell *et al.* 2001), with a preference for large and/or hydrophobic residues (Honeyborne *et al.* 2006). Collectively, B*35 alleles have been classified as B*35Py and B*35Px (Table 6.1). The differences in epitope binding are accounted for by minor differences in amino acid sequence of the HLA class I allele (Table 6.2).

	HLA alleles	Position 2	C terminus
B*35Py	B*3501	**P** (A, V)	**Y** (F, M, L)
B*35Px	B*3502 B*3503 B*5301	**P** (A)	**M, L** (F)

Table 6.1: Classification and binding motifs of HLA-B*35 alleles. Preferred amino acids for binding at HLA anchor positions are shown in bold, other options in brackets (Marsh *et al.* 2000; Gao *et al.* 2001; Honeyborne *et al.* 2006; Escobar *et al.* 2008).

Allotype	Residue																						
	16	24	45	63	67	77	80	81	82	83	94	95	97	99	103	109	114	116	131	152	156	163	171
B*3501	G	A	T	N	F	S	N	L	R	G	I	I	R	Y	L	L	D	S	S	V	L	L	Y
B*3502	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	N	Y	-	-	-	-	-
B*3503	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	-	-	-	-	-
B*5301	-	-	-	-	-	N	I	A	L	R	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 6.2 Amino acid sequence of common HLA-B*35 and B*53 alleles. Residues 77-83 determine the Bw4/Bw6 motif in the HLA α_1 helix (highlighted in black). Residues 114 and 116 lie in the α_2 domain.

6.1.3 HLA-B*35 restricted CD8$^+$ T cell responses in HIV

Previous studies have documented the distribution of immunodominant responses restricted by HLA-B*35. In a B-clade infected cohort in Japan, B*3501-restricted responses to two overlapping Nef epitopes VY8 (VPLRPMTY, Nef 78-85) and RY11 (RPQVPLRPMTY, Nef 75-85) were dominant in both acute and chronic infection

(Ueno *et al.* 2007). ELISpot responses and escape mutations also predominated in Nef in a longitudinal analysis of HLA-B*35 positive adolescent subjects with chronic B clade infection (Bansal *et al.* 2007). A similar distribution of responses is reported from a US cohort of acutely infected adults, in whom the majority of $CD8^+$ T cell responses in HLA-B*35 subjects targeted Env and Nef epitopes (Streeck *et al.* 2007).

6.1.4 The impact of HLA-B*35 on HIV disease outcome

HLA-B*35 has been consistently associated with poor control of viraemia and rapid disease progression in subjects infected with B-clade HIV (Kaslow *et al.* 1996; Carrington *et al.* 1999; O'Brien *et al.* 2001; Jin *et al.* 2002; Flores-Villanueva *et al.* 2003). In particular, alleles of the less common Px group have been associated with worse outcome; these subjects characteristically have an early decline in $CD4^+$ T cell counts (Gao *et al.* 2005) and perform worse than homozygotes, suggesting the Px allele may be actively detrimental rather than simply functioning as a 'null' allele (Jin *et al.* 2002).

The mechanisms for these effects are imprecisely understood, but may include the early selection of escape mutations leading to skewing of T cell responses towards less effective targets (Gao *et al.* 2005), highlighting how a single amino acid change in an HLA allele can alter disease outcome (Gao *et al.* 2001). B*35 alleles are members of the Bw6 group; the epitope repertoire and lack of interaction with KIR3DL1 and KIR3DS1 may also account for accelerated disease progression.

6.1.5 Hypothesis and aims

In the C-clade infected cohorts studied in this thesis, subjects with HLA-B*35 appeared to have viral loads approximating to, or lower than, that of the population median, and showed evidence of increased ELISpot responses to Gag peptides. In contrast, the published observations that HLA-B*35 alleles perform poorly in control of HIV viraemia in association with a $CD8^+$ T cell response weighted towards Env/Nef, have all been derived from the analysis of B-clade infected cohorts.

The studies here are therefore based on the hypothesis that subjects with HLA-B*35 may perform better in C-clade than B-clade infection due to different epitope availability. Specifically, the following aims were pursued:

i. To compare viraemic control in HLA-B*35–positive subjects with B- and C-clade infection in cross-sectional and longitudinal data sets;

ii. To identify and characterise HLA-B*35-restricted epitopes, with a particular focus on immune control effected through Gag epitopes;

iii. To identify sites of escape mutation selected by HLA-B*35, and to study the effects of these polymorphisms in Gag epitopes on disease outcome;

iv. To seek evidence for differences in HIV proteins targeted by HLA-B*35 subjects according to the clade of infection.

6.2 METHODS

6.2.1 Study cohorts

Subjects were recruited from cohorts as previously described (Chapter 2). Japanese subjects from Kumamoto were recruited from two distinct cohorts separated by 14-25 years, as described in Chapter 4. Nineteen subjects with HLA-B*35 alleles were recruited from the Thames Valley Cohort.

6.2.2 Generation of ELISpot data

IFN-γ ELISpot data for the South African cohort were generated in Durban. New epitopes were predicted from this data set using Fisher's Exact Test, corrected for multiple comparisons by a Bonferroni approach.

Identifying ELISpot responses made by subjects in the Thames Valley Cohort was undertaken first by screening overlapping peptides (OLPs) spanning the entire HIV proteome, at the baseline time-point. PBMCs were also tested against a panel of optimal epitopes restricted by HLA-B*35, using clade-specific sequences for B and C clade infected subjects. Optimal epitopes screened for were Gag WF9, HA9, PY9; Nef VF/Y8; Pol EW10, TY9, VY10, NY9, EY10; and Env VL11, DL9 and TW9 (full epitope sequences listed in section 6.3.4, Table 6.4).

6.2.3 Analysis of sequence data

Population sequences from Durban and Gaborone were analysed according to the lineage-corrected analysis as previously described (Chapter 3; Carlson *et al.* 2008). The number of taxa contributed from each cohort were matched (selecting at random

from a larger pool where necessary) in order to make results between the cohorts more closely comparable.

6.2.4 Binding studies

HLA-peptide binding studies were undertaken in the Buus lab, Copenhagen, Denmark, using luminescent oxygen channelling immunoassay (LOCI) technology, as previously described Harndahl *et al.* 2009.

6.2.5 Statistics and terminology

Box plots show the median plus 25th and 75th percentiles, with whiskers showing minimum and maximum values, unless otherwise stated. Throughout this chapter, the term 'HLA-B*35' is used to represent alleles from subtypes HLA-B*3501, B*3502 and B*3503 viewed collectively. As shown in Fig. 6.1, the vast majority of these subjects have HLA-B*3501. Four digit type is specified only when subanalysis has been undertaken based specifically on this high resolution type. Despite the overlapping binding motifs, subjects with HLA-B*5301 have not been included in these analyses.

6.3 RESULTS

6.3.1 HLA-B*3501 is the most frequent B*35 allele in HIV-infected cohorts in Europe, Asia and Africa

HLA-B*35 alleles account for between 3.9% (South Africa) and 19% (Japan, combined cohorts) of the cohorts studied here, with HLA-B*3501 predominating over other types (Fig. 6.1, Table 6.3).

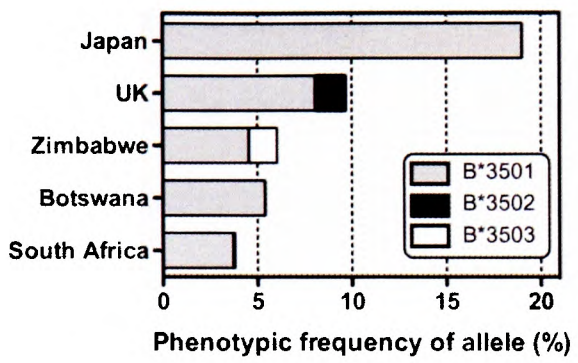


Figure 6.1: Phenotypic frequency of HLA-B*35 alleles in five study cohorts.

B-clade infected cohorts are represented by Japan and UK. C-clade cohorts are represented by Zimbabwe, Botswana, and South Africa.

Clade	Origin of cohort	All subjects		Subjects with HLA-B*35	
		n	median VL	n (%)	median VL
B	Japan (1983)	115	5100	11 (9.6%)	9400
	Japan (1997)	127	51000	26 (20.5%)	86000
	UK*	62	22167	6 (9.7%)	12549
C	Botswana	459	19200	24 (5.2%)	11902
	South Africa	1213	38200	47 (3.9%)	37900
	Zimbabwe*	66	5420	4 (6.1%)	2131
Mixed	Thames Valley (all)	264	15000	19 (7.2%)	15000

Table 6.3: HLA-B*35 phenotypic frequency and median viral load by country, for seven cohorts. Subjects from Japan are divided according to the starting date of recruitment. *Zimbabwe and UK subjects are subsets from within the Thames Valley Cohort.

6.3.2 HLA-B*35 has a different impact on viraemic control in B-clade compared to C-clade HIV

Unlike study subjects from previously published B-clade cohorts (O'Brien *et al.* 2001; Jin *et al.* 2002; Flores-Villanueva *et al.* 2003; Bansal *et al.* 2007), in chronically infected study subjects from South Africa and Botswana, HLA-B*35 is not associated with higher viraemia, and may even drive a trend towards lower viraemia (Fig. 6.2).

Pooling data from B- and C-clade cohorts studied here, there was a highly significant difference in viral loads among 82 subjects with HLA-B*3501 according to the clade of infection (p=0.0004, Mann-Whitney Test, Fig. 6.3). However, these data should be interpreted with caution, as viral loads between cohorts are not directly comparable (Table 6.3), and differences may simply reflect variation in viral load between cohorts. Overall, reflecting previous results, there was a highly significant association between HLA-B*35 and higher viral loads in B-clade subjects, while there was a trend towards lowered viraemia in the presence of HLA-B*35 in C-clade infection (Fig. 6.3).

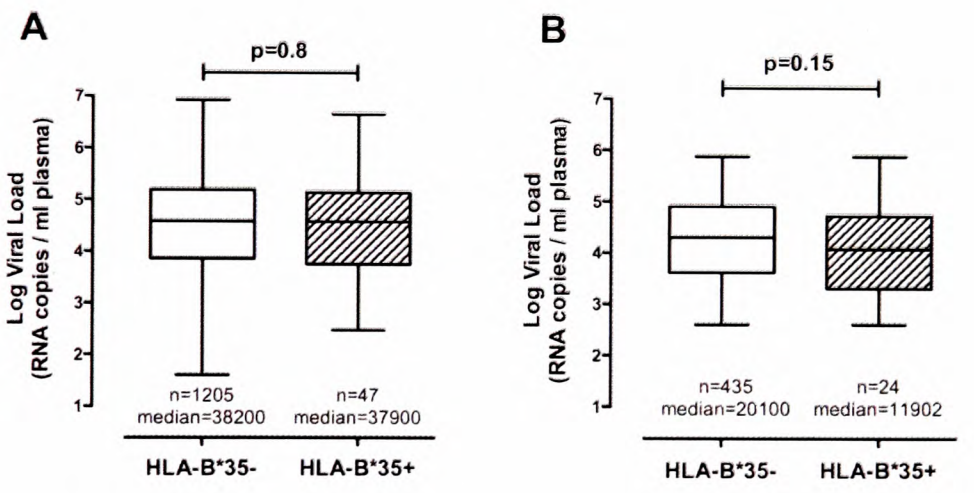


Figure 6.2: Viral load in association with the expression of HLA-B*35 in two C-clade cohorts in subjects with chronic infection. A: Durban, South Africa, B: Gaborone, Botswana.

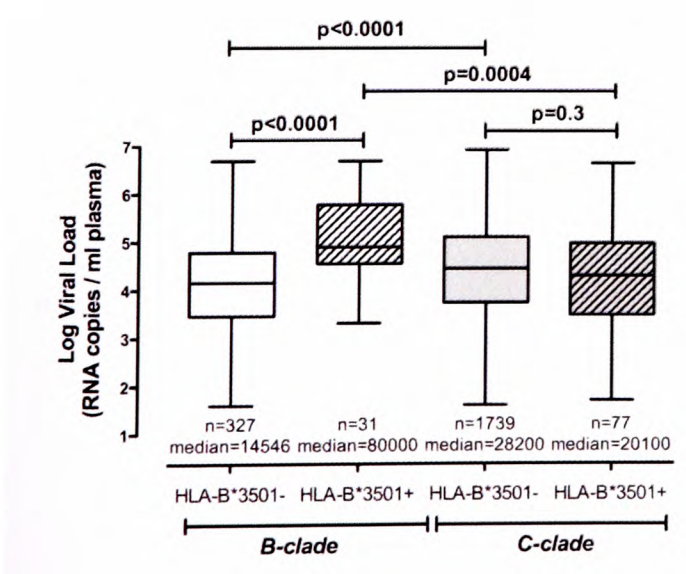


Figure 6.3: Viral load in pooled data from B and C-clade infected study subjects with HLA-B*3501.

B-clade data were taken from the recent Japanese cohort and Thames Valley Cohort; C-clade data were taken from Durban, Gaborone, and Thames Valley Cohort. p values by Mann Whitney Test. Non-HLA-B*3501 alleles were excluded from this analysis in order to reduce bias due to the variable proportion of Px alleles in different cohorts.

In a published analysis of B-clade infected subjects, HLA-B*35 alleles are ranked among the worst performing in relation to viraemic control (O'Brien *et al.* 2001). However, repeating this analysis for two African cohorts demonstrates that HLA-B*35 is associated with viral loads lower than the population median in both Durban and Gaborone (Fig. 6.4).

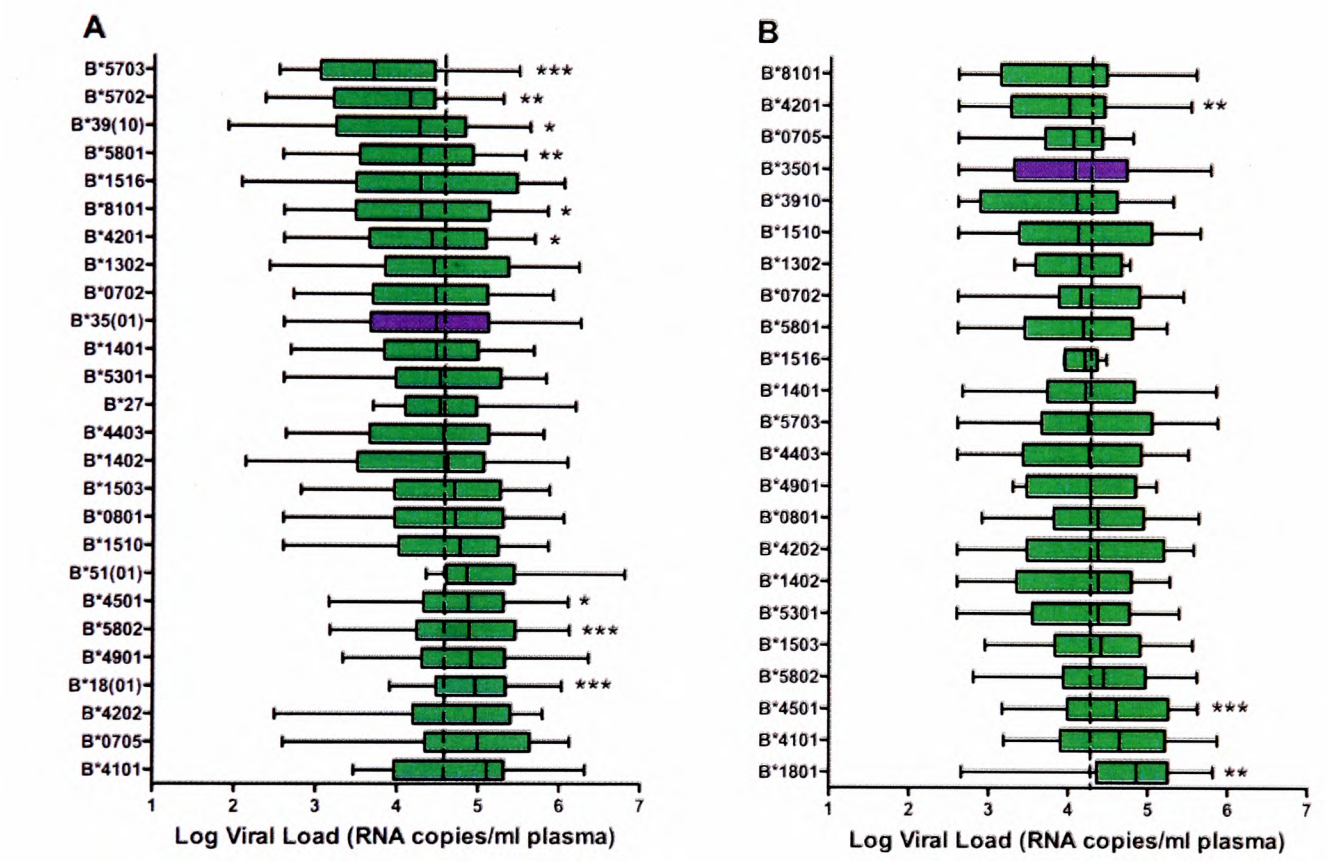


Figure 6.4: HLA-B alleles from two African cohorts ranked according to viral load. A: Durban, B: Gaborone. Boxes show median viral load, 25th and 75th centiles; whiskers 5-95% confidence intervals. HLA-B*35 highlighted in purple in each case. P values calculated by Mann Whitney test, comparing viral loads for specified allele to viral loads for whole cohort *** $p<0.0001$, **$p<0.001$, *$p<0.05$. The dotted line represents the median viral load for the whole cohort (VL 38200 RNA copies/ml in Durban; 19150 RNA copies/ml in Gaborone). HLA types in brackets show designation of allele where four digit type is extrapolated from two digit data in some instances. Alleles represented are those occurring at ≥0.5% phenotypic frequency, and for which a minimum of 5 subjects had viral load data available.

6.3.3 The impact of HLA-B*35 on disease control may change over time

To investigate whether there is evidence of a changing impact of HLA-B*35 over time, the relationship of the allele with viral load was further assessed in two Japanese cohorts, first presenting in 1983, and second from 1997 onwards. In both cohorts, subjects with HLA-B*35 had median viral loads approximately 2-fold greater than the HLA-B*35 negative group. However, this difference was only statistically significant in the later cohort (p=0.002, Mann-Whitney test, Fig. 6.5), and interpretation should be made with reference to the considerations discussed below.

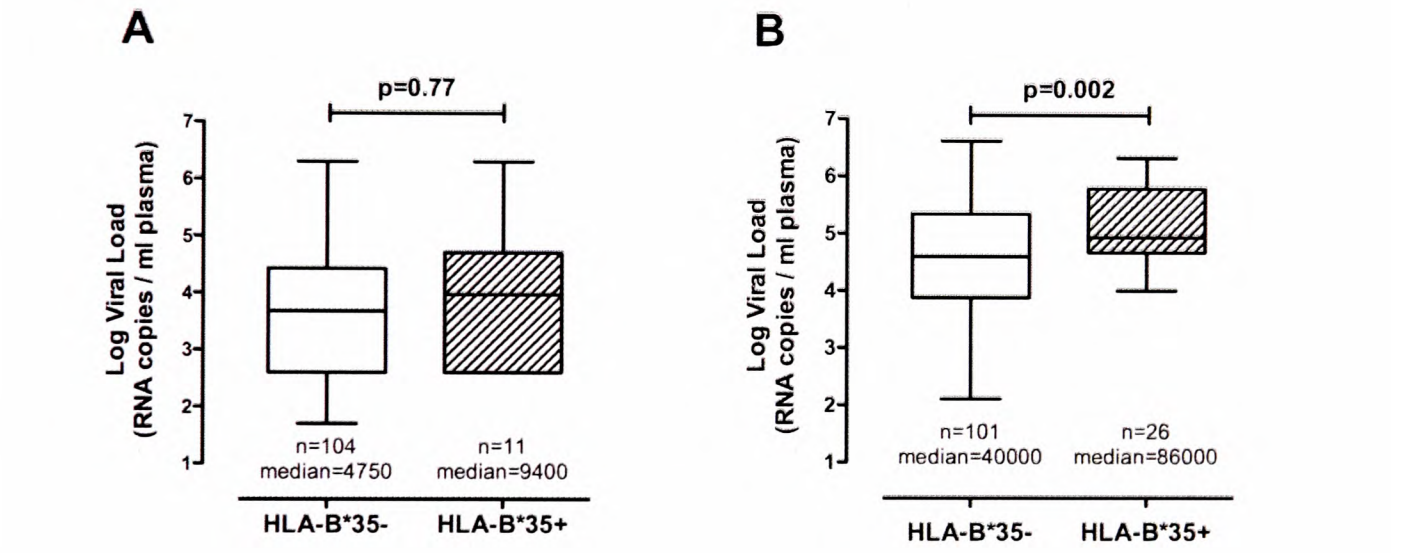


Figure 6.5: Viral load in association with the expression of HLA-B*35 in two Japanese cohorts separated by 14-25 years. A: 115 subjects with haemophilia, recruited in 1983. B: 127 subjects diagnosed after 1997.

The difference in the impact of HLA-B*35 between these two cohorts may be explained in several ways:

i. Fewer subjects (and a smaller proportion of subjects) in the earlier cohort have HLA-B*35; statistical significance may simply be be reduced in this cohort by small study numbers.

ii. The epidemiology of the cohorts is different, as haemophilia subjects may have been infected by a different viral strain and were screened after the risk of HIV acquisition was recognized. These subjects are likely to have been diagnosed earlier in the course of disease than later cohorts. This hypothesis is supported by the significantly lower viral loads in the first cohort compared to the later cohort (for all subjects $p<0.0001$, Mann-Whitney test).

iii. It is possible that any beneficial effect of HLA-B*35 is being lost as the epidemic evolves, due to longitudinal sequence changes in circulating viral strains. This possibility is addressed further in section 6.3.8.

6.3.4 HLA-B*35 alleles present multiple HIV epitopes

In order to investigate the mechanisms by which HLA-B*35 may be associated with this differential effect on disease control in different clades, a list of known and putative HLA-B*35 epitopes was assimilated, based on previously published epitope restrictions and on analysis of HLA associations with ELISpot responses to OLPs carried out in the Durban cohort (Table 6.4).

Protein	Epitope name	B-clade sequence '-P--------Y	C-clade sequence	OLP number
p17 Gag	WF9	WASRELER F	---------	5
p17 Gag	NY9	NSSQVSQN Y	n/a	17
p24 Gag	HA9*	LHPVHAGPI APGQMREPR	---------	29/30
p24 Gag	PY9	PPIPVGEI Y	------D--	35
Nef	VF8	VPLRPMT Y	-------F	76/77
Nef	YF9**	RYPLTFGWC FKLVPV	---------	84/85
Rev	K/QY10**	KTVRLIKFL YQSNPPPS	QA--L--I--	95/96
Tat	EW10	EPVDPRLEP W	-----N----	111
RT	TY9	TVLDVGDA Y	---------	180
RT	VY10	VPLDKDFRK Y	----E-----	182
RT	NQY9	NPDIVIYQ Y	--E------	190
RT	EY10	EPIVGAETF Y	---A------	223
RT	IY11**	GYIEAEVIPAETGQETAY	----------	252
gp120	VL11	VPVWKEATTTL	-------K---	294
gp120	DL9	DPNPQEVV L	------M--	299
gp41	TW9	TAVPWNAS W	-----S---	369

Table 6.4: Optimal epitopes presented by HLA-B*35 alleles (HIV Sequence Compendium 2009.; Brumme *et al.* 2007; Kiepiela *et al.* 2007; Streeck *et al.* 2008)**.** Consensus sequences are given for the B and C clade variants; '-' represents a conserved amino acid position between clades. Epitopes are aligned to the anchor-residues of the most common binding motif. Due to sequence differences between clades, the B-clade p17 Gag epitope NY9 does not exist in C clade. For unrestricted epitopes, the entire OLP sequence is given, with the putative optimal shown in red. *Epitope previously published (www.hiv.lanl.gov) but yet to be formally restricted. **Epitopes are newly defined from analysis of ELISpot responses made by subjects in the Durban cohort, and yet to be formally restricted.

6.3.5 Confirmation of HA9 (HPVHAGPIA, Gag 216-224) as an optimal epitope restricted by HLA-B*35

The HA9 epitope has not been previously confirmed as B*35-restricted, and other alleles (HLA-B*3910, HLA-B*07) may also present the same epitope (Kiepiela *et al.* 2007). Furthermore, with an alanine residue at the C-terminal position, the peptide does not match the recognized HLA-B*35 binding motif. Therefore, further confirmation of this response was sought using three different approaches.

(i) OLP screening by ELISpot

In the Durban cohort, HLA-B*35 was strongly statistically associated with ELISpot responses to OLP 29/30 ($p=7.9x10^{-11}$, Table 6.5). After removal of subjects with HLA-B*35 from the dataset, HLA-B*3910 and B*0702 were also significantly associated with responses to OLP 29/30. These findings confirm the likely presence of an epitope restricted by all three of these alleles within these OLPs. In order to exclude an artefactual association between HLA-B*35 and OLP 29/30 responses due to LD with other alleles, a relationship between linked molecules HLA-Cw*04 and HLA-A*74 (see Appendix II) and these OLP responses in the absence of HLA-B*35 was excluded (Table 6.5).

	Allele present		Allele absent		p value
	Response to OLP 29/30	No response to OLP 29/30	Response to OLP 29/30	No response to OLP 29/30	(Fisher's exact test)
HLA-B*35(01)	17	21	36	613	$7.9x10^{-11}$
HLA-B*0702[§]	11	65	25	548	$1.5x10^{-3}$
HLA-B*39(10)[§]	12	24	24	589	$6.3x10^{-8}$
HLA-Cw*04[§]	2	125	34	488	0.03
HLA-A*74[§]	2	63	34	550	0.56

Table 6.5: Relationship between HLA Class I expression and ELISpot response to OLP 29/30 in Durban cohort (n=687). § Analysis undertaken in HLA-B*35(01) negative subjects only.

(ii) Optimisation and definition of HLA restriction via ELISpot assay

Two Thames Valley Cohort subjects with HLA-B*35 who made responses to the HA9 peptide were screened for responses to this optimal epitope and truncations (Fig. 6.6).

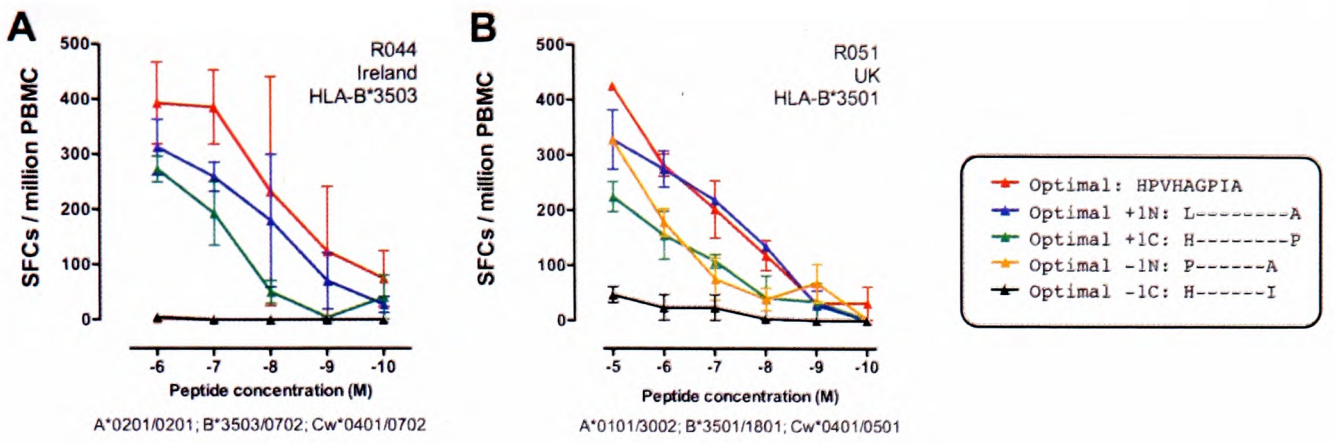


Figure 6.6: ELISpot testing of truncations of optimal HLA-B*35 epitope HA9 (Gag 216-224) in two B-clade infected subjects. The patient ID number, country of origin and HLA-B*35 allele are displayed on the top right of each panel; the full HLA type is displayed underneath.

Although LA10 was similarly recognised in one subject (R051) the addition of one amino acid at the N-terminus did not improve recognition compared to HA9. One subject (R044) has a B*07 allele that could also present this peptide, but the other (R051) has no alleles with an overlapping restriction. These data therefore point to HA9 as the optimal epitope. A $CD8^+$ T cell line primed for this response was grown from both subjects, but failed to reach >10% specificity (Appendix IV); the HLA-restriction of this response has not yet been successfully demonstrated owing to these difficulties in obtaining the appropriate effectors and panel of HLA-matched BCL at the same time.

(iii) HLA-binding data

With a binding (KD) value of 55nM, wild-type HA9 peptide was confirmed as binding to HLA-B*35 (classified as an 'intermediate' binder (Harndahl *et al.* 2009)). Common variants of the peptide show improved binding (Table 6.6). If any of these polymorphisms operate as escape variants, these data suggest a mechanism other than via an effect on HLA-binding.

HA9 Sequence	Binding value (KD, nM)	Binding classification
HPVHAGPIA		
---------	55	Intermediate
-------V-	22	Good
---Q-----	1	Very good

Table 6.6: Binding scores of wild-type HA9 and common peptide variants to HLA-B*35 *in vitro*.

6.3.6 HLA-B*35-restricted epitopes HA9 and PY9 are in highly polymorphic regions of Gag

Given previously published data that substantiate the importance of Gag-specific responses (Borghans *et al.* 2007; Kiepiela *et al.* 2007; Sacha, Chung, Rakasz *et al.* 2007; Goepfert *et al.* 2008; Matthews *et al.* 2008), further assessment of HLA-B*35 epitopes HA9 (Gag 216-224) and PY9 (Gag 254-262) was pursued. Quantification of the variability of p24 Gag at each residue demonstrated that both of these epitopes lie in variable regions of the protein (Fig. 6.7).

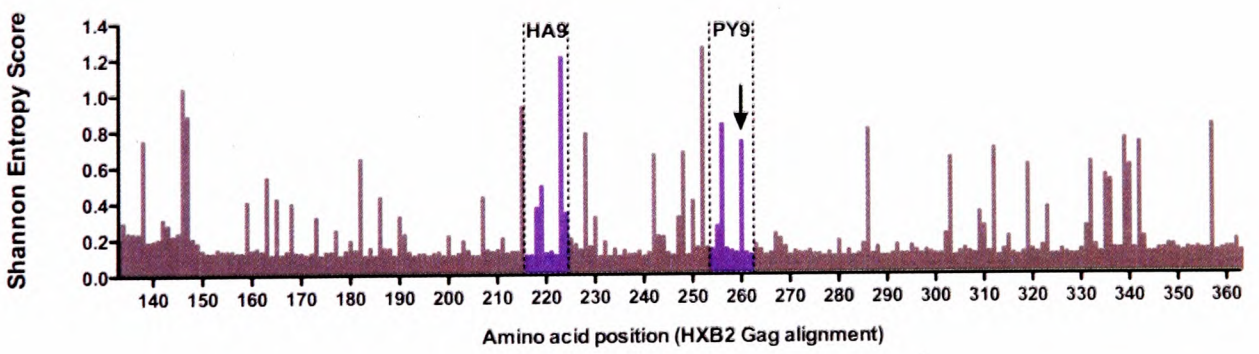


Figure 6.7: Shannon Entropy Score in p24 Gag, showing sites of HLA-B*35 epitopes. Entropy scores were calculated from an alignment of 687 Durban p24 sequences. The positions of the HLA-B*35 restricted epitopes HA9 (216-224) and PY9 (254-262) are marked (purple), and the position of the HLA-B*35-associated polymorphism D260X is indicated (arrow).

The HA9 epitope overlaps with the Cyp-A binding loop (Gag 217-225) (Gatanaga *et al.* 2006), a well-recognized area of variability (Hatziioannou *et al.* 2004; Gatanaga *et al.* 2006). Within this region, the H219Q substitution is associated with compensation

for mutations selected by a variety of HLA alleles, including HLA-B*57 (Gatanaga *et al.* 2006) and HLA-B*27 (Schneidewind *et al.* 2007). The amino acid variability within these individual epitopes was then determined more precisely, according to the presence or absence of HLA-B*35 (Tables 6.7, 6.8).

Following correction for multiple comparisons, none of the amino acid polymorphisms in the HA9 epitope was significantly associated with the expression of HLA-B*35. Together with binding data (Table 6.6), this analysis makes it less likely that that either of the two common HA9 variants, H219X or I225X, function as escape mutations. However, in the PY9 epitope, single mutations at position 256 and position 260, and the double mutant combining escape at these two positions were significantly associated with HLA-B*35 expression in at least one of the cohorts.

Analysis of variation in the HA9 epitope was repeated in Durban and Gaborone cohorts following removal of subjects with HLA-B*3910 and HLA-B*0702 in whom selection in this epitope may also occur. Again, no associations with $p<0.05$ were detected (data not shown).

A

DURBAN

HA9 sequence (Gag 216-224)	Number (%) of subjects with sequence B*35-positive	B*35-negative	P (Fisher's Test)
HPVHAGPIA			
---------	8 (36.4)	275 (42.6)	0.66
-------X-	9 (40.9)	254 (39.4)	1
---X-----	2 (9.1)	16 (2.5)	0.12
---X---X-	2 (9.1)	29 (4.5)	0.27
--X----XX	1 (4.5)	2 (<0.5)	0.1
--X----X-	0 (0)	15 (2.3)	1
-------XX	0 (0)	12 (1.9)	1
--XX---X-	0 (0)	11 (1.7)	1
--------X	0 (0)	8 (1.2)	1
---X---XX	0 (0)	5 (0.8)	1
---X----X	0 (0)	3 (<0.5)	1
--XX---XX	0 (0)	3 (<0.5)	1
--X------	0 (0)	3 (<0.5)	1
-----X---	0 (0)	2 (<0.5)	1
------X--	0 (0)	2 (<0.5)	1
--XX-----	0 (0)	2 (<0.5)	1
--X-----X	0 (0)	1 (<0.5)	1
---XX--XX	0 (0)	1 (<0.5)	1
---X--X--	0 (0)	1 (<0.5)	1
-----X-X-	0 (0)	0 (0)	1
--XX----X	0(0)	0(0)	1
Total	22	645	
H219X total	4 (18.2)	71 (11.0)	0.3
I223X total	12 (54.5)	332 (51.5)	0.8

B

GABORONE

HA9 sequence (Gag 216-224)	Number (%) of subjects with sequence B*35-positive	B*35-negative	P (Fisher's Test)
HPVHAGPIA			
---------	9 (45)	102 (36.2)	0.47
-------X-	4 (20)	119 (42.2)	0.06
---X-----	0 (0)	15 (5.3)	0.61
---X---X-	4 (20)	18 (6.4)	0.05
--X----XX	0 (0)	1 (<0.5)	1
--X----X-	0 (0)	6 (2.1)	1
-------XX	0 (0)	6 (2.1)	1
--XX---X-	2 (10)	4 (1.4)	0.05
--------X	0 (0)	2 (0.7)	1
---X---XX	0 (0)	4 (1.4)	1
---X----X	0 (0)	0 (0)	1
--XX---XX	1 (5)	1 (<0.5)	1
--X------	0 (0)	1 (<0.5)	1
-----X---	0 (0)	1 (<0.5)	1
------X--	0 (0)	0 (0)	1
--XX-----	0 (0)	0 (0)	1
--X-----X	0 (0)	0 (0)	1
---XX--XX	0 (0)	0 (0)	1
---X--X--	0 (0)	0 (0)	1
-----X-X-	0 (0)	1 (<0.5)	1
--XX----X	0 (0)	1 (<0.5)	1
Total	20	282	
H219X total	7 (35.0)	43 (15.2)	0.03
I223X total	11 (55.0)	160 (56.7)	1

Table 6.7: Variants of Gag HA9 epitope in population sequences from (A) Durban, and (B) Gaborone.

A

DURBAN

PY9 sequence (Gag 254-262)	Number (%) of subjects with sequence B*35-positive	B*35-negative	P (Fisher's Test)
PPIPVGDIY			
---------	2 (9.1)	227 (35.2)	0.01
--X------	0 (0)	200 (31.0)	5.30×10^{-4}
------X--	7 (31.8)	94 (14.6)	0.04
--X---X--	12 (54.6)	95 (14.7)	2.60×10^{-3}
-X-------	1 (4.5)	3 (0.5)	0.13
-X----X--	0 (0)	11 (1.7)	1
X--------	0 (0)	3 (0.5)	1
-XX------	0 (0)	3 (0.5)	1
-XX---X--	0 (0)	2 (<0.5)	1
--XX-----	0 (0)	1 (<0.5)	1
--X----X-	0 (0)	1 (<0.5)	1
---X--X--	0 (0)	1 (<0.5)	1
-----XX--	0 (0)	1 (<0.5)	1
----X----	0 (0)	1 (<0.5)	1
---X-----	0 (0)	1 (<0.5)	1
-X--X-X--	0 (0)	1 (<0.5)	1
Total	22	645	
I256X total	12 (54.5)	302 (46.8)	0.52
D260X total	19 (86.4)	205 (31.8)	3.09×10^{-7}

B

GABORONE

PY9 sequence (Gag 254-262)	Number (%) of subjects with sequence B*35-positive	B*35-negative	P (Fisher's Test)
PPIPVGDIY			
---------	4 (20.0)	148 (52.9)	0.005
--X------	1 (5.0)	62 (22.1)	0.09
------X--	9 (45.0)	25 (8.9)	7.49×10^{-5}
--X---X--	6 (30.0)	33 (11.8)	0.03
-X-------	0 (0)	8 (2.9)	1
-X----X--	0 (0)	0 (0)	1
X--------	0 (0)	1 (<0.5)	1
-XX------	0 (0)	2 (0.7)	1
-XX---X--	0 (0)	0 (0)	1
--XX-----	0	0 (0)	1
--X----X-	0	0 (0)	1
---X--X--	0	0 (0)	1
-----XX--	0	0 (0)	1
----X----	0	0 (0)	1
---X-----	0	0 (0)	1
-X--X-X--	0	1 (<0.5)	1
Total	20	280	
I256X total	7 (35.0)	97 (34.6)	1
D260X total	15 (75.0)	59 (21.1)	1.24×10^{-4}

Table 6.8: Variants of Gag PY9 epitope in population sequences from (A) Durban and (B) Gaborone.

Greater variation in the PY9 epitope is seen in C-clade compared to B-clade infection

In order to account for possible clade-specific differences in the targeting of Gag epitopes, the extent of variability in the HA9 and PY9 epitopes and flanking residues was assessed in subjects from different clades (Fig. 6.8). In the case of HA9, there was significantly greater variability in B-clade at positions 211 and 219, and in C-clade at position 215. In PY9, positions 256 and 260 were significantly more variable in C-clade infected subjects. This latter observation again supports the hypothesis that the PPIPVGDIY epitope is more frequently targeted in the context of C-clade infection, accounting for the increased selection of escape variants.

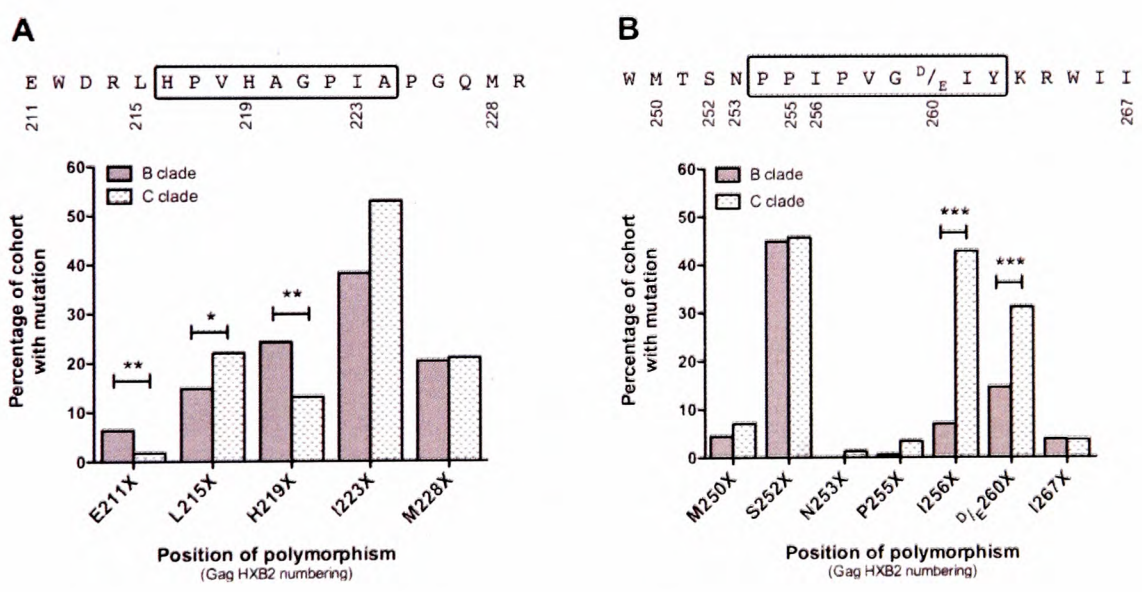


Figure 6.8: Variability of the region spanning the Gag HLA-B*35 restricted epitopes, HA9 and PY9 epitope (±5 amino acids), in B and C clade study cohorts. Sites with variability at >1% of residues were included in the analysis. Variation is considered compared to clade consensus, which differs at position 260, but is otherwise conserved. A: Variability in and flanking HA9 epitope. B-clade sequences from Japan and UK, C-clade sequences from South Africa, Botswana and Zimbabwe. B: Variability in and flanking PY9 epitope. B-clade sequences from UK only, C-clade sequences as before. *$p<0.05$, **$p<0.01$, ***$p<0.0001$.

6.3.7 Gag-D260X is among the most frequent polymorphisms selected by HLA-B*35

Having identified epitopes that are presented by HLA-B*35 alleles, and associated sequence variability, sites of escape mutation from HLA-B*35 restricted responses were additionally sought using the lineage-corrected algorithm described in Chapter 3 (Carlson *et al.* 2008) (Table 6.9). In keeping with the results of Fisher's Exact Test (Table 6.8), the D260X mutation in the PY9 epitope was strongly and consistently associated with the presence of HLA-B*35. There is no evidence of reversion at this position, suggesting that the mutation does not impose a significant fitness cost and is likely to accumulate in populations over time.

Protein/Sequence		Gaborone *(n=322)*		Durban subset *(n=322)*		Durban full cohort *(n=666)*		Reversion
		p	q	p	q	p	q	(any cohort)
p24	P I A P G Q M R E P **R** G S D I A G T T S T	NS	NS	NS	NS	5.64E-04	0.10	R
	M T S N P P I P V G **D** I Y K R W I I L G L	1.50E-06	2.15E-03	3.23E-05	0.01	1.09E-07	0.00	
	M T A C Q G V G G P **G** H K A R V L	6.71E-04	0.19	NS	NS	NS	NS	R
RT	G D A Y F S V P L D **E** D F R K Y T A F T I	1.89E-05	0.04	1.88E-04	0.08	5.39E-07	0.00	
	L W Y Q L E K E P I **A** G A E T F Y V D G A	NS	NS	5.63E-04	0.13	NS	NS	
	L E P F R A K N P E **I** V I Y Q Y M D D L Y	NS	NS	NS	NS	5.61E-04	0.15	
Nef	R P Q V P L R P M T **Y** K A A F D L S F F L	1.59E-08	5.25E-05	1.06E-03	0.19	1.13E-07	0.00	R
	D W Q N Y T P G P G **V** R Y P L T F G W C F	3.77E-06	2.93E-03	NS	NS	6.46E-06	0.00	R
	K F D S S L A R R H **M** A R E L H P E Y Y K	5.70E-05	0.03	8.97E-05	0.04	NS	NS	R

Table 6.9: Sequence polymorphisms in HIV-1 Gag, Pol and Nef identified in association with HLA-B*35 in subjects from Gaborone and Durban. Analysis of Durban subjects based on two cohorts: 'subset' is a cohort of antenatal subjects with sequence numbers matched to Gaborone; 'full' is whole cohort of subjects as presented in Chapter 3. Site of polymorphism marked in black bar. Known epitopes highlighted in grey. p and q values computed by lineage-corrected algorithm. NS = not significant (q>0.2). R = reversion predicted following transmission to HLA-mismatched recipient.

This analysis demonstrates that HLA-B*35 imposes strong immunological selection pressure, leading to escape mutations not only in p24 Gag, but also in RT and Nef. Of nine HLA-associated polymorphisms identified, seven are within or flanking (±8 amino acids) known HLA-B*35 epitopes. This evidence of potent immune selection

pressure supports the hypothesis that HLA-B*35 alleles can potentially make an important contribution to viraemic control.

6.3.8 Frequency of Gag-D260X polymorphism varies according to the population prevalence of HLA-B*35

Having demonstrated a strong selection pressure for the D260X mutation, the prevalence of this variant in different populations was assessed. As described in Chapter 5, the frequency of the Gag variant D260X was shown to be proportional to the phenotypic frequency of HLA-B*35 in the population (Fig. 6.9).

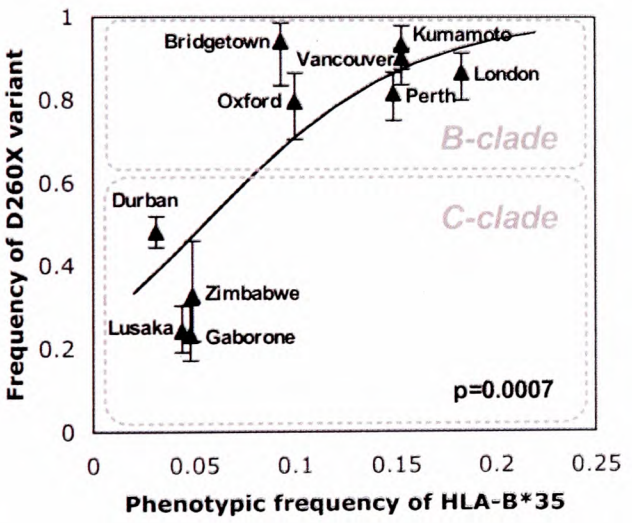


Figure 6.9: Relationship between phenotypic frequency of HLA-B*35 and frequency of Gag D260X in 10 cohorts.

Study subjects as described in Chapter 5, including additional cohort from Zimbabwe. In B-clade infected cohorts, the E260 form accounts for a minimum of 80% of all circulating strains, while in C-clade, D260 is still present in the majority. P=0.0007, logarithmic regression.

The striking difference in the distribution of variants of the PY9 epitope between clades may underpin differences in control of viraemia by subjects with HLA-B*35 in the context of B- and C-clade infection. The high frequency of the E260 variant in B-clade infected subjects may could account for diminished recognition of this epitope compared to that seen in C-clade infection. Likewise, the wildtype (D260) epitope is largely unavailable in currently circulating viruses in Japan, but may have been more widely prevalent earlier in the course of the epidemic, potentially explaining the

longitudinal difference in disease outcomes described above (section 6.3.3). However, founder effect may also have an effect in mediating this clade-specific difference.

6.3.9 Consequences of Gag-D260X polymorphism in subjects with HLA-B*35

Having identified strong selection pressure for Gag-D260X substitution, both using Fisher's test and using a lineage-corrected algorithm, the dynamics and consequences of this HLA-selected mutation were further explored.

The C-clade escape variant Gag D260E is the B-clade consensus sequence

Interestingly, the presence of glutamic acid (E) at Gag position 260 is the HLA-B*35 escape variant in C-clade infection, while it represents the consensus residue in B-clade taxa. This may be as a result of founder effect, with or without an original HLA-B*35 footprint in the ancestral B-clade sequence, or it may reflect accumulation of the D260E variant that has occurred in B-clade subjects over time. This recapitulates the thesis proposed in Chapter 4, that clade specific amino acid differences frequently occur at sites of $CD8^+$ T-cell escape.

The D260E substitution reduces $CD8^+$ T cell recognition of the PY9 epitope in C-clade infection

The effect of amino acid substitutions at Gag-260 was examined in a South African subject from the Thames Valley Cohort who made an ELISpot response to the wildtype PY9 epitope. The response to the optimal peptide was markedly diminished in the presence of the D260E substitution (Fig. 6.10).

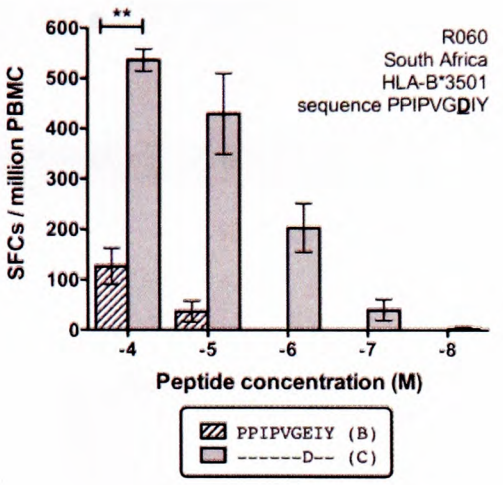


Figure 6.10: ELISpot responses to two variants of the PY9 epitope made by an HIV-infected subject. Error bars show mean ± SEM of three experiments. Patient's ID number, country of origin, HLA-B*35 allele and autologous population sequence are shown at the top right corner. **p<0.001, paired t-test.

Additional evidence for the impact of the D260E mutation on the recognition of the PY9 epitope was assessed using ELISpot data from 32 subjects with HLA-B*35. Subjects making an IFN-γ ELISpot response to OLP 35 (PY9 epitope) were enriched for the presence of D260; these data suggest that the presence of the D260E substitution makes response to the PY9 epitope less likely, although this association did not reach statistical significance (p=0.2, Fisher's exact test, Fig. 6.11).

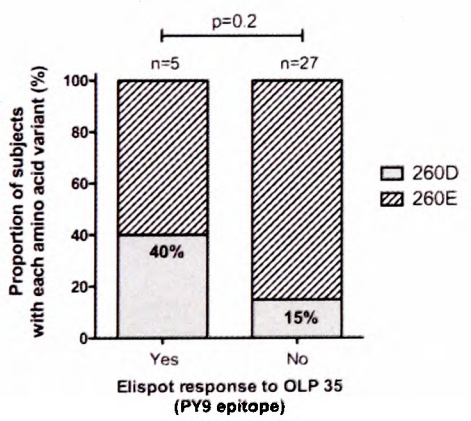


Figure 6.11: Relationship between ELISpot response to OLP 35 and amino acid residue at Gag position 260 in subjects with HLA-B*35.
Pooled data for HLA-B*35 positive subjects in Durban (n=16, C-clade) and Thames Valley (n=16, mixed clade) cohorts. p value by Fisher's exact test.

Mechanism of escape in association with Gag D260E mutation

In order to investigate the mechanism of $CD8^+$ T cell escape associated with the Gag D260E substitution, binding studies were carried out (Harndahl *et al.* 2009) (Fig. 6.12). Both variants of the PY9 epitope (differing only at position 260) were classified as 'intermediate' binders). Additional binding data are shown in Appendix V.

A

PY9 Sequence	Binding value (KD, nM)	Binding classification
PPIPVGEIY		
---------	76	Intermediate
------D--	407	Intermediate

B

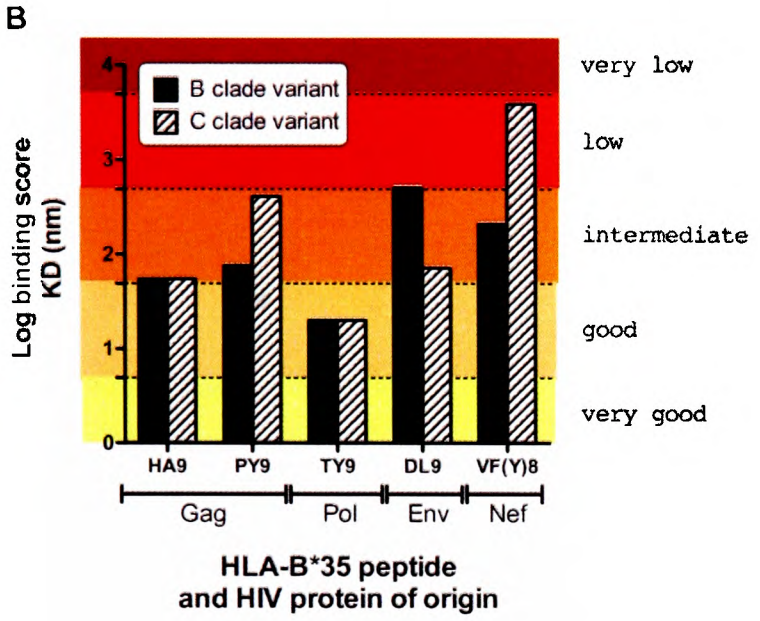


Figure 6.12: Binding of HLA-B*35 B and C clade epitopes to HLA-B*35 Class I molecule.

A: Binding scores of B-clade wildtype peptide (E260) and C-clade wildtype peptide (D260) to HLA-B*35.

B: Binding of B and C-clade peptide variants for epitopes HA9, PY9, TY9, DL9 and VF(Y)8 to HLA-B*35. Sequences as listed in table 6.4.

Together, these ELISpot and binding data suggest that the D260E escape mutation is most likely to allow escape from the $CD8^+$ T cell response by reducing TCR recognition rather than by altering antigen processing or HLA-binding.

The Gag-D260E polymorphism is not associated with altered viral load

Having highlighted the potential contribution of the Gag PY9 epitope in control of viraemia by subjects with HLA-B*35, the effect of the escape mutation, D260E, on viral load was examined (Fig. 6.13). Selection of an escape mutation at D260 might be expected to lead to increased viraemia resulting from a loss of HLA-B*35-mediated immune control. However, there was no significant effect of this mutation on viral load, either in individual cohorts, or in amassed cross-cohort data. Multiple different polymorphisms in this region may affect HLA Class I presentation, and the

presence of other epitopes in p24 Gag (including HA9) may allow preservation of immune control despite PY9 escape.

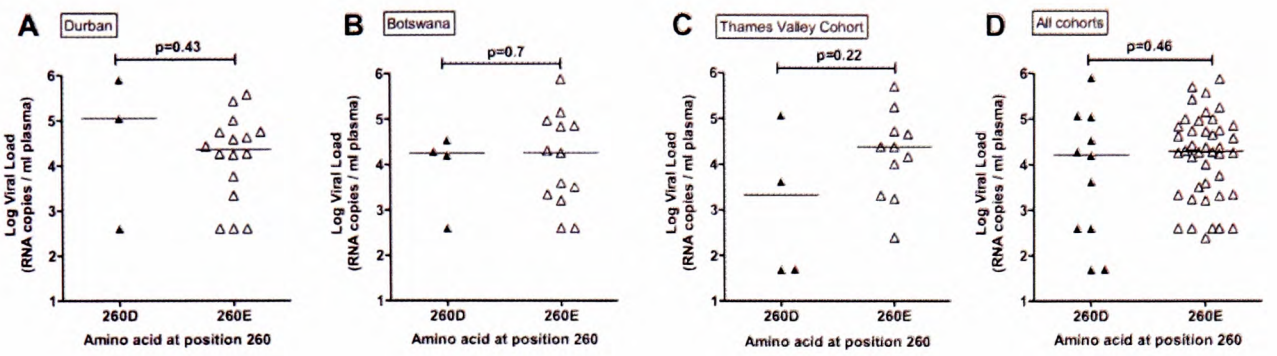


Figure 6.13: The effect of the D260E Gag polymorphism on viral load in Durban (A), Gaborone (B) and Thames Valley Cohorts (C), and all data pooled (D).

6.3.10 Co-variation of Gag-D260X with other Gag mutations

In subjects with HLA-B*35, the PY9 variant I256X is almost exclusively seen in the presence of D260X (table 6.8). This observation suggests that D260X is likely to arise first, and may be required to compensate a fitness cost associated with subsequent selection of I256X. In HLA-B*35-negative subjects, the I256X mutation commonly arises (in >30% of subjects in Durban) in the absence of D260E, but may be compensated by other changes outside the epitope. The analysis in Chapter 3 identified I256X mutations in association with HLA-B*57, suggesting that it may be part of the suite of Gag mutations – combining escape and compensatory changes – selected by this allele.

To identify the relationship between combinations of mutations in p24, an analysis of co-variation was undertaken (Table 6.10). This analysis confirms significant co-variation across p24 in association with these PY9 polymorphisms, with a clear link to mutations selected by HLA-B*57 as well as to other HLA-B*35 epitopes.

Index polymorphism	Co-varying residue (HXB2)	Sequence	Restriction of known epitope(s)	p	q
256	147	NLQGQMVHQAISPRTLNAWVK	B*57	1.2×10^{-3}	0.12
	163	NAWVKVVIEKAFSPEVIPMFS	B*57	7.4×10^{-4}	0.09
	182	FTALSEGATPQDLNTMLNTVG	B*07, B*39, B*81	2.2×10^{-3}	0.18
	186	SEGATPQDLNTMLNTVGGHQA	B*07, B*39, B*81	6.7×10^{-5}	0.02
	190	TPQDLNTMLNTVGGHQAAMQM	B*07, B*39, B*81	2.9×10^{-4}	0.05
	207	AMQMLKETINEEAAEWDRLHP	B*07, B*35, B*39	4.6×10^{-5}	0.01
	215	INEEAAEWDRLHPVHAGPIAP	B*07, B*35, B*39	1.8×10^{-4}	0.04
	223	DRLHPVHAGPIAPGQMREPRG	B*07, B*35, B*39	8.2×10^{-5}	0.02
	225	LHPVHAGPIAPGQMREPRGSD	B*07, B*35, B*39	1.2×10^{-3}	0.12
	228	VHAGPIAPGQMREPRGSDIAG	B*07, B*35, B*39	6.9×10^{-5}	0.02
	246	IAGTTSTLQEQIAWMTNNPPI	B*35, B*57, B*58	2.5×10^{-3}	0.19
	247	AGTTSTLQEQIAWMTNNPPIP	B*35, B*57, B*58	4.0×10^{-8}	3.5×10^{-5}
	248	GTTSTLQEQIAWMTNNPPIPV	B*35, B*57, B*58	5.3×10^{-10}	7.8×10^{-7}
	250	TSTLQEQIAWMTNNPPIPVGD	B*35, B*57, B*58	3.5×10^{-4}	0.06
	252	TLQEQIAWMTNNPPIPVGDIY	B*35, B*57, B*58	3.6×10^{-4}	0.06
260	138	PIVQNLQGQMVHQAIS		3.5×10^{-4}	0.06
	146	QNLQGQMVHQAISPRTLNAWV	B*57	1.2×10^{-4}	0.03
	147	NLQGQMVHQAISPRTLNAWVK	B*57	8.5×10^{-5}	0.02
	168	VVIEKAFSPEVIPMFTALSEG	B*57	2.8×10^{-6}	0.001
	207	AMQMLKETINEEAAEWDRVHP		2.5×10^{-5}	0.009
	215	INEEAAEWDRLHPVHAGPIAP	B*07, B*35, B*39	4.3×10^{-10}	6.5×10^{-7}
	216	NEEAAEWDRLHPVHAGPIAPG	B*07, B*35, B*39	1.7×10^{-4}	0.04
	224	RVHPVHAGPIAPGQMREPRGS	B*07, B*35, B*39	1.7×10^{-6}	0.009
	228	VHAGPIAPGQMREPRGSDIAG	B*07, B*35, B*39	2.0×10^{-4}	0.04
	250	TSTLQEQIAWMTNNPPIPVGD	B*35, B*57	2.8×10^{-10}	6.0×10^{-7}
	268	VGDIYKRWIILGLNKIVRMYS		1.9×10^{-3}	0.16
	277	ILGLNKIVRMYSPVSILDIKQ		0.01	0.1

Table 6.10: Co-variation of polymorphisms in p24 Gag with mutations at positions 256 and 260 in the PY9 epitope (PPIPVGDIY, 254-262) in the Durban cohort. The position of known epitopes in the region of the co-varying polymorphism are highlighted in grey. Sequences are C-clade consensus; p and q values according to lineage-corrected analysis (Carlson *et al.* 2008).

6.3.11 Subjects with HLA-B*35 make ELISpot responses weighted towards Gag/Pol in C-clade and towards Env/Nef in B-clade infection

The evidence presented thus far demonstrates improved outcomes in association with HLA-B*35 in C- compared to B-clade infection and highlights a heightened immune selection pressure on the Gag PY9 epitope in C-clade infection.

In order to investigate the extent to which PY9 and other HLA-B*35-restricted epitopes are targeted, Thames Valley Cohort subjects were screened for IFN-γ ELISpot responses to HIV OLPs spanning the entire proteome. Sites of six HLA-B*35 epitopes in Gag, Nef, Pol and Env were targeted more frequently by subjects with HLA-B*35 (n=17) than by HLA-B*35 negative subjects (n=178) (Fig. 6.14).

Unlike other HLA-B*35 epitopes, targeting the region of the PY9 epitope (OLP 35) was not significantly associated with the expression of HLA-B*35. This may relate to the mixed clades represented in this cohort, and also to the common selection of escape mutations that abrogate responses even among C-clade infected study subjects.

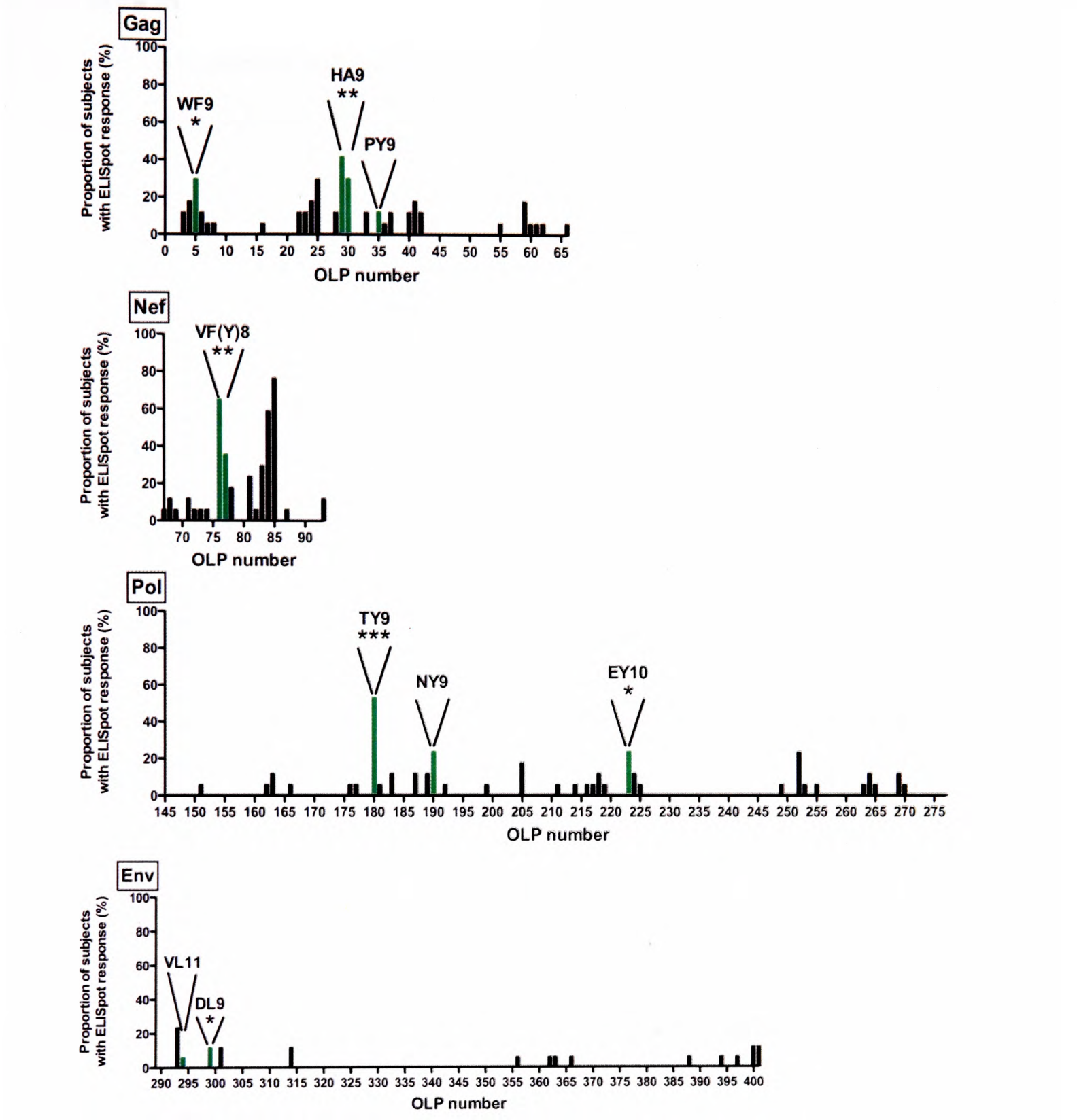


Figure 6.14: Proportion of HLA-B*35-positive subjects making IFN-γ ELISpot responses to overlapping HIV peptides in Gag, Nef, Pol and Env. Sites of HLA-B*35 restricted epitopes are shown in green. Statistical significance was assessed by comparing response made by HLA-B*35 positive subjects (green bars) to responses made by HLA-B*35 negative subjects (data not shown), using Fisher's Exact Test. * $p<0.05$, ** $p<0.001$, *** $p<0.0001$.

To focus more specifically on HLA-B*35-restricted responses, ELISpot screening was also undertaken using a panel of 12 optimal epitopes. As seen in OLP screening, responses were again distributed across Gag, Pol, Env and Nef (Fig. 6.15). When divided according to clade, there was a trend towards increased Gag/Pol responses, and away from Env/Nef responses in C-clade compared to B-clade infected subjects, although this did not reach statistical significance (Fig. 6.16).

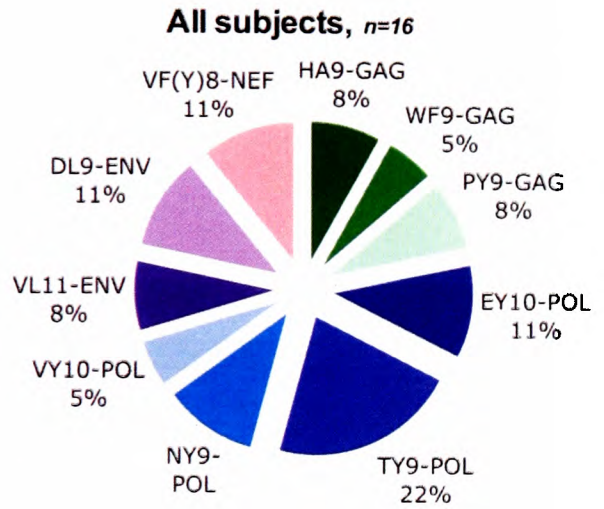


Figure 6.15: Distribution of IFN-γ ELISpot responses to HLA-B*35 optimal epitopes made by 16 subjects with HLA-B*35. 13 subjects had B*3501, 2 had B*3502 and 1 had B3503. Each subject made 1-4 responses (median 2) to B*35 optimals; a total of 37 responses are represented. No subject made responses to Pol EW10 or Env TW9.

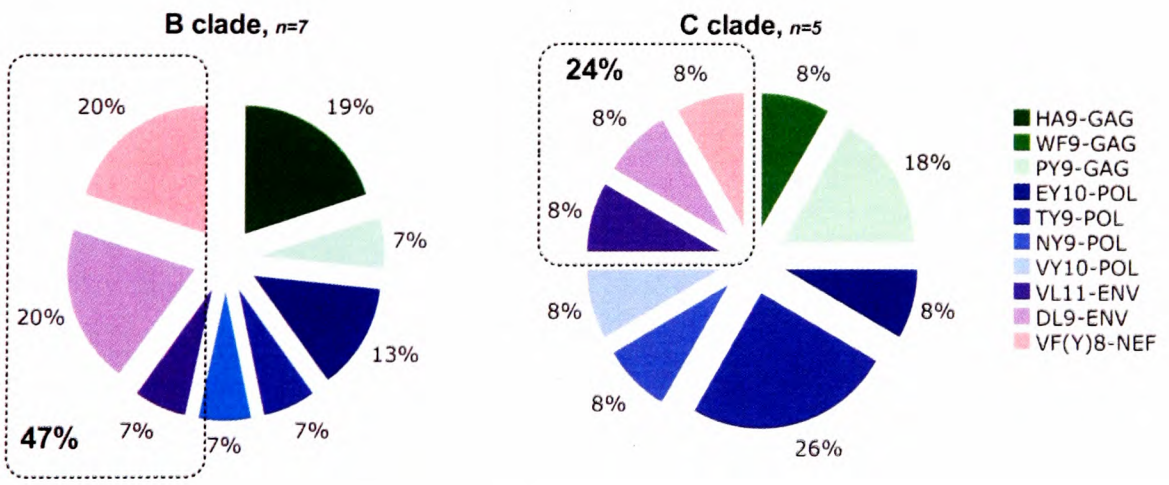


Figure 6.16: Distribution of IFN-γ ELISpot responses to HLA-B*35 optimal epitopes made by subjects with HLA-B*35 according to clade of infection. Subjects with probable nonB/C-clade infection are excluded. Responses to Nef/Env are enclosed within dashed boxes, showing reduction in responses to these proteins in C-clade, paralleled by a greater proportion of Gag/Pol responses in C-clade. However, this difference is not statistically significant (p=0.4, Fisher's Exact test).

As well as quantifying the proportion of all HLA-B*35 optimal responses that are accounted for by each epitope, the proportion of HLA-B*35-positive subjects making each individual response was considered (Fig. 6.17).

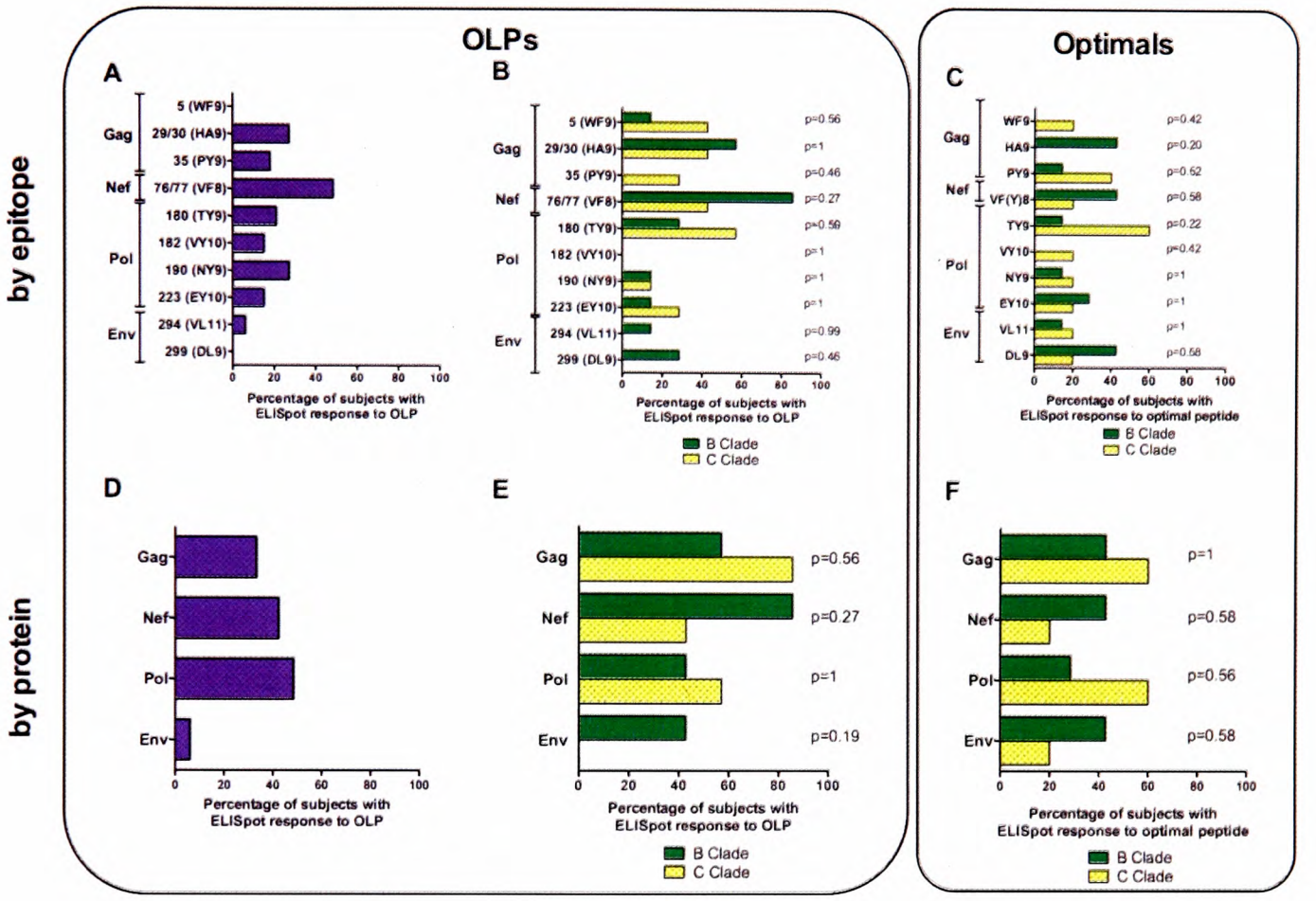


Figure 6.17: Percentage of HLA-B*35-positive subjects from Durban (purple) and Thames Valley (green/yellow) cohorts making responses to HLA-B*35 OLPs and optimal epitopes. Durban subjects were screened by megamatrix only (individual optimal responses were not tested). p values calculated by Fisher's Exact Test.

As expected, the PY9 epitope was more frequently targeted in C-clade infected subjects (Fig. 6.17B,C), while a greater proportion of B-clade subjects made responses to epitopes in Nef and Env. HA9 was targeted by subjects with both B and C clade infection, although B-clade responses were more frequent overall. However, this is confounded by the presence of overlapping epitopes restricted by HLA-B*0702 and B*3910 as previously discussed.

6.3.12 HLA-B*35 restricted responses to OLPs containing Gag epitopes HA9 and PY9 are associated with reduction in viral load

The relationship between increased breadth of Gag responses and improved viraemic control has previously been demonstrated in the Durban cohort (Kiepiela *et al.* 2007). In order to establish the extent to which $CD8^+$ T cell control of viraemia in subjects with HLA-B*35 might hinge on responses to Gag epitopes, the relationship between breadth of ELISpot responses and viral load was assessed in this subset of patients. As for the whole cohort, a significant correlation was observed between number of Gag responses and lowered viral load (p=0.01, Fig. 6.18A), that was not significant for Pol, Env or Nef (Fig. 6.18B-D).

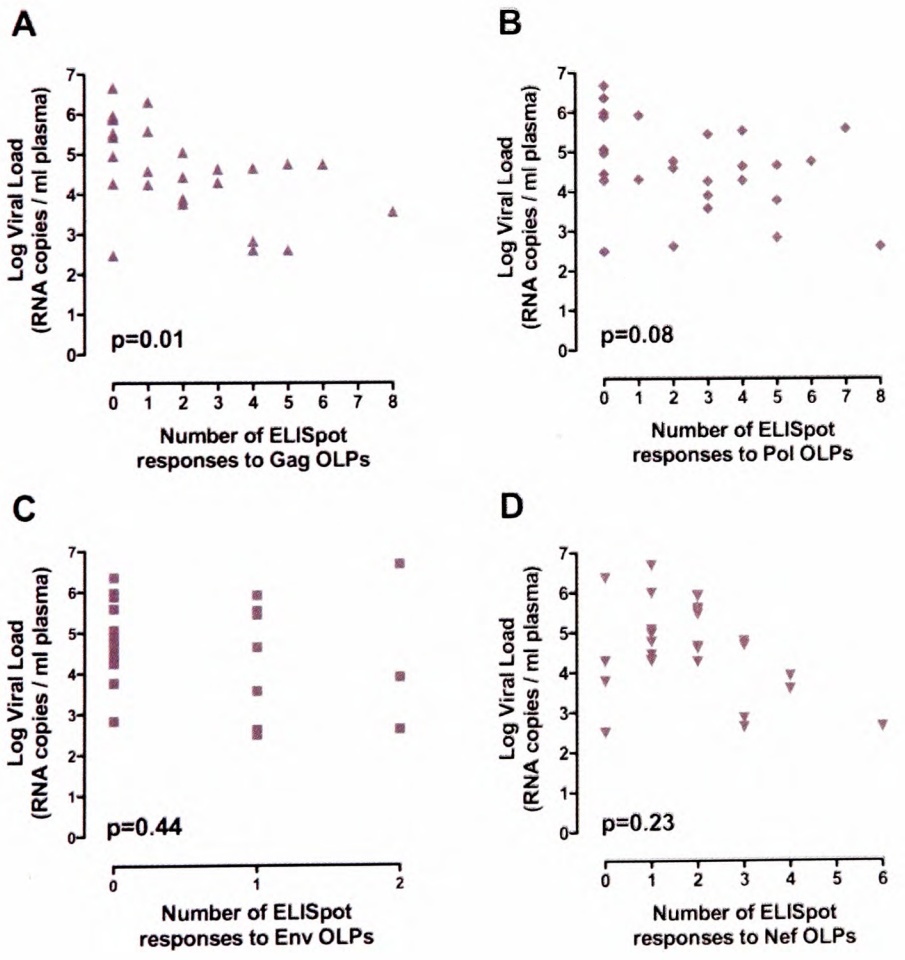


Figure 6.18: Relationship between number of OLPs targeted and viral load among subjects with HLA-B*35 in Durban cohort. A: Gag, B: Pol, C: Env, D: Nef.

p values by Spearman rank correlation.

In order to ascertain to what extent targeting of the epitopes HA9 and PY9 might contribute to this viraemic control, the relationship between ELISpot responses to these specific peptides and viral load was examined. In Durban, subjects targeting any of the OLPs containing these epitopes (OLPs 29, 30 and 35) had significantly lowered viraemia compared to subjects not making these responses (p=0.03, Fig. 6.19A). In the Thames Valley Cohort, a similar trend was observed, although this did not reach statistical significance (p=0.15, Fig. 6.19B). Analysing the C-clade data together, the difference became more strongly significant (p=0.005, Fig. 6.19C). These data again highlight the importance of $CD8^+$ T cell responses to Gag epitopes HA9 and PY9 in control of viraemia.

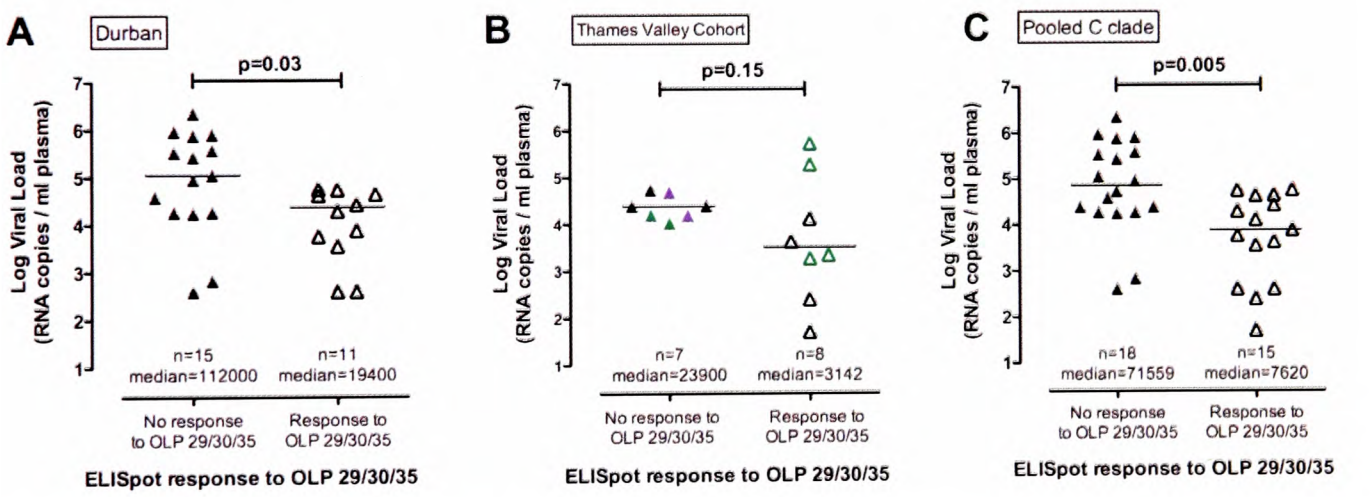


Figure 6.19: HLA-B*35-positive subjects with ELISpot responses to OLPs 29, 30 or 35, containing the Gag epitopes HA9 and PY9, have lowered viral loads compared to subjects not making these responses. p values by Mann-Whitney test. A: Durban cohort. B: Thames Valley Cohort, (A-clade – green; B-clade – purple; C-clade – black). C: Pooled C-clade data from Durban and Thames Valley Cohorts.

6.4 DISCUSSION

In summary, the data presented in this chapter demonstrate that subjects with HLA-B*35 control viraemia better in the context of C-clade than B-clade HIV infection. Over time, changes in B*35-mediated control are likely, due to the accumulation of escape polymorphisms in populations in proportion to the phenotypic frequency of HLA-B*35 alleles. HLA-B*35 restricts multiple $CD8^+$ T cell responses across HIV Gag, Pol, Env and Nef proteins, with control of disease likely to hinge primarily on responses to Gag epitopes.

6.4.1 Evidence that HLA-B*35 may contribute to viraemic control and HIV sequence evolution

In contrast to previously published work based on B-clade analysis (O'Brien *et al.* 2001; Jin *et al.* 2002; Bansal *et al.* 2007), these data highlight that HLA-B*35 is not necessarily detrimental to disease control in subjects with C-clade infection. Cross-sectional analysis of ten different study cohorts, and longitudinal analysis of B-clade infected subjects in Japan, support the hypothesis that the contribution of this allele to viraemic control may be changing over time, with a possible loss of critical $CD8^+$ T cell responses due to HLA adaptation at a population level.

6.4.2 Identification of HLA-B*35 epitopes and escape mutations

Location of HLA-B*35 epitopes, and evidence for Gag HA9 epitope

HLA-B*35 epitopes have been identified across HIV Gag, Pol, Env and Nef, suggesting a great diversity of potential $CD8^+$ T cell responses (www.lanl.gov). Although the Gag epitope HA9 (HPVHAGPIA) deviates from the recognised B*35

binding motif, several pieces of evidence presented here suggest that this peptide is presented by HLA-B*35. First, recognition of the relevant Gag OLPs in subjects with HLA-B*35 in the Durban cohort is strongly statistically significant. Second, in two Thames Valley Cohort subjects, ELISpot screening of the 9-mer peptide and truncations support HA9 as the most likely optimal epitope. Finally, binding data confirm that this epitope is bound by HLA-B*35 with KD values similar to other HLA-B*35 epitopes.

Characterising sites and consequences of HLA-B*35 selection pressure

HLA-B*35 epitopes HA9 and PY9 both lie in polymorphic regions of p24 Gag, causing potential loss of statistical power in attempts to define novel HLA-B*35 escape mutations. However, sequence analysis of C-clade infected cohorts demonstrated immune selection pressure imposed by HLA-B*35 operating in p24, RT and Nef. Among the strongest associations was the selection of the D260E mutation in the Gag PY9 epitope, highlighting the dominance of the PY9 response in these subjects. In previously published B-clade data, where glutamic acid (E) is the consensus amino acid at Gag-260, the presence of consensus is statistically associated with HLA-B*35 (Brumme, Tao *et al.* 2008), suggesting this allele may have contributed to the overall accumulation of E260. Thus, overall, there is evidence for the role of HLA-B*35 in driving HIV sequence evolution over time. The overall consequence is the potential loss of this Gag epitope, leading to reduction in control of viraemia.

Location and significance of amino acid polymorphisms in the HA9 epitope

Identifying sites of HLA-B*35 selection pressure within and flanking this epitope is difficult due to the marked variability of this region (HA9 overlaps with the location of the Gag Cyp-A binding loop, a well-recognised region of variability (Gatanaga *et al.* 2006)), and by the presence of overlapping epitopes restricted by other alleles (www.hiv.lanl.gov). No polymorphisms within the HA9 epitope were found to be associated with the presence of HLA-B*35, although a downstream mutation was identified at position Gag-232 in the extended Durban cohort.

Common polymorphisms in this epitope (H219Q and I223V) are shown to be associated with improved binding. Although the consequences of this effect are unknown, this is a potential example of how HLA-mediated selection (in this case, B*57-mediated selection of H219Q as a compensatory mutation) may alter epitopes available for targeting by other alleles (in this case, affecting the HLA-B*35 HA9 epitope).

Mechanism of immune escape in association with D260X substitution

ELISpot data presented here suggest a significant diminution in recognition of the PY9 epitope following D260E substitution in an individual subject with C-clade infection. This suggests that the polymorphism either impacts on HLA binding of this epitope, or on TCR recognition. Binding studies suggest no significant difference between the variants, leading to the overall conclusion that the mechanism of escape associated with the D260E substitution is likely to be via reduction in TCR recognition.

Impact of D260X on recognition of PY9 epitope

At a cross-cohort level, the Gag D260E substitution was associated with a trend towards reduction of ELISpot responses to the relevant OLP, although this did not reach statistical significance. Due to strong immune selection pressure, the D260E mutation is likely to occur early in the context of infection (substantiated by its appearance in ≥75% of HLA-B*35-positive study subjects in chronic infection). Therefore, by the time of recruitment to the study, many subjects may have already acquired the mutation. However, some circulating $CD8^+$ T cells may still recognize the wild-type variant, leading to disparity between the predominant sequence detected by population sequencing (bearing a mutation) and yet the persistence of a response to the infecting (wild-type) sequence. Further study of sequence changes and ELISpot responses in the setting of acute infection with longitudinal follow-up, and studies using clonal sequence analysis, would help to inform this analysis.

Co-variation of p24 Gag polymorphisms

Cross-sectional analysis of combinations of PY9 mutations suggests selection of D260X prior to I256X in subjects with HLA-B*35, possibly pointing to a fitness cost of I256X alone. In addition, there are multiple amino acid residues in p24 that co-vary with the HLA-B*35-associated escape mutations I256X and D260X. In particular, there is a strong association between these mutations and the presence of escape mutations selected by HLA-B*57. This finding demonstrates that many HLA-selected polymorphisms are inter-dependent, with compensatory mutations particularly occurring in the three tropism loops in p24 in which variation can most be tolerated.

6.4.3 Use of ELISpot data to determine HLA-B*35-restricted responses in B and C-clade infection

The ELISpot data presented here point to a preference for HLA-B*35 restricted responses to Gag and Pol in subjects with C-clade infection, in contrast to the Env and Nef targeting seen in B-clade. This observation suggests that the mechanism for the favourable effect of HLA-B*35 alleles in C-clade infected subjects is by providing increased breadth of Gag-specific responses, known to be associated with enhanced control of viraemia (Edwards *et al.* 2002; Borghans *et al.* 2007; Geldmacher *et al.* 2007; Honeyborne *et al.* 2007; Kiepiela *et al.* 2007; Brumme, Tao *et al.* 2008; Matthews *et al.* 2008). Similarly, the more favourable control of viraemia in subjects with Py than Px genotypes has been previously attributed to higher proportions of Gag-specific $CD8^+$ T cell responses (Jin *et al.* 2002).

Although subjects with both B and C-clade infection make responses to the OLPs containing the HA9 epitope, the optimal epitope was only recognized by subjects with B-clade infection. However, irrespective of this differential targeting of this epitope between clades, the overall trend was in favour of Gag and Pol responses in C-clade infection, associated with better disease outcomes.

Small study numbers limit the statistical power of the optimal ELISpot data; repeating analyses in extended cohorts would add further confidence to the findings. In particular, no conclusions can be drawn about differences between HLA-B*3501 and the less common HLA-B*35-Px variants, HLA-B*3502 and HLA-B*3503, due to small numbers of subjects with these alleles. In order to extend the analysis in future,

subjects with the closely related HLA-B*5301 allele could be added to the B*35-Px group.

6.4.4 Summary comments

Overall, these data highlight that alleles may operate differently in different populations, suggesting that caution should be exercised in making generalisations about the impact of a given allele on viraemic control. In the case of HLA-B*35, the likely reason for this difference is the skew towards targeting of Gag and Pol epitopes in C-clade and towards Nef and Env in B-clade. Targeting of the Gag PY9 epitope in C-clade infection may be particularly critical in viraemic suppression.

These studies support the view that the clade of the population targeted should be taken into consideration when selecting immunogens for vaccine design. Vaccine strategies may also need to account for longitudinal sequence changes that can be predicted to occur over time, exemplified by potential loss of the wild-type PY9 response due to accumulation of polymorphisms in the epitope.

CHAPTER 7: *In vitro* quantification of the $CD8^+$ T cell response using Viral Suppression Assays (VSA)

7.1 INTRODUCTION

7.1.1 Limitations of ELISpot assays in assessing the $CD8^+$ T cell response

IFN-γ ELISpot assays are among the most widely used methods for evaluating the $CD8^+$ T cell response *in vitro*. However, the results are generally not well correlated with individual disease outcomes *in vivo,* either in terms of breadth or magnitude of responses (Gea-Banacloche *et al.* 2000; Betts *et al.* 2001; Addo *et al.* 2003; Chen *et al.* 2009). Furthermore, in recent T cell vaccine trials, ELISpot responses induced in vaccinees failed to correlate with disease protection or with disease control following infection (Johnston and Fauci 2007; McElrath *et al.* 2008).

The limitations of ELISpot assays may be explained both by the persistent production of IFN-γ even in functionally exhausted T-cells (Bennett *et al.* 2008), and by methodology that exposes cells to supra-physiologic concentrations of synthetic peptide, thus circumnavigating the critical steps of infection, antigen processing and presentation that must occur for T cell recognition to take place *in vivo* (Chen *et al.* 2009). Therefore, other methods that focus on the killing or suppressive capacity of $CD8^+$ effector cells - rather than on quantifying the breadth or magnitude of $CD8^+$ T

cell responses - are crucial for assessing disease control in both natural and vaccine-induced immunity.

7.1.2 Use of Viral Suppression Assays (VSA) for quantifying the $CD8^+$ T cell response

A method for measuring HIV suppression *in vitro*, using co-culture of HLA-transfected target cells with $CD8^+$ T cell clones, was first published in 1997 (Yang *et al.* 1997). In these assays, viral growth in target cells was measured in the presence and absence of HLA-matched $CD8^+$ effector cells. Using this method, Yang *et al.* demonstrated viral suppression mediated by a single $CD8^+$ T cell clone, attributable to HLA Class-I restricted cytolysis. More recently, these assays were adapted to use autologous $CD4^+$ T cells as infected targets and viral suppression was again measured following co-culture with $CD8^+$ T cell lines and clones (Chen *et al.* 2009). Similar techniques have also been employed in the study of SIV-specific T cell clones in the macaque model (Loffredo *et al.* 2005; Chung *et al.* 2007).

7.1.3 Hypothesis and aims

The general aim of the studies in this chapter were to optimise an *in vitro* system, based on previously published methods (Yang *et al.* 1997), for assessing viral suppression mediated by HLA-Class I restricted $CD8^+$ T cell responses, and to investigate the extent to which the results of these assays correlate with disease outcome *in vivo*.

Specific aims were considered in further developing these assays:

i. To broaden the scope of the assay by using bulk autologous $CD8^+$ T cells as effectors, such that the full repertoire of $CD8^+$ T cell responses can be assessed, followed by experiments to isolate the impact of individual Class I restricted responses from within this total response.

ii. To optimise the assay for direct *ex vivo* use, with the aims both of reducing the outgrowth of autologous virus and minimising phenotypic changes that inevitably occur in cultured cell populations over time.

iii. To address the question of whether Gag-specific $CD8^+$ T-cell responses are more effective at suppression of viral replication than non-Gag-specific responses.

7.2 METHODS

7.2.1 Cell enrichment protocol and set-up of VSA

The methods used are based on those previously published (Yang *et al.* 1997; Chen *et al.* 2009). However, cell enrichment was undertaken using magnetic cell separation (MACS) technology rather than using bi-specific antibodies as most recently described (Chen *et al.* 2009). Basic methods are given in Chapter 2, (section 2.13).

Broadly speaking, the assays rely on enrichment of a purified $CD8^+$ T cell population and an autologous target population enriched for $CD4^+$ T cells. The latter population is superinfected with laboratory virus, and viral growth is measured in the presence and absence of the $CD8^+$ T cell fraction. All study subjects in this chapter were recruited from the Thames Valley Cohort.

7.2.2 Refinements to methods for cell enrichment

Positive vs. negative selection of T cell populations

To optimise the enrichment of pure $CD4^+$ and $CD8^+$ T cell populations, while minimising cell loss, a comparison of positive and negative selection was undertaken on HIV+ subjects (Fig. 7.1; table 7.1). Based on the data summarised in Fig. 7.1 and Table 7.1, and further experimental optimisation (data shown in Appendix VI), a final algorithm for cell enrichment was produced (Fig. 7.2).

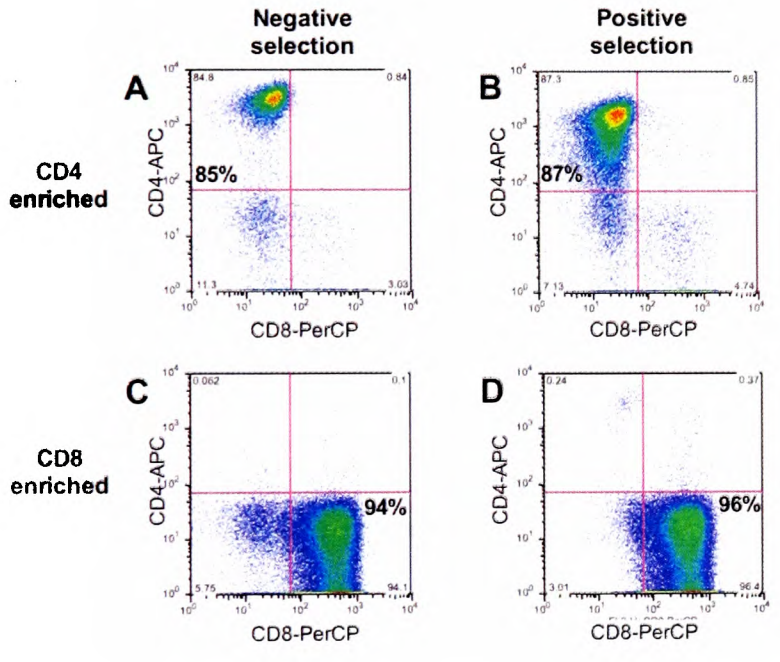


Figure 7.1: Characteristic FACS plots showing results of simultaneous enrichment of $CD8^+$ (A, B) and $CD4^+$ (C, D) cell populations by negative (A, C) and positive (B, D) selection. Enrichment undertaken from fresh PBMC from HIV-infected subject, N041.

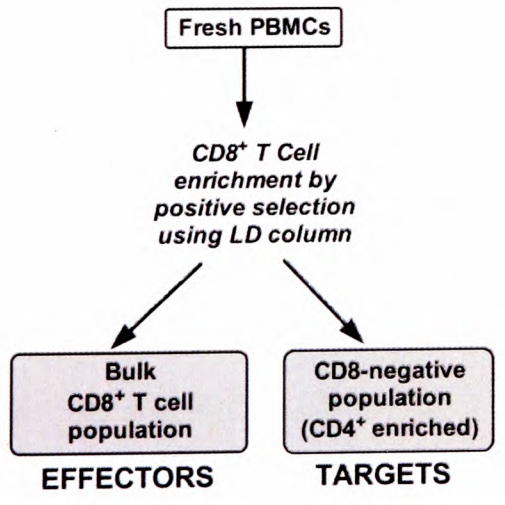


Figure 7.2: Algorithm for derivation of enriched $CD8^+$ and $CD4^+$ T cell populations from PBMCs.

Consideration	Positive selection	Negative selection
Purity of yield attained following enrichment	Higher purity of yield, as only the specifically labelled cells are retained.	Potential for lower purity of yield, as some cells may not be successfully labelled by an antibody cocktail. Susceptible to reduction in yield when cell ratios are altered (E.g. altered $CD4^+$:$CD8^+$ ratio seen in HIV).
Features of population to be sorted	Optimum for selection of rare cells.	Optimum when no antibody label available for the desired cell population.
Subsequent growth in culture	Selected cells selected carry a magnetic label, with the potential for interference with subsequent cell-cell interactions in culture*.	Cells selected are 'untouched' by beads, a theoretical advantage for subsequent culture.
Ease / speed of enrichment process	Quick, single-step labelling process. Additional advantage that this reduces number of washing steps, therefore minimizing cell loss.	Longer two-step process. More washing steps lead to potentially greater cell loss.
Cost	Cheaper**.	More expensive**.

Table 7.1: Comparison of positive and negative cell-selection strategies using MACS

* Manufacturer states that microbeads are biodegradable, and disappear after a few days in culture.

** Based on price per Miltenyi MACS kit, July 2009.

7.2.3 Viral growth in culture

Subjects were infected with wild-type (WT) B-clade virus irrespective of the clade of their autologous infection. At a multiplicity of infection (MOI) of 0.01 and 0.001, logarithmic growth of virus was observed over 0-4 days, followed by gradual establishment of a plateau phase up to day 12 (Fig. 7.3). In subsequent experiments throughout this chapter an intermediate MOI of 0.03 was used.

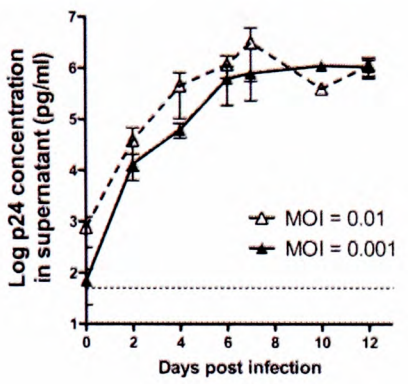


Figure 7.3: Effect of infection at different MOI (Multiplicity of Infection) on viral growth in culture.

Mean ± SEM shown for four subjects over 12 days in culture. Dashed line shows lower threshold of p24 antigen detection by the assay. Median growth at 7 days 5.89log for MOI of 0.01 compared to 5.37log for MOI of 0.001.

7.2.4 'Subtraction' approaches to quantify role of individual $CD8^+$ T cell responses

In order to assess the contribution of individual epitope-specific $CD8^+$ T cell responses to the total viral suppression brought about using the bulk $CD8^+$ T cell population, two 'subtraction' approaches were considered (so named as each aims to subtract a given epitope-specific response from the total):

i. Use MHC Class I tetramers to deplete an epitope-specific $CD8^+$ T cell fraction;

ii. Use a mutant (MT) virus bearing escape mutations known to abrogate a specific $CD8^+$ T cell response.

Tetramerisation of MHC Class I monomers

MHC Class I monomers were tetramerised on Extravidin-PE and tested using anti-HLA beads by FACS analysis (Fig. 7.4).

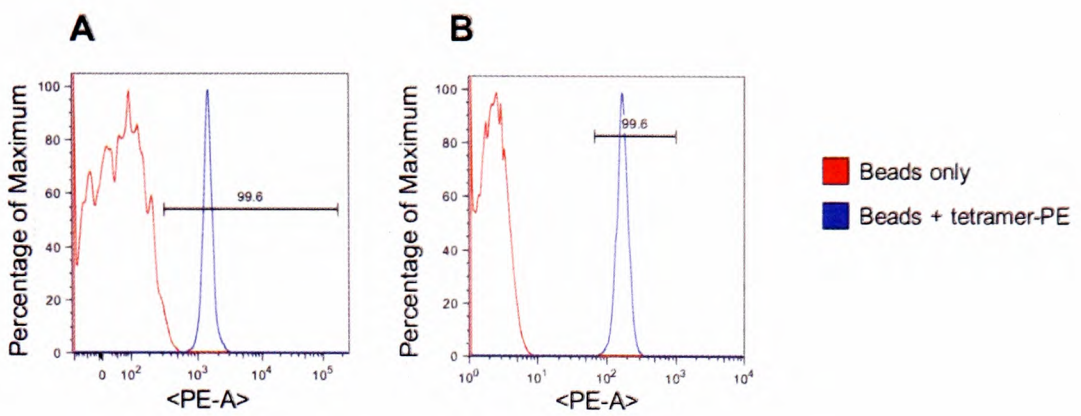


Figure 7.4: Histograms generated by FACS to show tetramerisation of Class-I MHC tetramers on extravidin-PE. Right-shift of peak (blue) indicates successful tetramerisation compared to negative control (red). A: HLA-B*57-KF11. B: HLA-B*35-PY9. Gated on FITC-positive beads.

Depletion of specific $CD8^+$ T cell populations was undertaken by incubating PBMCs with MHC Class I tetramer as described in Chapter 2 (section 2.10.3). The enrichment protocol was undertaken according to the algorithm shown in Fig. 7.5.

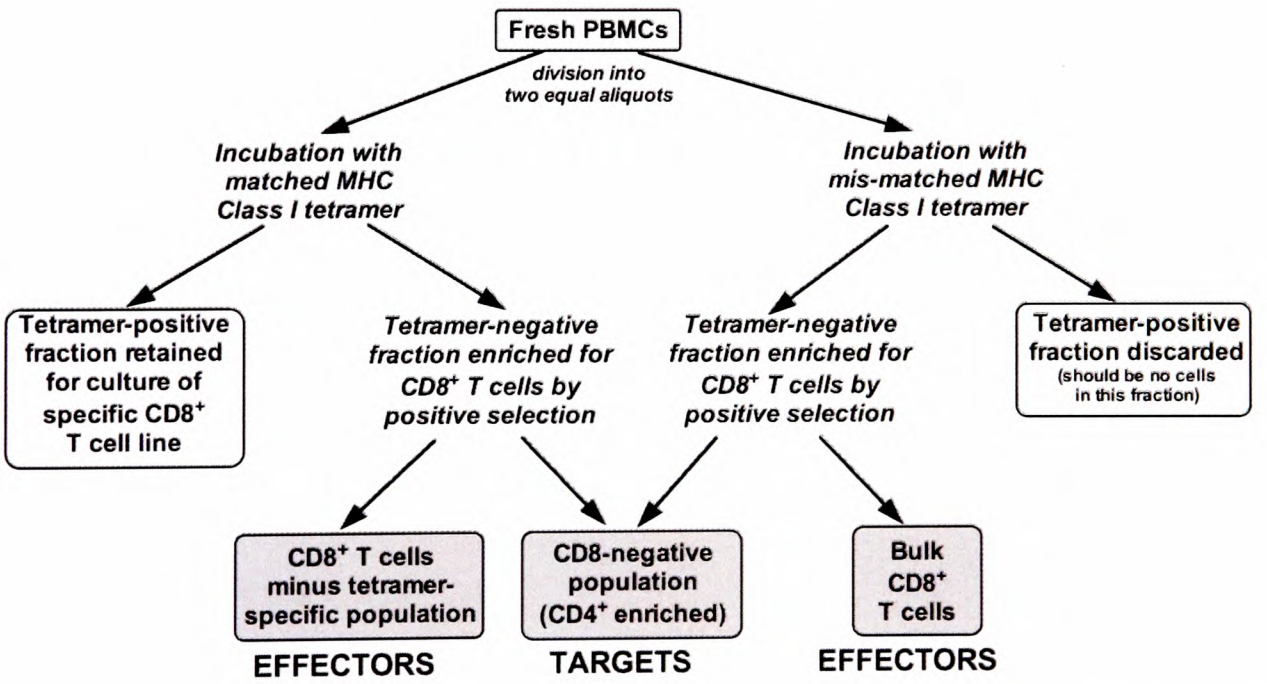


Figure 7.5: Algorithm showing protocol for enrichment of CD4 and CD8 T cell populations, with and without tetramer depletion. In order to rule out a non-specific effect of the tetramer on the depleted $CD8^+$ T cell fraction, the 'bulk' fraction was treated in parallel with a mismatched tetramer.

(ii) KF11 mutant virus

A KF11 mutant virus was synthesised by Julia Prado, using site-directed mutagenesis to introduce mutations A163G and S165N into the HLA-B*57 KF11 epitope (KAFSPEVIPMF, Gag 162-172) (Fig. 7.6), as used in published fitness assays (Crawford *et al.* 2007). This double escape variant was selected for use in subtraction assays in order to focus on a commonly targeted epitope, while minimising the cost to viral fitness (the double mutant replicates better than the A163G variant (Crawford *et al.* 2007)).

	133	143	153	163	173	183
NL4-3 sequence	PIVQNLQGQM	VHQAISPRTL	NAWVKVVEEK	AFSPEVIPMF	SALSEGATPQ	DLNTMLNTVG
Wild-type GFP (WT)	----------	----------	----------	----------	----------	----------
KF11 mutant GFP (MT)	----------	----------	----------	G-N-------	----------	----------

Figure 7.6: Gag p24 sequences of wild-type (WT) and mutant (MT) viruses used for viral suppression assays, aligned to NL4-3. Numbering with respect to HXB2. Sequences derived from viral RNA extracted from culture supernatant and amplified using B-clade PCR primers.

7.2.5 Analysis of p24 ELISA results

Viral p24 antigen concentration was measured by ELISA (see Chapter 2 section 2.13.3). Some inter-assay variability occurred in the thresholds for p24 detection. In order to standardise the estimation of a lower threshold, the mean of 20 assays was calculated as 5.00 pg/ml. The majority of these assays were performed in 200 µl total volume, and the minimum p24 detection was based on a 1:10 dilution (i.e. 50 pg/ml). For p24 antigen results below the lower threshold of detection (assumed to fall in the range 0-50 pg/ml), the median value of 25 pg/ml was used to plot growth curves.

Viral suppression was measured according to two end-point measures, based on viral growth (measured by the concentration of p24 antigen in culture supernatant) in the presence and absence of an effector ($CD8^+$ T cell) population:

i. The magnitude of suppression after 7 days in culture. This was calculated as: $\log_{10}$ (p24 concentration in absence of $CD8^+$ T cells) – $\log_{10}$ (p24 concentration in presence of $CD8^+$ T cells).

ii. The gradient of the slope of viral growth during the exponential growth phase was calculated from viral growth at days 0, 2 and 4 (determined in section 7.3.1 below), using regression analysis to fit the data to an exponential curve (LOGEST function in Microsoft Excel). The natural log of the slope value was then calculated (LN function in Microsoft Excel).

7.3 RESULTS

7.3.1 *In vitro* HIV infection of enriched $CD4^+$ T cells results in a logarithmic growth phase up to day 4

Initial experiments established that successful *in vitro* infection of target cells with WT HIV could be achieved in cells obtained from subjects with and without underlying HIV infection.

Determination of the duration of the logarithmic growth phase was required in order to perform further analysis based on the slope gradient. In these experiments, this exponential phase spanned days 0-4 (Fig. 7.7). Outgrowth of autologous virus was

most pronounced in subjects with viral loads >2000 RNA copies/ml plasma at the time of the assay (Fig. 7.7C).

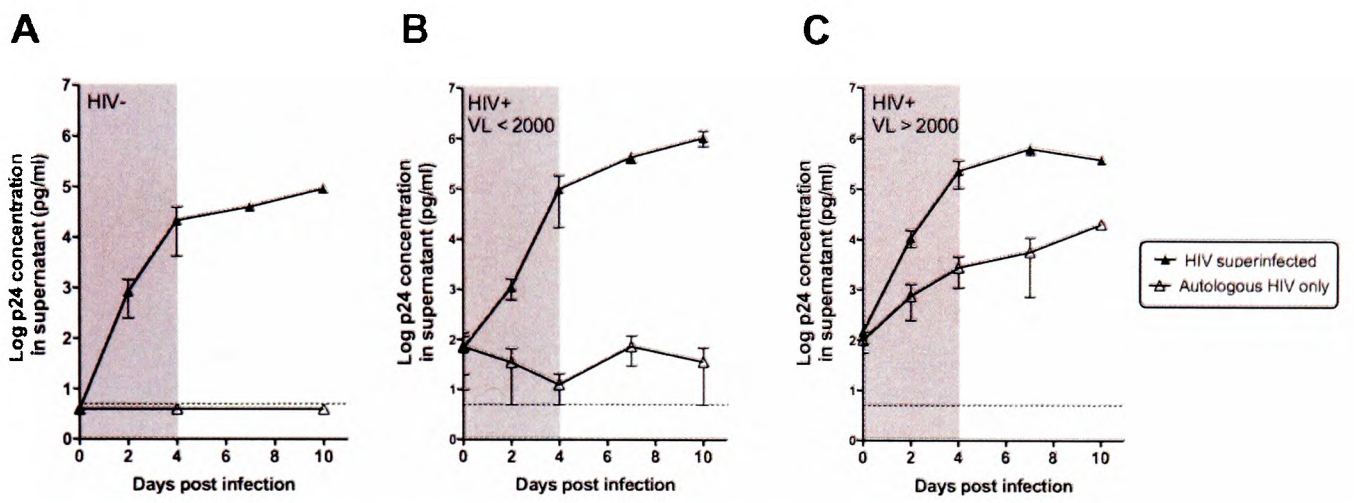


Figure 7.7: Measurement of viral growth by p24 ELISA following *in vitro* infection of CD4$^+$ target population, in cells obtained from study subjects with and without HIV infection. A: 2 HIV-negative donors. B: 7 HIV positive subjects with VL <2000 RNA copies/ml plasma (median 732, IQR 133-1836). C: 8 HIV positive subjects with VL >2000 RNA copies/ml plasma (median 69986, IQR 8447-113121). Grey bars show estimated duration of viral logarithmic growth phase. Dotted line shows mean lower threshold of p24 detection. Error bars show mean ± SEM of 3 repetitions, apart from D10 timepoint in panels A and C where only one subject was represented in each case.

7.3.2 Viral load (VL) and CD4$^+$ T cell count predict the outgrowth of autologous virus from enriched CD4$^+$ cells in culture

The outgrowth of autologous virus in these assays potentially confounds results obtained for viral suppression. Therefore, factors predicting the detection of autologous virus in the absence of an effector (CD8$^+$) T cell population were further investigated.

First, a correlation was sought between VL and autologous p24 detection in an extended cohort of 27 subjects. Overall, a significant correlation was observed (Fig.

7.8). However, detection of autologous virus is not inevitable even in cells from subjects with higher VL, and conversely, subjects with lower VL (<10^3 RNA copies/ml) may still produce detectable p24 antigen.

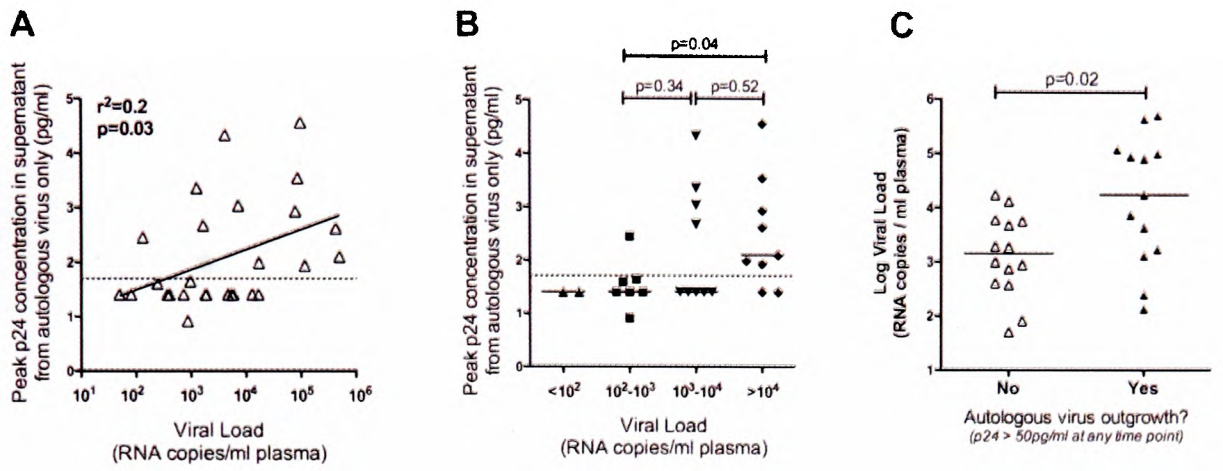


Figure 7.8: Relationship between viral load (VL) and p24 detection from autologous virus over 7 days in culture in cells obtained from 27 subjects. The mean p24 concentration was taken from 1-3 repetitions of each experiment, and the peak was taken as the highest of these mean values over 7 days. Dotted line shows mean lower threshold of p24 detection. A: Relationship between VL and p24 by linear regression. B: Subjects grouped according to VL, p values by Mann-Whitney test. C: Subjects grouped according to detection of p24 antigen from autologous virus above mean lower threshold for detection (50pg/ml), p value by Mann-Whitney test.

As expected, the reciprocal relationship was observed for $CD4^+$ T cell counts, such that higher CD4 counts were predictive of less autologous viral p24 detection (r^2=0.16, p=0.04, linear regression, data not shown). This is likely to be due to the well-established inverse relationship between CD4 count and VL (in this group of 27 patients, p=0.001, linear regression) rather than being an independent predictor of the growth of autologous virus.

As B-clade viruses are often considered 'fitter' or more pathogenic than A and C-clade strains (Abraha *et al.* 2009; Kanki *et al.* 1999; Spira *et al.* 2003), clade of

autologous infection was also considered as a possible factor predicting viral outgrowth. However, comparing 8 subjects with B-clade infection with 13 C-clade infected subjects, there was no significant difference in detection of autologous virus over seven days. In considering *in vitro* factors, viral outgrowth from $CD4^+$ T cells enriched from fresh or frozen PBMCs was also not statistically significant (table 7.2).

Factor predicting growth of autologous virus	p value (statistical test used)
Viral load	0.03 (linear regression)
$CD4^+$ T cell count	0.04 (linear regression)
Clade of autologous virus (B vs. non-B)	0.4 (Mann-Whitney test)
Fresh/frozen PBMC	0.2 (Mann-Whitney test)

Table 7.2: Factors potentially associated with outgrowth of autologous virus from $CD4^+$ T cells in culture. Based on the detection of autologous virus by p24 ELISA in up to 10 days in culture.

7.3.3 Inhibition of viral growth is not mediated by the addition of non-HIV-specific $CD8^+$ T cells

Prior to running co-culture experiments from HIV-infected subjects, preliminary experiments were carried out in HIV-negative subjects in order to establish that $CD8^+$ T cells were not mediating an inhibitory effect independent of HIV-specific responses. There was no inhibition of viral growth on co-culture of infected target cells with enriched effectors, even at effector:target (E:T) ratios >1, and/or in the presence of cellular activation by exogenous IL-2 and PHA (Fig. 7.9). Further experiments were all performed in 'CD4 culture media' (R10 + IL-2 + PHA, see methods section 2.14).

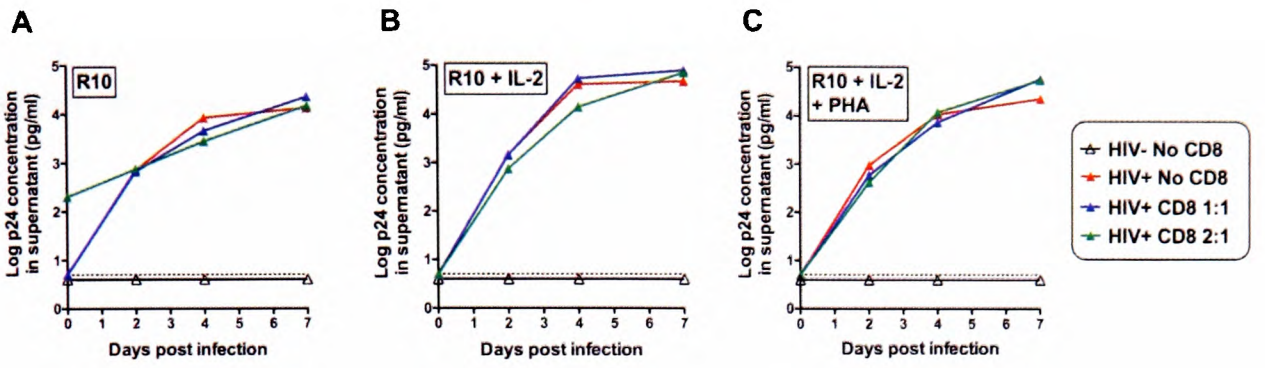


Figure 7.9: *In vitro* infection of HIV-negative target cells, showing equivalent viral replication in the presence and absence of autologous CD8$^+$ T cells in different culture media. Dotted line shows mean lower threshold of p24 detection. Effector to target ratio at 1:1 (blue) and 2:1 (green). A: Co-culture in R10 media only. B: Co-culture in R10 + IL-2. C: Co-culture in R10 + IL-2 + PHA.

7.3.4 Inhibition of viral growth in superinfected cells is mediated by HIV-specific CD8$^+$ T cells in a dose-dependent manner

Viral inhibition assays were optimised using different E:T ratios (Fig. 7.10), to investigate the dose-response relationship.

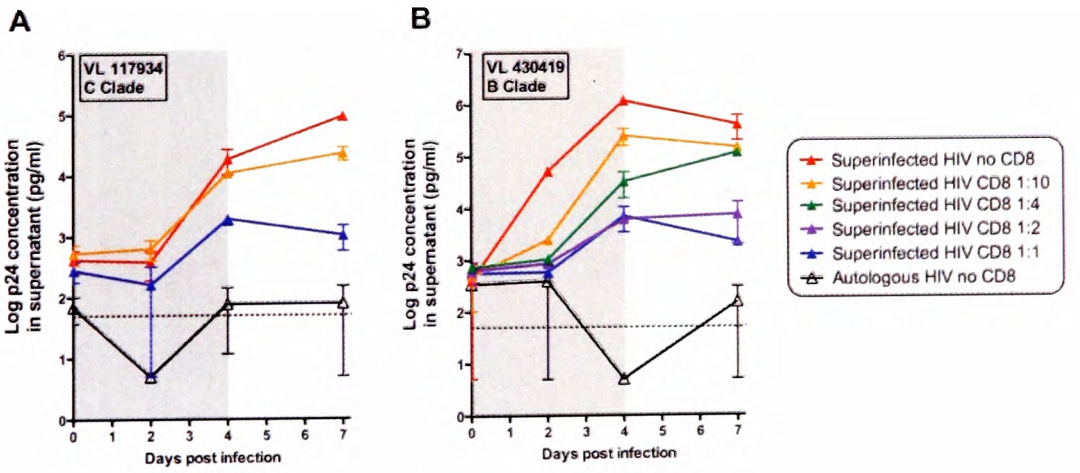


Figure 7.10: Viral suppression over 7 days by bulk CD8$^+$ T cells at different E:T ratios. Experiment undertaken in triplicate in C-clade infected subject R060, and B-clade infected subject R050. Error bars show mean ± SEM. As previously, the logarithmic growth phase of the virus is between days 0-4 (grey panels).

For further experiments, E:T ratios of 1:1 were used, based both on the approximate physiologic ratio of bulk CD8: CD4 seen *in vivo*, and the optimum suppression detected by these *in vitro* assays. This E:T ratio is also comparable to that used in recently published studies (Chen *et al.* 2009).

7.3.5 Significant viral suppression occurs *in vitro* in the presence of autologous $CD8^+$ T cells

Viral suppression assays were performed over 0-7 days in 19 HIV-infected study subjects (median VL 1791 RNA copies/ml, IQR 380-63697). As previously, the exponential growth phase of superinfecting virus occurred between days 0-4 (Fig. 7.11). Significantly greater suppression was measured at an E:T ratio of 1:1 compared to 1:10 at all time-points (Fig. 7.11A and 7.12). Over the exponential growth phase (day 0-4), suppression at an E:T ratio of 1:10 was not statistically significant, while an E:T ratio of 1:1 produced highly significant suppression (Fig. 7.11B and 7.12).

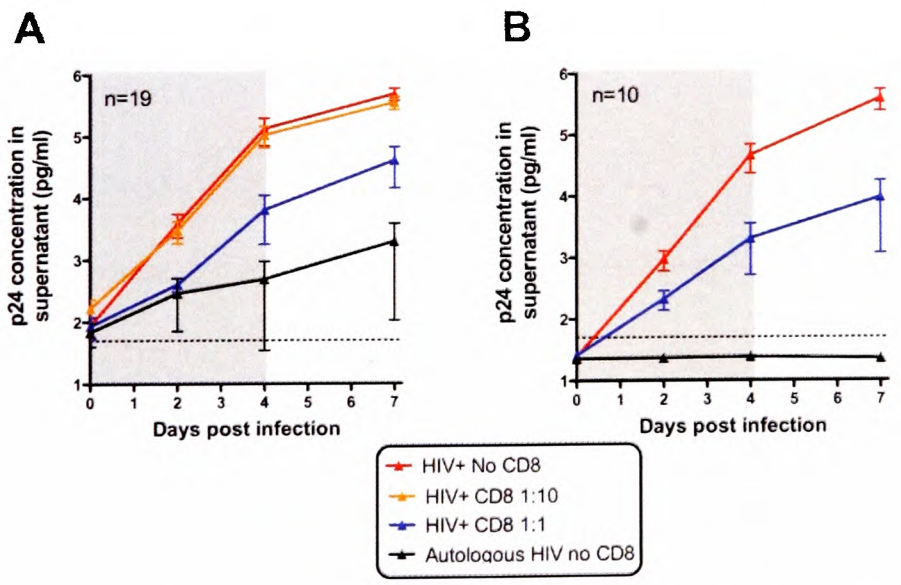


Figure 7.11: Viral suppression data from 19 HIV-infected study subjects. Dotted line shows mean lower threshold of p24 detection. Error bars show mean ± SEM of 3 replicates. Grey bars show estimated duration of viral logarithmic growth phase. A: Pooled data from all 19 subjects (except E:T ratio of 1:10 limited to 10 subjects). B: Data limited to subset of 10 subjects in whom no outgrowth of autologous virus was observed.

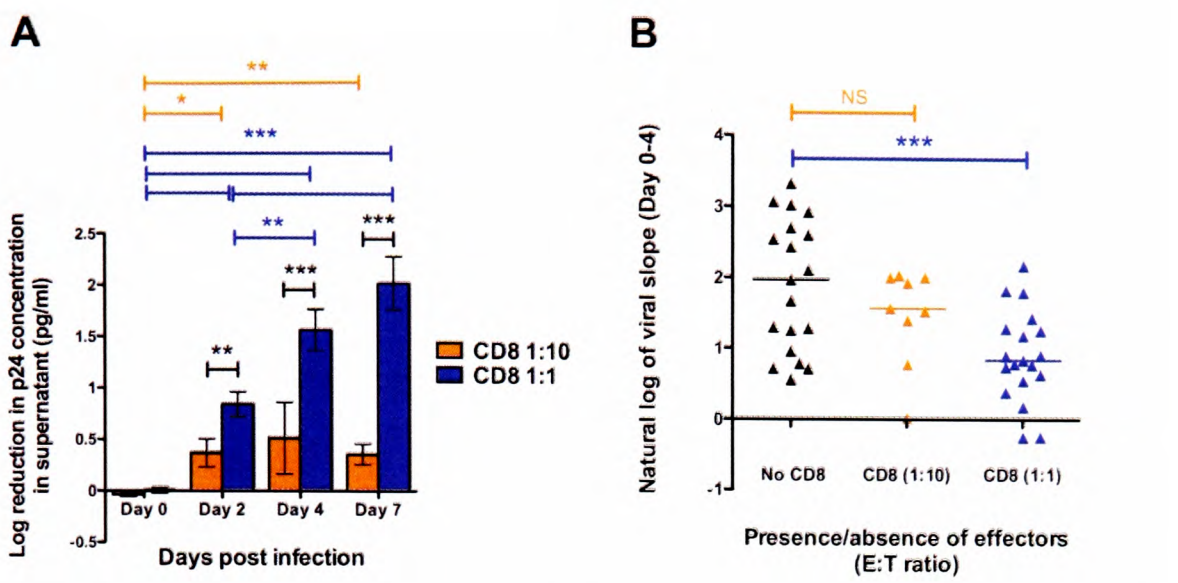


Figure 7.12: Viral suppression in 19 subjects quantified by two different methods. A: Log reduction in p24 (compared to superinfection in the absence of CD8+ T cell effectors). B: Natural log of viral slope in the presence and absence of effector population. p values by paired t-test; *p<0.05, **p<0.001, ***p<0.0001.

7.3.6 Magnitude of viral suppression is predicted to varying extents by HLA genotype and clade of autologous infection

Several factors were considered to be possible predictors of *in vitro* control of viraemia:

i. Viral load of study subject
ii. Presence/absence of outgrowth of autologous virus in culture
iii. HLA-type
iv. Clade of autologous infecting virus
v. Breadth of ELISpot responses to Gag

In this section, the impact of each of these variables is assessed using data from viral suppression assays carried out in 19 study subjects. Clearly, these factors may not be

independent of one another, but the number of study subjects was too small to permit individual sub-group analysis in many cases.

(i) Viral load (VL) of study subject

Unexpectedly, no relationship was identified between *in vitro* viral suppression and either VL (Fig. 7.13) or CD4 count (data not shown, p=0.64 linear regression).

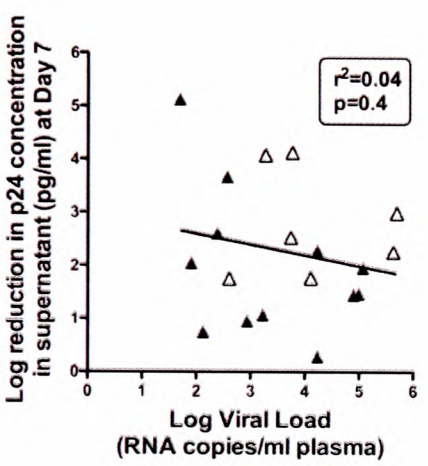


Figure 7.13: Relationship between VL of study subject and *in vitro* viral suppression.
Open symbols: B-clade autologous infection. Closed symbols: non-B-clade autologous infection. p value by linear regression for all 19 subjects. Repeating the analysis just for the subset of 7 subjects with B-clade infection, there was still no significant correlation (r^2=0.002, p=0.9).

(ii) Outgrowth of autologous infecting virus

Outgrowth of autologous virus has already been shown to be related to viral load (Fig. 7.8). However, the presence of detectable autologous virus did not significantly increase the total peak p24 detected at day 7 in culture, and no statistically significant difference was detected in the extent of viral suppression measured in experiments with and without the appearance of autologous virus (Fig. 7.14).

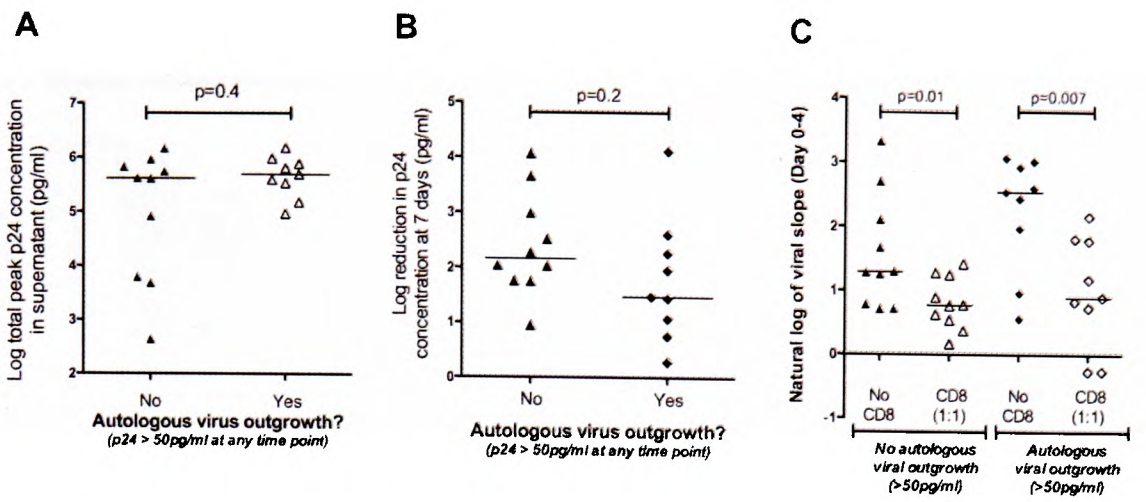


Figure 7.14: Effect of detectable autologous virus on *in vitro* viral suppression. A: Comparison of peak p24 antigen detected according to the presence/absence of detectable autologous virus (p=0.4 Mann-Whitney test). B: Comparison between reduction in p24 concentration in 10 experiments with no autologous virus detected, compared to 9 experiments in which viral outgrowth was detected (p=0.2, Mann-Whitney test). C: Comparison between reduction in viral slope in 10 experiments with no autologous virus detected, and in 9 experiments in which viral outgrowth was detected (p values by paired t-test).

(iii) HLA type of study subject

In order to ascertain whether the HLA genotype of the study subject is related to improved control of viraemia, subjects were ranked according to the log suppression at 7 days in the presence of bulk $CD8^+$ T cells (Fig. 7.15). This suggests a possible weak trend towards better suppression in 7 subjects with HLA-B*57 alleles.

The subset of subjects with HLA-B*57 (n=7) was further analysed for evidence of improved viral suppression in this group compared to HLA-B*57-negative subjects. There was a trend towards improved *in vitro* viral suppression based on log reduction in p24 in these subjects, but this did not reach statistical significance (Fig. 7.16).

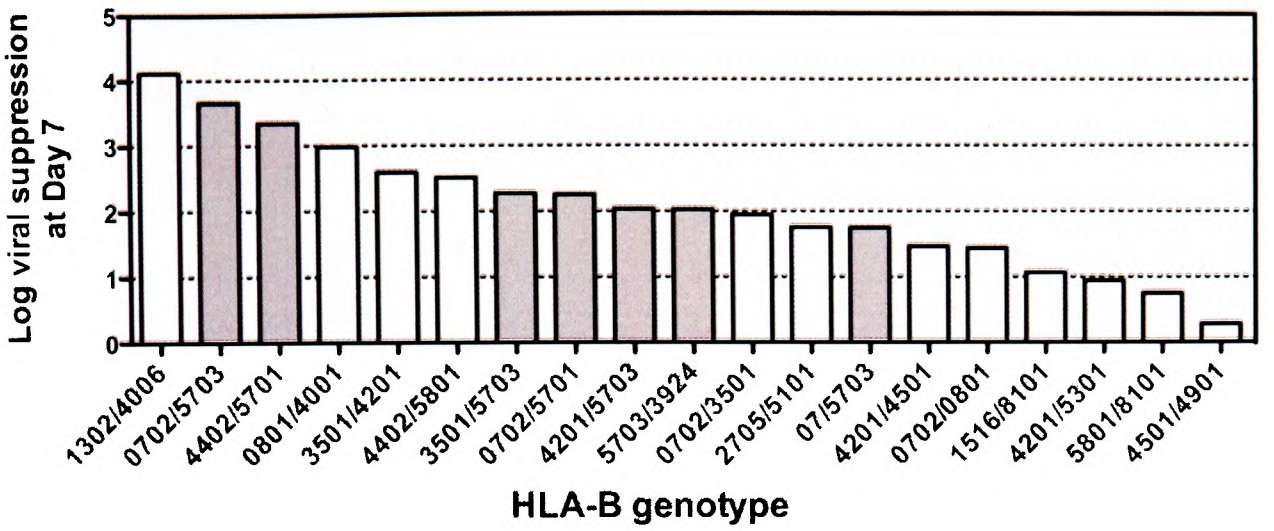


Figure 7.15: HLA-B genotype and extent of viraemic suppression at Day 7. Bars denoting subjects with HLA-B*57 are shaded grey.

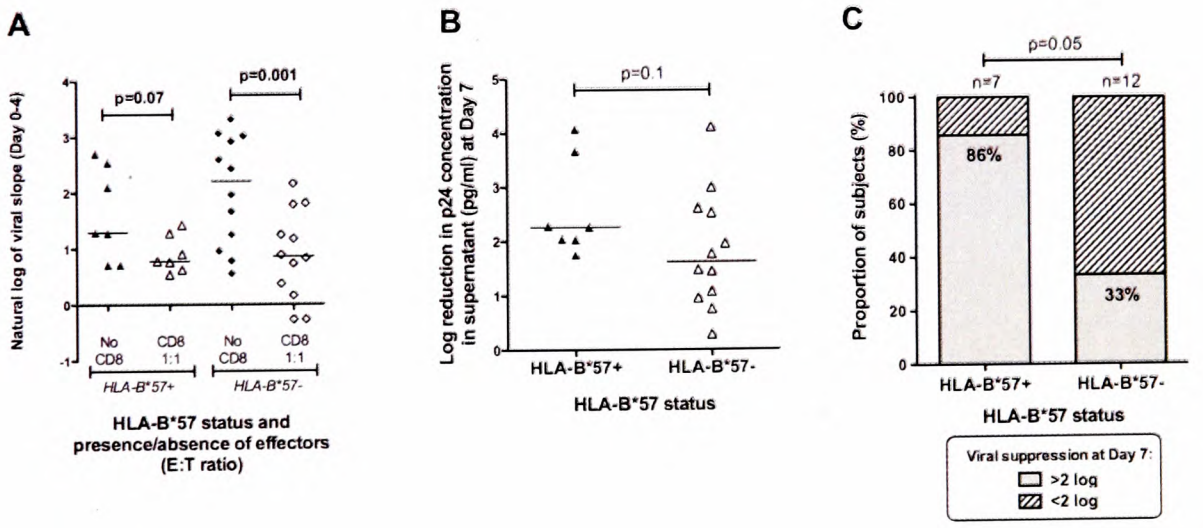


Figure 7.16: Relationship between HLA-B*57 and viral suppression. A: Viral slope in subjects with and without HLA-B*57 according to the presence or absence of effector ($CD8^+$) T cells, (p values by paired t-test). B: log reduction in p24 concentration in culture supernatant at day 7 in subjects with and without HLA-B*57 (p=0.1, Mann-Whitney test). C: Proportion of subjects with and without HLA-B*57 with >2 log suppression of viraemia at day 7 (p=0.05, Fisher's Exact Test).

(iv) Clade of autologous infecting virus

The clade of autologous infection is a likely predictor of the extent of suppression of a B-clade superinfecting virus, as subjects with autologous B-clade infection are most likely to possess $CD8^+$ T cell responses primed against the superinfecting strain.

As expected, there was a significant difference in viral suppression at day 7, with B-clade-infected subjects effecting better suppression (Fig. 7.17). These data suggest that a laboratory strain matching the patient's autologous infection should ultimately be used to optimise the specificity of measurement of viral suppression.

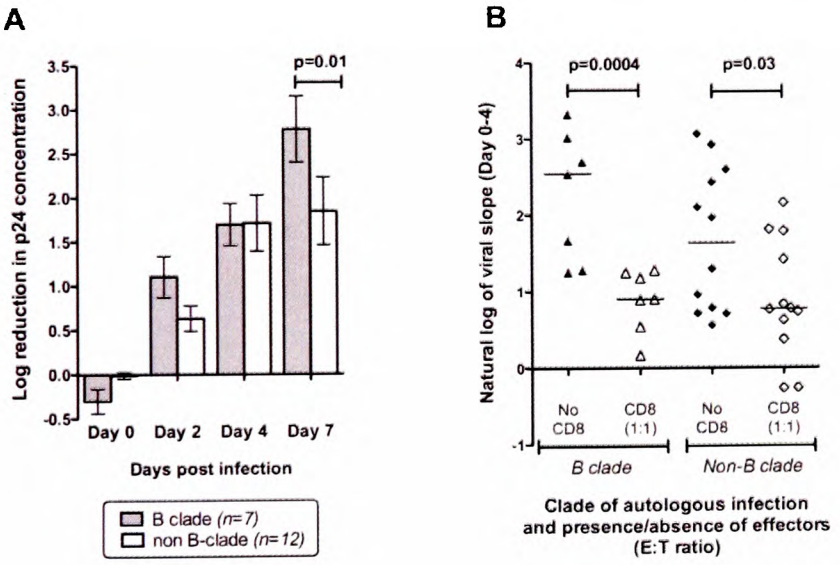


Figure 7.17: Relationship between clade of autologous infection and viral suppression following superinfection with NL4-3 (B-clade) HIV *in vitro*. A: Reduction in p24 concentration over 7 days. B: Difference between viral slope at Day 4 in the presence and absence of effectors in B and non-B clade infection.

(v) Breadth of ELISpot responses to Gag

Given the previously reported correlation between a greater breadth of ELISpot responses to Gag and improved control of viraemia (Edwards *et al.* 2002; Kiepiela *et al.* 2007), the relationship between these ELISpot responses and viral suppression was investigated.

In this small cohort (n=18) in keeping with previous data, there was an overall trend towards improved control of viraemia with increasing Gag responses although this did not reach statistical significance (Fig. 7.18A). However, unexpectedly, there was no relationship between the number of Gag ELISpot responses and viral suppression (Fig. 7.18B,C).

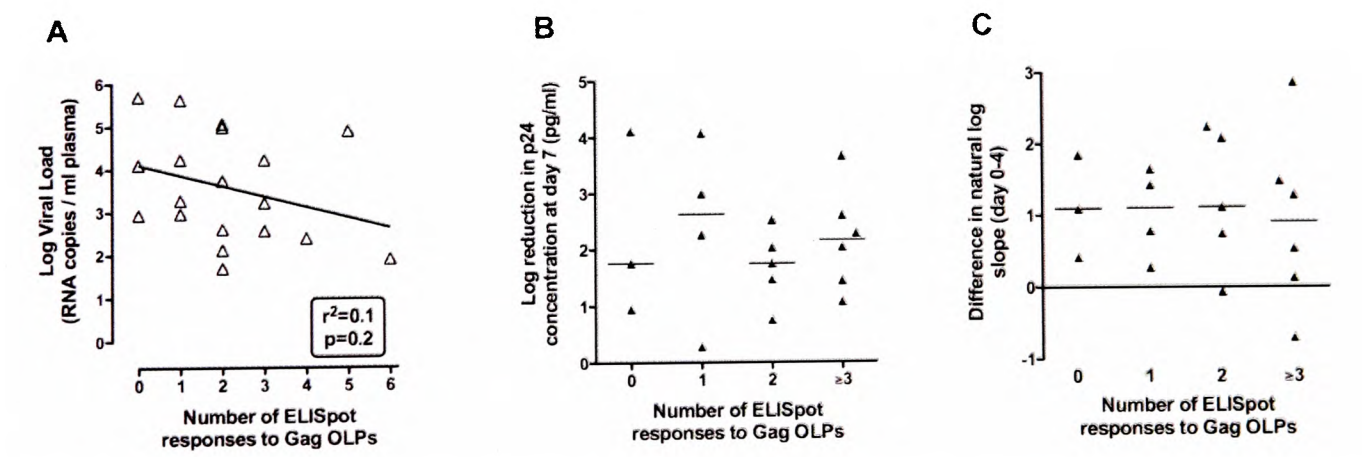


Figure 7.18: Relationship between number of ELISpot responses to OLPs in Gag and extent of viral suppression. A: Relationship between breadth of ELISpot responses to Gag and viral load (p=0.2 linear regression). B: Number of Gag ELISpot responses and log p24 reduction at day 7. C: Number of Gag ELISpot responses and reduction in viral slope (days 0-4).

7.3.7 Subtraction assays: tetramer-depletion approach

In order to investigate the contribution of the HLA-B*57 KF11 response to the overall magnitude of viraemic suppression, a subtraction approach was used. PBMCs from subjects with HLA-B*57 were screened for specific $CD8^+$ T cell responses using IFN-γ ELISpot assays to identify those with a dominant response to the Gag epitope KF11 (Table 7.3 / Fig. 7.19).

Subject ID	Clade	HLA type	Baseline VL (RNA copies/ ml plasma)	ELISpot responses to HLA-B*57 optimals in Gag
H024	?B	A*03/74 B*5703/07 Cw*0701/0702	2266	KF11 TW10
H037	C	A*2301/3002 B*3924/5703 Cw*0701/1801	50	KF11 TW10
LD011	C	A*2301/3004 B*0702/5703 Cw*0701/0802	372	ISW9 KF11 TW10
LD015	C	A*3601/74 B*5703/4201 Cw*1701/1801	81	KF11
R059	C	A*0103/6802 B*3501/5703 Cw*0401/0701	17000	ISW9 KF11

Table 7.3: Characteristics of Thames Valley Cohort subjects with HLA-B*57 used for viral suppression assays. Likely clade of infection is inferred from country of origin. Responses to optimal epitopes based on the same time-point as cells used in the viral suppression assay.

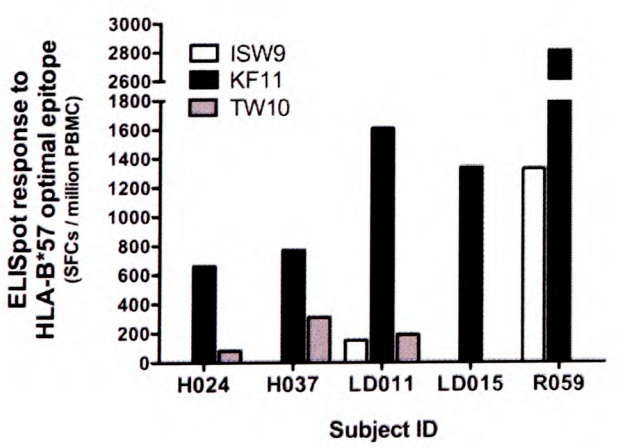


Figure 7.19: ELISpot responses to Gag optimal epitopes restricted by HLA-B*57 in five subjects. Screening was performed with clade-specific optimal peptides; responses are shown to the immunodominant Gag epitopes ISW9 (ISPRTLNAW Gag 147-155), KF11 (KAFSPEVIPMF, Gag 162-172) and TW10 (TSTLQEQIAW/TSTLQEQIGW, Gag 240-249).

For two subjects, H024 and H037, tetramer-depletion was undertaken from fresh PBMCs, leading to removal of the cells specific for this response from the bulk $CD8^+$ T cell population (Fig. 7.20). Co-culture experiments undertaken with bulk $CD8^+$ T cells and tetramer-depleted $CD8^+$ T cells demonstrated a small reduction in viral suppression in the absence of the KF11 response in two study subjects (Fig. 7.21). This was statistically significant only at Day 7 in subject H037 (Fig. 7.21D).

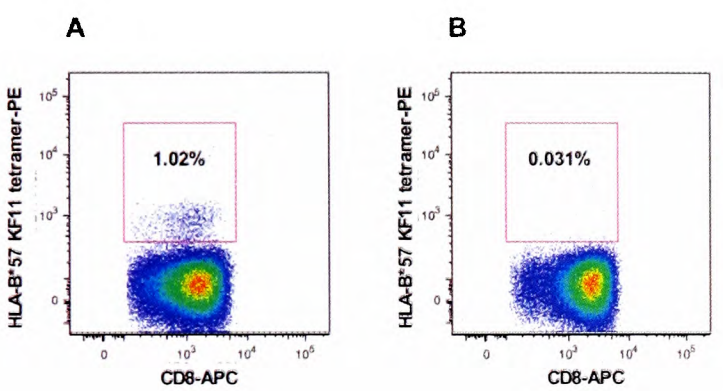


Figure 7.20: FACS plots showing tetramer-depletion of HLA-B*57 restricted KF11 response in subject H024. $CD8^+$ T cells enriched by positive selection. A: $CD8^+$ T cell population following incubation with a mismatched tetramer. B: $CD8^+$ T cell population following tetramer-depletion of the B57-KF11 response.

Based on the analysis in Chapter 6, further consideration was also given to the HLA-B*35 restricted response to the PY9 epitope (Gag 254-262) with a view to establishing the contribution of this response to overall viraemic suppression. Viral suppression assays from one study subject, R019, were carried out in quadruplicate. In this instance, there was no significant difference in viral inhibition in the presence/absence of the B*35-PY9 response at any time point (Fig. 7.22).

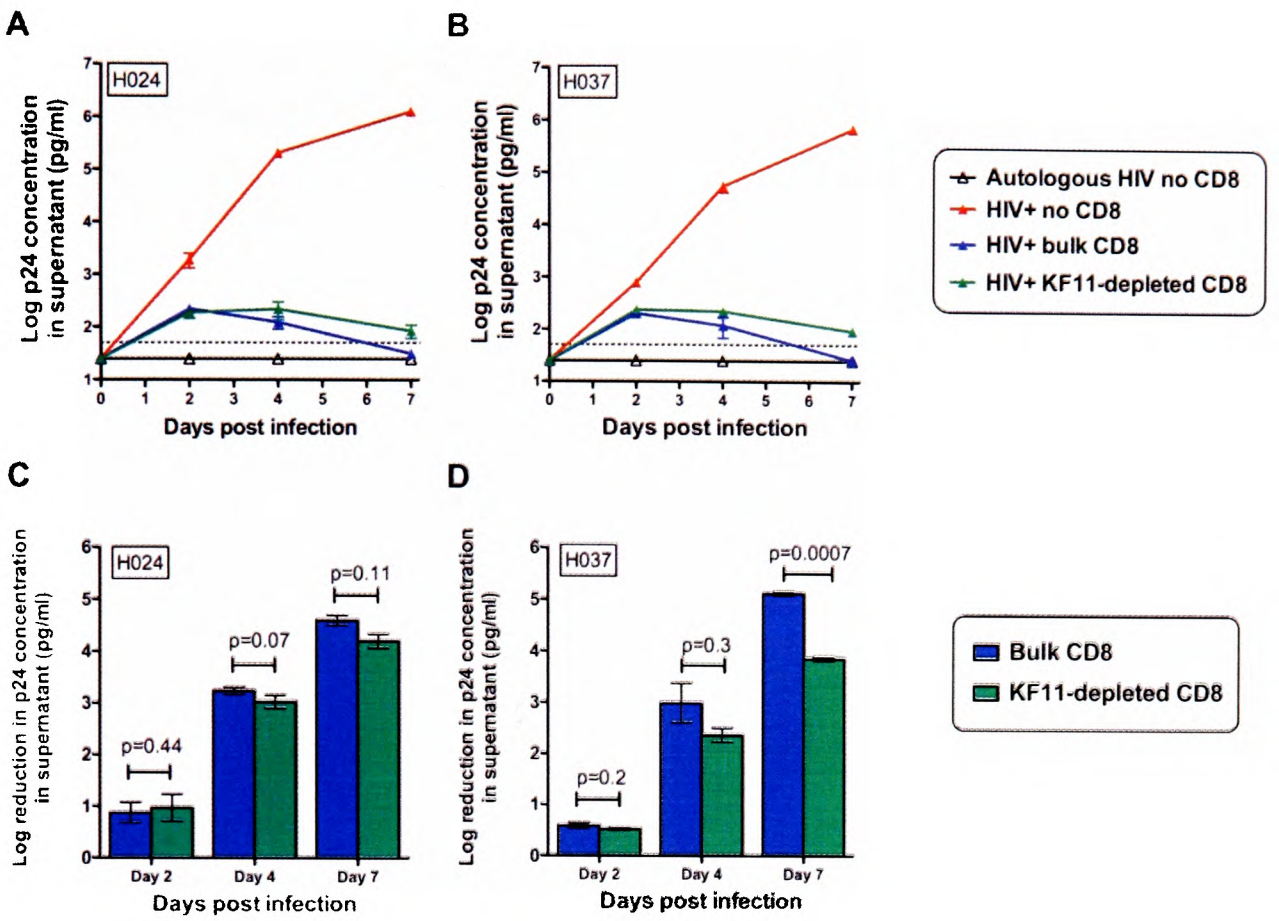


Figure 7.21: Viral inhibition over seven days in culture in the presence and absence of the HLA-B*57-KF11 response in two HIV-infected subjects. Mean ± SEM of three replicates is shown in each case. E:T ratio 1:1 throughout. Panels A and B show viral growth in the presence and absence of HLA-B*57-KF11 specific cells in subjects H024 and H037 respectively. Dotted line shows mean lower threshold of p24 detection. Panels C and D show log reduction in p24 concentration at three time points in H024 and H037 respectively, p values by paired t-test.

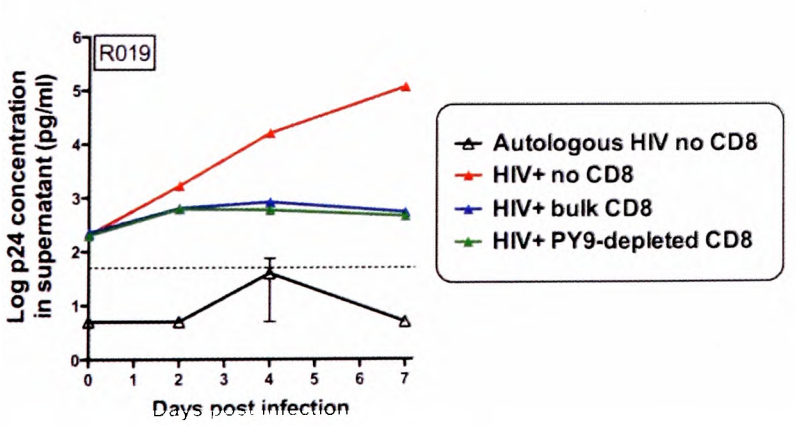


Figure 7.22: Viral inhibition over seven days in culture in the presence and absence of the HLA-B*35-PY9 response in an HIV-infected subject. Mean ± SEM of four replicates is shown. Dotted line shows mean lower threshold of p24 detection.

7.3.8 Subtraction assays: viral escape approach

Due to technical difficulties encountered in running these tetramer-depletion experiments (addressed further in section 7.4.4), further subtraction experiments were performed by superinfecting target cells with a virus bearing two escape mutations in the HLA-B*57 KF11 epitope (A163N and S165G).

In these assays, the suppression mediated by $CD8^+$ T cell responses to the KF11 epitope was assessed by comparing suppression in the presence of the wild-type (WT) epitope with an escape mutant (MT) variant of the epitope. These experiments were carried out in the five subjects detailed in table 7.3; sample data are shown for subject H024 (Fig. 7.23).

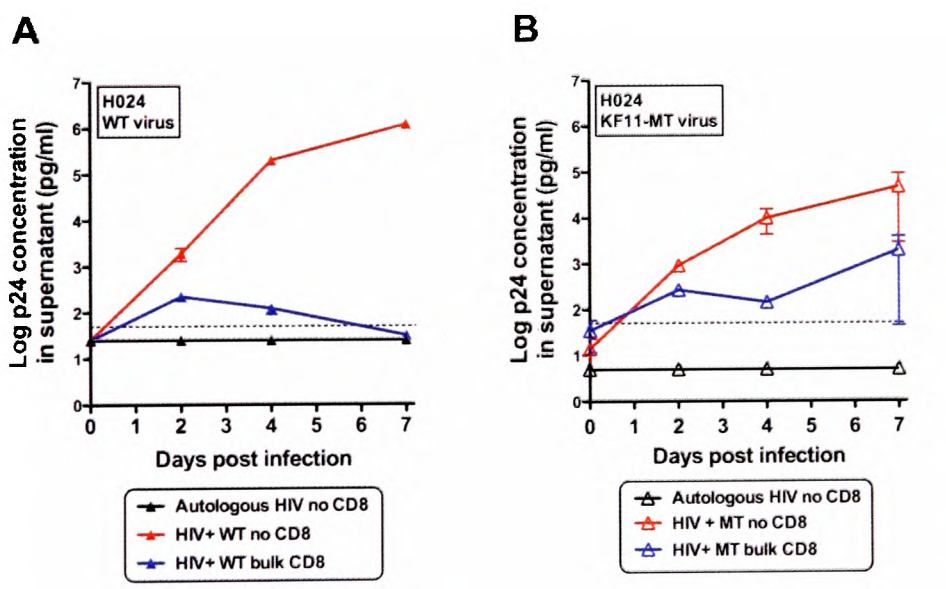


Figure 7.23: Viral inhibition over seven days in culture by bulk $CD8^+$ T cells in the presence of wild-type (WT) virus (A) and HLA-B*5703 KF-11 mutant (MT) virus (B). Mean ± SEM for 3 replicates shown in each case. Dotted line shows mean lower threshold of p24 detection.

These data suggest better suppression in the presence of the WT virus. However, direct comparison is difficult due to the different growth curves of the WT and MT viruses. Therefore, rather than using absolute values, suppression data were used to

calculate the difference between viral growth in the presence and absence of bulk $CD8^+$ T cells. These values are shown below for five suppression experiments in HLA-B*5703 subjects, each carried out in triplicate (Fig. 7.24).

Initial data again demonstrate poorer growth of the MT virus compared to WT, and demonstrate suppression of viral growth in the presence of $CD8^+$ T cell effector cells both in the presence of WT and MT viral strains (Fig. 7.24A). At days 2, 4 and 7 post-infection, there was a consistent trend towards improved suppression in the context of WT infection, but this was only statistically significant at day 2 (Fig. 7.24B). Comparing the natural logs of the slopes, a reduction in growth was only statistically significant in the presence of WT virus (Fig. 7.24C).

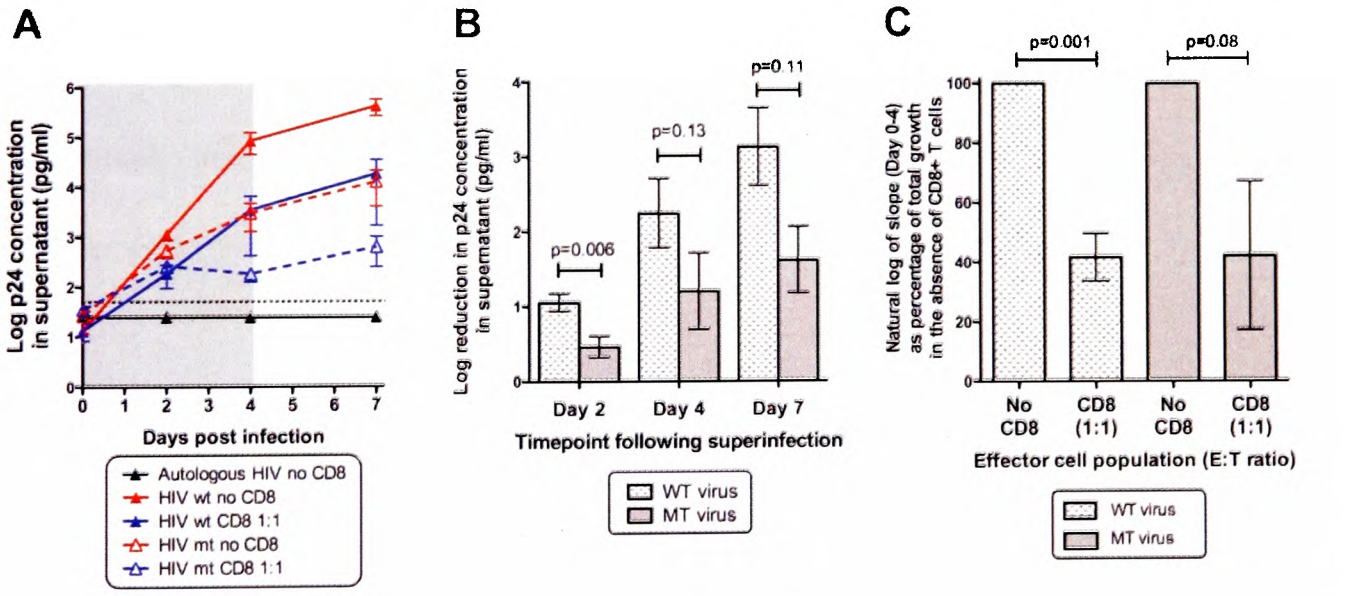


Figure 7.24: Results of viral suppression assays in five HIV-infected subjects with an HLA-B*57 restricted KF11 response showing the reduction in viral suppression associated with the introduction of Gag mutations A163G and S165N. A: Pooled results of viral growth in all 5 subjects showing mean ± SEM. Exponential growth phase marked with grey panel between days 0-4. B: Log reduction in p24 in supernatant; figure plotted is difference between viral growth (either WT or MT) in the presence/absence of $CD8^+$ T cells. p values by paired t-test. C: Natural log of the slope between days 0-4, normalising viral growth to 100% in the absence of $CD8^+$ T cell population. p values by paired t-test.

One possible reason for intra-patient variability, and lack of consistent effect of loss of the KF11 epitope, is the presence of other Gag responses (ISW9 and TW10). The study numbers here are too small for meaningful assessment, but there is a probable role for other HLA-B*57-restricted Gag responses in maintaining viraemic control even after loss of the KF11 response.

7.4 DISCUSSION

7.4.1 Summary of VSA as an *in vitro* approach to quantifying $CD8^+$ T cell responses

This chapter focuses on the development and optimisation of *in vitro* assays that can contribute to dissecting the components of the $CD8^+$ T cell response to HIV. Unlike previous studies, these experiments have the benefit of being performed *ex vivo*, are controlled by the use of autologous target and effector cells, and are strengthened by an approach using bulk $CD8^+$ T cell populations followed by 'subtraction' assays to allow the contribution of an individual response to be precisely defined.

7.4.2 Conclusions from assays using bulk $CD8^+$ T cell populations

Overall, significant suppression can be measured after two days in culture using a bulk $CD8^+$ T cell population at an E:T ratio of 1:1, with progressive suppression up to day 7. The extent of viral suppression at 7 days was statistically associated with the clade of autologous infection, with improved suppression observed in B-clade infected subjects. This is likely to reflect the priming of $CD8^+$ T cell responses in autologous infection that more closely reflect the wild-type sequence of the NL4-3 superinfecting strain, allowing superior control.

There was also a borderline statistical correlation between the presence of HLA-B*57 and improved *in vitro* suppression. A more strongly significant result was originally anticipated. However, the majority of subjects with HLA-B*57 were C-clade infected, thus giving them a potential disadvantage in control of the B-clade lab strain virus used in these assays. Furthermore, due to selective recruiting of healthy (HAART-naïve) subjects, this study cohort is also artificially enriched for subjects with favourable alleles. Thus, HLA-B*5801, -B*27, -B*51 -B*8101 and -B*13 are all represented in the HLA-B*57-negative group but may also exert potent control of viraemia through HLA-B restricted $CD8^+$ T cell epitopes.

Surprisingly, there was lack of correlation of viral suppression with VL, CD4 count, or breadth of Gag ELISpot responses. The lack of any statistical relationship between the results of viral suppression assays and these predictors of improved *in vivo* outcomes is disappointing, suggesting that caution must be applied to the use of these assays in inferring clinical outcome. The reasons for lack of correlation are likely to be multifactorial, and include technical considerations (e.g. cell enrichment, use of frozen cells, quality control of media), as well as intra-subject variability (e.g. HLA type, stage of disease, intercurrent infection).

7.4.3 Conclusions from 'subtraction' assays

The 'subtraction' approach aims to remove an epitope-specific $CD8^+$ T cell response, either by selectively depleting the $CD8^+$ T cell population, or by introducing viral mutants that abrogate HLA-class I presentation of the epitope. The data presented

here suggest that both of these methods can be successfully employed to help to dissect the constituents of total $CD8^+$ mediated viral suppression.

Although the number of subjects studied here is small, there is statistical evidence pointing to some reduction in viraemic control when responses to the HLA-B*5703 KF11 epitope are abrogated. Achieving more statistically significant results may be limited in these small cohorts by intra-patient variability, including the presence of responses to other HLA-B*57 epitopes that may allow maintenance of viraemic control even after loss of the KF11 response. These findings highlight that control of viraemia frequently does not hinge on a single $CD8^+$ T cell response, but that favourable alleles are likely to operate through the recognition of multiple HIV epitopes.

7.4.4 Caveats to the use of 'subtraction' approaches

Technical difficulties and limitations

Particularly in experiments using the tetramer-depletion approach, substantial technical difficulties were encountered in refining these assays. The enrichment processes used here required a minimum of 30 x 10^6 PBMCs to yield enough cells to run each experiment in triplicate *ex vivo*, limiting the numbers of subjects from whom the assays could be successfully undertaken. Several attempts to enrich and deplete cell populations from smaller cell numbers, or from frozen PBMCs, failed due to loss of cells during the enrichment process. In particular, approaches that depend on running cells twice through a magnetic column routinely led to significant cell death when enriching cells from frozen. The success of tetramer-depletion was variable, as – despite short incubation times at 4°C – there is a risk of tetramer-internalisation

following surface binding, such that the tetramer-labelled cells do not subsequently become attached to an anti-PE magnetic bead for depletion in the column.

As well as being limited by the available volume of blood, meaningful comparison between subjects is made difficult by the low prevalence of certain responses (including, in these experiments, HLA-B*35-restricted responses to the Gag PY9 epitope). For this reason, the studies focused largely on the HLA-B*57 KF11 response that is highly prevalent among subjects with this allele.

Outgrowth of autologous virus

Outgrowth of autologous virus is a potentially significant limitation of these assays, particularly (but not exclusively) in subjects with higher viral loads (>10^4 RNA copies / ml plasma). This alters the dynamics of the assay by altering the infecting inoculum and by broadening the repertoire of epitopes being presented by surface MHC molecules on target cells. Although these factors remain incompletely controlled, it is nevertheless reassuring that the presence of detectable autologous virus did not alter the peak p24 concentration, nor did it predict better or worse viraemic suppression by $CD8^+$ T cells in culture.

7.4.5 Further optimisation of VSA and future work

It is doubtless the case that these assays would be optimised by routine use of larger volumes (≥50ml) of fresh blood from HIV-infected study subjects. At present, the use of a B-clade superinfecting virus is a bias towards better suppression in subjects whose autologous infection is with B-clade HIV, as these individuals will have $CD8^+$ T cell responses primed towards B-clade epitopes. Therefore, the specificity of the

approach could be improved by the use of a C-clade laboratory strain virus in order to match the super-infecting strain with the autologous virus as closely as possible.

One potential approach to mitigate against the outgrowth of autologous virus is to add an anti-retroviral agent, such as tenofovir*, to the culture media to suppress autologous virus for the duration of the assay and to use a drug-resistant strain with which to superinfect. However, attempts at this approach by our collaborators have also proved subject to technical difficulties and have not yielded successful results to date (B. Julg, *personal communication*).

The ultimate aim of these experiments would be to use serial depletion of different $CD8^+$ T cell responses, either by using a panel of Class I tetramers or by generating multiple viral mutants, to assess the relative contribution of each response to the overall magnitude of suppression.

7.4.6 Summary comments

These data demonstrate the use of a potentially useful *in vitro* technique of measuring viral suppression, and suggest strategies for dissecting out the importance of different components of the overall $CD8^+$ T cell response. The reasons for lack of correlation of data from these viral suppression assays with markers of improved disease control are uncertain, and suggests this technique may not meet the remit of predicting clinical outcome. However, given the small study numbers, and enrichment for favourable alleles in this cohort, further studies are warranted.

* A nucleotide analogue reverse transcriptase inhibitor

CHAPTER 8: Discussion

8.1 What do these studies add?

The work presented in this thesis aims to study the interplay between host HLA class I expression, and HIV disease progression and evolution, both in individual patients and across large cohorts. Sequence data derived for these studies focused largely on Southern African subjects, given the predominance of C-clade viruses in the global epidemic and the relative under-representation of these cohorts in published studies. Previous studies have highlighted the impact of certain alleles, such as HLA-B*57 (Goulder *et al.* 1996; Migueles *et al.* 2000; Crawford *et al.* 2007; Fellay *et al.* 2007), HLA-B*27 (Goulder, Phillips *et al.* 1997; Almeida *et al.* 2007; Schneidewind *et al.* 2007) and HLA-B*51 (Tomiyama *et al.* 1999; Frater *et al.* 2007; Kawashima *et al.* 2009) in mediating successful viraemic control. The starting point for this work was to develop further understanding of the mechanisms underpinning this natural immune containment.

8.1.1 Relationship between reverting mutations and viraemic control

The data presented in Chapter 3 contribute to understanding how favourable alleles operate, by demonstrating the selection of multiple escape variants that are associated with a cost to viral fitness. These results highlight the delicate balance that exists between viral evasion of the $CD8^+$ T cell response, on the one hand, and fitness constraints caused by selection of escape mutations on the other hand. Paradoxically, the HLA-selection of escape mutations may not lead to uncontrolled viral replication,

but may actually contribute to viraemic suppression due to detrimental effects on viral infectivity or replicative capacity.

These studies also describe the refinement of an algorithm for the identification of HLA-selected mutations, developed in close collaboration with Microsoft (Matthews *et al.* 2008). Despite certain limitations of this approach, it remains a widely used strategy for gene- or genome-wide searches for HLA associations (Brumme, Tao *et al.* 2008; Carlson *et al.* 2008; Rousseau *et al.* 2008).

8.1.2 Influence of HLA on viral evolution within and between clades

The studies described in Chapter 4 (Matthews *et al.* 2009) aimed to describe the influence of HLA Class I on HIV evolution. The analysis demonstrated that the footprints of HLA-selection on the virus may underpin differences between clades, operating in conjunction with founder effect in the establishment of viral subtypes (Matthews *et al.* 2009). Although non-synonymous differences at the nucleotide level undoubtedly determine clade-specificity (Korber *et al.* 2001), founder effect is not independent of HLA-selection: for example, in AE CRF strains there is evidence of an ancestral HLA-B*57 footprint. The potency of a single HLA allele in driving convergent evolution was shown to relate both to the frequency of the selecting allele and number of mutations selected (Matthews *et al.* 2009).

Chapter 5 builds further on the hypothesis that HLA is a significant mediator of viral evolution by investigating a multi-centre cohort for evidence of HIV adaptation to HLA. The study confirmed a significant global, cross-clade impact of HLA in driving HIV sequence changes, in keeping with previous work that has suggested the

potential for epitope loss in the setting of high prevalence of the restricting allele (Leslie *et al.* 2005; Frahm *et al.* 2006). Preliminary evidence is also presented for changing disease outcomes in association with adaptation to HLA-B*51 in Japan, suggesting that alleles associated with good control of viraemia may lose their favourable effect over time (Kawashima *et al.* 2009).

An additional cohort of Zimbabwean Gag sequences was used to investigate phylogeny, sequence adaptation and *ex vivo* $CD8^+$ T cell responses in a population with an increased phenotypic frequency of HLA-B*57. Gag sequence data from this cohort corroborate the earlier findings of HLA-B*57 adaptation (Kawashima *et al.* 2009). Insufficient VL data were available for Zimbabweans with HLA-B*57 to determine whether (as seen for HLA-B*51 in Japan) a possible loss of immune control arises in parallel with the accumulation of escape mutations in the population. However, as ELISpot data suggest that immunodominant HLA-B*57-restricted responses are currently preserved, longitudinal follow-up of the Zimbabwean population would be valuable to identify evolving changes in viraemic control.

8.1.3 Differences between viraemic control in association with HLA-B*35 in different populations

HLA-B*35 was selected for further study based on the high frequency of associated HIV polymorphisms (Chapter 3, (Matthews *et al.* 2008)), evidence of HIV adaptation to HLA-B*35 in the Gag PY9 epitope (Chapter 5, (Kawashima *et al.* 2009)), and the hypothesis that viraemic control in association with this allele might vary between populations. In previous studies of B-clade infected cohorts, HLA-B*35 has widely been associated with poor control of viraemia (Shiga *et al.* 1996; Carrington *et al.*

1999; O'Brien *et al.* 2001; Bansal *et al.* 2007). Conversely, the data presented in Chapter 6 demonstrate that subjects with HLA-B*35 infected with C-clade HIV have viral loads equivalent to, or lower than, the population median. The mechanism for this inter-clade difference appears – at least in part – to be attributable to the $CD8^+$ T cell response to the PY9 epitope. These data thus demonstrate that the effects of an allele on disease control may hinge on the availability of key epitopes (particularly in Gag), rather than on inherent properties of the restricting allele.

8.1.4 Use of *in vitro* assays to quantify viraemic suppression in association with bulk and individual $CD8^+$ T cell responses

The lack of correlation between ELISpot assays and disease control has been reported in HIV and SIV studies, and highlights a need for development of alternative *in vitro* methods for measuring the efficacy of the $CD8^+$ T cell response (Walker and Burton 2008). In Chapter 7, the optimization of viral suppression assays was considered, and 'subtraction' methods were developed for dissecting the overall $CD8^+$ T cell response, whereby a single epitope-specific response can be removed in order to quantify the impact of this individual response to total viraemic control. Such methods are pertinent to future vaccine trials, in which assays that reflect disease outcome are likely to be critical in determining vaccine responses (Barouch 2008; Watkins *et al.* 2008).

8.2 Related studies currently in progress

Following on from these studies, I am currently working on additional projects that are not presented in this thesis but will be the subject of ongoing investigation.

8.2.1 The effect of HLA haplotypes on HIV viraemic control

From the analysis in Chapter 3, HLA-A*74 emerged as an allele associated with selection of escape mutations in Gag, and one that is characteristically associated with low viral loads. In the Durban cohort, HLA-A*74 is transmitted in linkage disequilibrium with HLA-B*5703, and the viraemic control of subjects with HLA-B*5703 is stratified by the presence or absence of HLA-A*74, suggesting a beneficial effect of this haplotype. Computational analysis for other alleles that operate synergistically is currently in progress, and work is being undertaken to identify the mechanism by which HLA-A*74 and HLA-B*5703 operate in tandem to lower viraemia.

8.2.2 Study of diversity in Southern African *nef* genes

In the course of the work presented in Chapters 3-5, a large pool of sequence data for the *nef* gene has been generated for subjects from South Africa, Botswana and Zimbabwe. Multiple copies of Nef are produced early in the course of HIV infection (Shugars *et al.* 1993; Tolstrup *et al.* 2006), making it an abundant source of epitopes for HLA Class I presentation (Lichterfeld *et al.* 2004; Wang *et al.* 2009); Nef is also crucial to the pathogenic functioning of HIV, making it a potentially attractive vaccine target. Previous studies of Nef diversity in C-clade infection have been limited by small sample numbers (van Harmelen *et al.* 2001; Bredell *et al.* 2007). Additional studies therefore aim to identify $CD8^+$ T cell epitopes and escape

mutations, characterise the significance of insertions and deletions, and determine the phylogeny of Nef sequences in larger African cohorts.

8.2.3 *In vitro* comparison of HLA-B*35-restricted $CD8^+$ T cell responses

MHC Class I tetramers for HLA-B*35 epitopes will be used to enrich or deplete specific HLA-B*35 responses using MACS technology, for use in viral suppression assays. Such experiments will allow further development of the observation that viraemic control in these subjects hinges on a Gag response to the PY9 epitope.

8.2.4 Epitope presentation and recognition in acute infection

Methods used by Sacha *et al.* to demonstrate the early presentation and recognition of Gag epitopes in SIV (Sacha, Chung, Rakasz *et al.* 2007; Sacha, Chung, Reed *et al.* 2007) are currently being applied to models of HIV infection *in vitro*; staining for cytokines, chemokines and degranulation markers at intervals over a 24 hour period is used to define the timing of recognition of HIV epitopes by different $CD8^+$ T cell lines.

8.3 Questions needing further investigation

8.3.1 What are the correlates of immune protection?

To date, the precise mechanisms of immune control remain incompletely characterised. Although the importance of HLA Class I restricted $CD8^+$ T cell responses is well documented, there is considerable variation in disease outcome among subjects with 'good' HLA alleles, highlighting the role of other factors in

contributing to overall disease outcome. Conflicting or inconclusive results have arisen from studies that have aimed to determine *in vitro* correlates of effective $CD8^-$ T cell responses, including polyfunctionality (Betts *et al.* 2006; Almeida *et al.* 2007; Streeck *et al.* 2008), avidity (Yang, Sarkis, Trocha *et al.* 2003; Almeida *et al.* 2007; Bennett *et al.* 2008), lytic capacity / killing efficiency (Rollman *et al.* 2007; Ueno *et al.* 2007; Migueles *et al.* 2008) and proliferative capacity (Migueles *et al.* 2002; Day *et al.* 2007). Characterising the specific attributes of a successful $CD8^+$ T cell response remains a critical goal for future studies.

8.3.2 How significant is HLA in driving HIV evolution in the long term?

Although the work presented here provides evidence for the role of $CD8^+$ T cell-mediated selection pressure in driving HIV evolution, HLA is only one of many factors that are pertinent to driving viral evolution. In particular, in the era of HAART, the selection mediated by anti-retroviral agents is potentially dominant in driving sequence evolution (Martinez-Picado and Martinez 2008). To address more completely the impact of HLA-mediated selection on disease control, detailed longitudinal studies in diverse study populations are required.

8.3.3 Why are certain HLA alleles favourable or unfavourable in effecting viraemic control?

Alleles associated with good immune control generally restrict epitopes that lie in conserved regions of the genome, such that selection of escape variants either cannot occur, or occurs only at a detriment to viral fitness (Fernandez *et al.* 2005; Martinez-Picado *et al.* 2006; Brockman *et al.* 2007; Crawford *et al.* 2007; Liu *et al.* 2007;

Boutwell *et al.* 2009). Responses to Gag epitopes provide the immune system with a target that is available within hours of acute infection (Sacha, Chung, Rakasz *et al.* 2007), thus allowing killing of infected cells before the release of progeny virions.

Determinants of immune control may also lie in the HLA peptide-binding groove of individual Class I molecules. HLA-B*57/5801, and the related Mamu-B*17, preferentially bind tryptophan (W) at the C-terminal position (Marsh *et al.* 2000; Honeyborne *et al.* 2006). This residue is encoded only by a single codon, TGG. Escape mutations encoded by single nucleotide changes to TGA or TAG result in stop codons, while alternative nucleotide changes lead to non-conservative amino acid substitutions. Thus any given change to the TGG codon may render the protein non-viable. As tryptophan is one of the most conserved amino acids in the HIV proteome (Trachtenberg *et al.* 2003), selection of escape mutations at this position may be functionally difficult for the virus. Another potential benefit of the tryptophan-preference is the multiple van der Waals contacts that come into play to bind this bulky residue, enhancing the stability of the peptide-MHC complex (Stewart-Jones, Gillespie *et al.* 2005).

The effect of alleles that are characteristically associated with poor immune control of HIV can be considered according to two possible mechanisms. First, the allele restricts no $CD8^+$ T cell epitopes in functionally important regions of the protein, providing no significant contribution to killing of infected cells. These alleles can be regarded as 'null', effectively rendering the individual homozygous at this locus. Secondly, an allele associated with poor outcomes may be considered as actively detrimental to viraemic control (Jin *et al.* 2002). These alleles potentially operate via a

dominant response that inhibits other subdominant responses, as exemplified by the dominant Env response restricted by HLA-B*5802 (Ngumbela *et al.* 2008).

8.4 The future of vaccine design

8.4.1 Is a preventative vaccine possible?

The results of many studies, and observation of natural infection, suggest that – to varying extents - immunological amelioration of HIV is possible (Berkley and Koff 2007). The potential for antibody-mediated neutralization, success of live-attenuated vaccines in monkeys, and populations of elite controllers with HIV infection all give rise to optimism (Walker and Burton 2008). However, recent reviews by experts in the field have questioned whether a vaccine is currently possible (Steinbrook 2007; Walker and Burton 2008). In particular, the failure of the Merck STEP trial highlights the challenges of effecting protective immunity using a T cell vaccine (McElrath *et al.* 2008). Further doubts about the potential use of $CD8^+$-induced immunity arise from a study that demonstrated vaccine-mediated induction of immunodominant $CD8^+$ responses in a subject with HLA-B*27. Following infection in this individual, immune escape and progression to AIDS followed swiftly (Betts *et al.* 2005), highlighting that even responses associated with immune control following natural infection cannot be reliably correlated with vaccine-induced protection.

Induction of sterilising immunity using passive transfer of broadly neutralizing antibodies has been demonstrated in animal models (Baba *et al.* 2000; Mascola *et al.* 2000; Mascola 2002), but no vaccine strategy has managed to elicit such an antibody response (Burton *et al.* 2004; Watkins *et al.* 2008). SIV-macaque studies have demonstrated that control of viraemia is also possible in the absence of an antibody

response (Wilson *et al.* 2006), but it is now generally agreed that a vaccine has the best chance of success using a combination of antibody- and cell-mediated immunity (Watkins *et al.* 2008), coordinated by a $CD4^+$ T cell response (Walker and Burton 2008), and that a realistic initial goal is reduction of viraemia rather than sterilising immunity (Barouch 2008).

With respect to the development of a $CD8^+$ T cell vaccine, several key challenges have emerged:

(i) Eliciting appropriate breadth of responses: one of the reasons cited for the failure of the Merck vaccine was the lack of quality and quantity of $CD8^+$ T cell responses in vaccinees (Barouch 2008; Fauci *et al.* 2008; Plotkin 2009). Using this as a benchmark, any future candidate vaccine would need to produce an increased breadth and/or magnitude of responses (McElrath *et al.* 2008; Watkins *et al.* 2008), both to maximize immune responses and to reduce escape options for the virus (Barouch 2008). Reliable stimulation of immunodominant responses to Gag peptides is also likely to be a crucial feature of a successful vaccine (Edwards *et al.* 2002; Kiepiela *et al.* 2007; Sacha, Chung, Rakasz *et al.* 2007).

(ii) Identifying an appropriate vector: using an Ad5 vector in macaques, only animals expressing the favourable Mamu-A*01 allele were significantly protected (Wilson *et al.* 2006), suggesting limited protection following this approach. In the STEP trial, also based on a homologous Ad5 prime/boost regimen, prior Ad5 immunity may have directly contributed to vaccine failure (Watkins *et al.* 2008). Subsequent SIV studies have demonstrated superiority of an Ad26 prime/Ad5 boost

regimen (Liu *et al.* 2009), suggesting that further exploration of heterologous prime/boost strategies is warranted (Barouch 2008).

(iii) Tackling reservoirs of latency: a vaccine-induced $CD8^+$ T cell response may not be able to expand quickly enough to protect against establishment of infection (Barouch 2008), given the early establishment of latent viral reservoirs following acute infection (Chun *et al.* 1998; Alexaki *et al.* 2008; Groot *et al.* 2008). Strategies that promote development of mucosal immunity may be required to protect against these critical steps in acute pathophysiology. This hypothesis is supported by the success of vaccine strategies against other pathogens, which rely on the induction of NAbs as 'immunological gatekeepers' (Walker and Burton 2008).

(iv) Overcoming viral diversity and clade-specificity: Even single amino acid substitutions can lead to immune escape (Yang, Sarkis, Ali *et al.* 2003). Sequence mismatch between the vaccine strain and an infecting virion, allowing viral escape from vaccine-induced responses, is therefore a significant concern given the overall diversity of circulating viruses (Gaschen *et al.* 2002; Brander *et al.* 2006; Barouch 2008; Bennett *et al.* 2008; Watkins *et al.* 2008).

8.4.2 What modifications might be required for a successful vaccine?

$CD8^+$ T cell vaccines

'Mosaic' vaccine antigens that incorporate wild type and escape epitopes (Plotkin 2009) or multi-epitope constructs to alter natural immunodominance patterns to target optimum epitopes have been suggested (Brander *et al.* 2006; Watkins *et al.* 2008). The use of new vectors is being explored, including Adenovirus serotypes that are

rare in the human population (Barouch 2008; Liu *et al.* 2009), as well as CMV, measles and vaccinia (Plotkin 2009). DNA priming prior to Adenovirus-vector boost has been associated with improved outcomes in some SIV trials (Casimiro *et al.* 2005; Barouch 2008), but outcomes in human infection have yet to be assessed.

Other vaccine strategies

Further research is ongoing into the use of vaccine adjuvants to harness other immune responses, e.g. toll-like receptor agonists that activate DCs and B-cells (Wille-Reece *et al.* 2006). The development of other novel approaches to stimulating vaccine-induced NAbs is also crucial, and for this, a greater understanding of the structure of Env trimers is required (Walker and Burton 2008).

8.5 Concluding remarks

The work presented here aimed to further an understanding of the role and impact of $CD8^+$ T cell responses both in control of HIV infection and in directing the future course of the epidemic. The basic science questions addressed are motivated by the pressing need for improved immunological insights into the control of HIV to inform future vaccine strategies, with particular reference to the devastating epidemic in Southern Africa. Sustained, collaborative efforts by scientists and clinicians over the coming years will be required to halt this sophisticated and rapidly evolving pathogen.



CHAPTER 8: DISCUSSION

'Science appears calm and triumphant when it is completed;

but science in the process of being done is only

contradiction and torment, hope and disappointment.'

P. Émile Roux, 1853-1933

Pasteur Institute, Paris

BIBLIOGRAPHY

Abraha, A., Nankya, I. L., Gibson, R., Demers, K., Tebit, D. M., Johnston, E., Katzenstein, D., Siddiqui, A., Herrera, C., Fischetti, L., Shattock, R. J. and Arts, E. J. (2009). "CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic." J Virol **83**(11): 5592-605.

Addo, M. M., Yu, X. G., Rathod, A., Cohen, D., Eldridge, R. L., Strick, D., Johnston, M. N., Corcoran, C., Wurcel, A. G., Fitzpatrick, C. A., Feeney, M. E., Rodriguez, W. R., Basgoz, N., Draenert, R., Stone, D. R., Brander, C., Goulder, P. J., Rosenberg, E. S., Altfeld, M. and Walker, B. D. (2003). "Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load." J Virol **77**(3): 2081-92.

Ako-Adjei, D., Johnson, M. C. and Vogt, V. M. (2005). "The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles." J Virol **79**(21): 13463-72.

Alexaki, A., Liu, Y. and Wigdahl, B. (2008). "Cellular reservoirs of HIV-1 and their role in viral persistence." Curr HIV Res **6**(5): 388-400.

Ali, A., Ng, H. L., Dagarag, M. D. and Yang, O. O. (2005). "Evasion of cytotoxic T lymphocytes is a functional constraint maintaining HIV-1 Nef expression." Eur J Immunol **35**(11): 3221-8.

Allen, T. M., Altfeld, M., Geer, S. C., Kalife, E. T., Moore, C., O'Sullivan K, M., Desouza, I., Feeney, M. E., Eldridge, R. L., Maier, E. L., Kaufmann, D. E., Lahaie, M. P., Reyor, L., Tanzi, G., Johnston, M. N., Brander, C., Draenert, R., Rockstroh, J. K., Jessen, H., Rosenberg, E. S., Mallal, S. A. and Walker, B. D. (2005). "Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution." J Virol **79**(21): 13239-49.

Allen, T. M., Altfeld, M., Yu, X. G., O'Sullivan, K. M., Lichterfeld, M., Le Gall, S., John, M., Mothe, B. R., Lee, P. K., Kalife, E. T., Cohen, D. E., Freedberg, K. A., Strick, D. A., Johnston, M. N., Sette, A., Rosenberg, E. S., Mallal, S. A., Goulder, P. J., Brander, C. and Walker, B. D. (2004). "Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection." J Virol **78**(13): 7069-78.

Almeida, J. R., Price, D. A., Papagno, L., Arkoub, Z. A., Sauce, D., Bornstein, E., Asher, T. E., Samri, A., Schnuriger, A., Theodorou, I., Costagliola, D., Rouzioux, C., Agut, H., Marcelin, A. G., Douek, D., Autran, B. and Appay, V. (2007). "Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover." J Exp Med **204**(10): 2473-85.

Alter, G. and Altfeld, M. (2009). "NK cells in HIV-1 infection: evidence for their role in the control of HIV-1 infection." J Intern Med **265**(1): 29-42.

Altfeld, M., Addo, M. M., Rosenberg, E. S., Hecht, F. M., Lee, P. K., Vogel, M., Yu, X. G., Draenert, R., Johnston, M. N., Strick, D., Allen, T. M., Feeney, M. E., Kahn, J. O., Sekaly, R. P., Levy, J. A., Rockstroh, J. K., Goulder, P. J. and Walker, B. D. (2003). "Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection." Aids **17**(18): 2581-91.

Appay, V., Douek, D. C. and Price, D. A. (2008). "CD8+ T cell efficacy in vaccination and disease." Nat Med **14**(6): 623-8.

Appay, V. and Rowland-Jones, S. L. (2002). "The assessment of antigen-specific CD8+ T cells through the combination of MHC class I tetramer and intracellular staining." J Immunol Methods **268**(1): 9-19.

Baba, T. W., Liska, V., Hofmann-Lehmann, R., Vlasak, J., Xu, W., Ayehunie, S., Cavacini, L. A., Posner, M. R., Katinger, H., Stiegler, G., Bernacky, B. J., Rizvi, T. A., Schmidt, R., Hill, L. R., Keeling, M. E., Lu, Y., Wright, J. E., Chou, T. C. and Ruprecht, R. M. (2000). "Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection." Nat Med **6**(2): 200-6.

Baba, T. W., Liska, V., Khimani, A. H., Ray, N. B., Dailey, P. J., Penninck, D., Bronson, R., Greene, M. F., McClure, H. M., Martin, L. N. and Ruprecht, R. M. (1999). "Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques." Nat Med **5**(2): 194-203.

Bansal, A., Yue, L., Conway, J., Yusim, K., Tang, J., Kappes, J., Kaslow, R. A., Wilson, C. M. and Goepfert, P. A. (2007). "Immunological control of chronic HIV-1 infection: HLA-mediated immune function and viral evolution in adolescents." Aids **21**(18): 2387-97.

Barouch, D. H. (2008). "Challenges in the development of an HIV-1 vaccine." Nature **455**(7213): 613-9.

Barouch, D. H., Kunstman, J., Kuroda, M. J., Schmitz, J. E., Santra, S., Peyerl, F. W., Krivulka, G. R., Beaudry, K., Lifton, M. A., Gorgone, D. A., Montefiori, D. C., Lewis, M. G., Wolinsky, S. M. and Letvin, N. L. (2002). "Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes." Nature **415**(6869): 335-9.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 868-71.

Bennett, M. S., Ng, H. L., Ali, A. and Yang, O. O. (2008). "Cross-clade detection of HIV-1-specific cytotoxic T lymphocytes does not reflect cross-clade antiviral activity." J Infect Dis **197**(3): 390-7.

Benson, J., Chougnet, C., Robert-Guroff, M., Montefiori, D., Markham, P., Shearer, G., Gallo, R. C., Cranage, M., Paoletti, E., Limbach, K., Venzon, D., Tartaglia, J. and Franchini, G. (1998). "Recombinant vaccine-induced protection against the highly pathogenic simian immunodeficiency virus SIV(mac251): dependence on route of challenge exposure." J Virol **72**(5): 4170-82.

Berkley, S. F. and Koff, W. C. (2007). "Scientific and policy challenges to development of an AIDS vaccine." Lancet **370**(9581): 94-101.

Berman, P. W., Murthy, K. K., Wrin, T., Vennari, J. C., Cobb, E. K., Eastman, D. J., Champe, M., Nakamura, G. R., Davison, D., Powell, M. F., Bussiere, J., Francis, D. P., Matthews, T., Gregory, T. J. and Obijeski, J. F. (1996). "Protection of MN-rgp120-immunized chimpanzees from heterologous infection with a primary isolate of human immunodeficiency virus type 1." J Infect Dis **173**(1): 52-9.

Betts, M. R., Ambrozak, D. R., Douek, D. C., Bonhoeffer, S., Brenchley, J. M., Casazza, J. P., Koup, R. A. and Picker, L. J. (2001). "Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection." J Virol **75**(24): 11983-91.

Betts, M. R., Exley, B., Price, D. A., Bansal, A., Camacho, Z. T., Teaberry, V., West, S. M., Ambrozak, D. R., Tomaras, G., Roederer, M., Kilby, J. M., Tartaglia, J., Belshe, R., Gao, F., Douek, D. C., Weinhold, K. J., Koup, R. A., Goepfert, P. and Ferrari, G. (2005). "Characterization of functional and phenotypic changes in anti-Gag vaccine-induced T cell responses and their role in protection after HIV-1 infection." Proc Natl Acad Sci U S A **102**(12): 4512-7.

Betts, M. R., Nason, M. C., West, S. M., De Rosa, S. C., Migueles, S. A., Abraham, J., Lederman, M. M., Benito, J. M., Goepfert, P. A., Connors, M., Roederer, M. and Koup, R. A. (2006). "HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells." Blood **107**(12): 4781-9.

Bhattacharya, T., Daniels, M., Heckerman, D., Foley, B., Frahm, N., Kadie, C., Carlson, J., Yusim, K., McMahon, B., Gaschen, B., Mallal, S., Mullins, J. I., Nickle, D. C., Herbeck, J., Rousseau, C., Learn, G. H., Miura, T., Brander, C., Walker, B. and Korber, B. (2007). "Founder effects in the assessment of HIV polymorphisms and HLA allele associations." Science **315**(5818): 1583-6.

Bodmer, W. F. (1987). "The HLA system: structure and function." J Clin Pathol **40**(9): 948-58.

Borghans, J. A., Molgaard, A., de Boer, R. J. and Kesmir, C. (2007). "HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24." PLoS ONE **2**(9): e920.

Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. and Oldstone, M. B. (1994). "Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection." J Virol **68**(9): 6103-10.

Boutwell, C. L., Rowley, C. F. and Essex, M. (2009). "Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein." J Virol **83**(6): 2460-8.

Bracho, M. A., Moya, A. and Barrio, E. (1998). "Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity." J Gen Virol **79 (Pt 12)**: 2921-8.

Brander, C., Frahm, N. and Walker, B. D. (2006). "The challenges of host and viral diversity in HIV vaccine design." Curr Opin Immunol **18**(4): 430-7.

Bredell, H., Martin, D. P., Van Harmelen, J., Varsani, A., Sheppard, H. W., Donovan, R., Gray, C. M. and Williamson, C. (2007). "HIV type 1 subtype C gag and nef diversity in Southern Africa." AIDS Res Hum Retroviruses **23**(3): 477-81.

Brenchley, J. M., Hill, B. J., Ambrozak, D. R., Price, D. A., Guenaga, F. J., Casazza, J. P., Kuruppu, J., Yazdani, J., Migueles, S. A., Connors, M., Roederer, M., Douek, D. C. and Koup, R. A. (2004). "T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis." J Virol **78**(3): 1160-8.

Brenchley, J. M., Paiardini, M., Knox, K. S., Asher, A. I., Cervasi, B., Asher, T. E., Scheinberg, P., Price, D. A., Hage, C. A., Kholi, L. M., Khoruts, A., Frank, I., Else, J., Schacker, T., Silvestri, G. and Douek, D. C. (2008). "Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections." Blood **112**(7): 2826-35.

Brenchley, J. M., Price, D. A., Schacker, T. W., Asher, T. E., Silvestri, G., Rao, S., Kazzaz, Z., Bornstein, E., Lambotte, O., Altmann, D., Blazar, B. R., Rodriguez, B., Teixeira-Johnson, L., Landay, A., Martin, J. N., Hecht, F. M., Picker, L. J., Lederman, M. M., Deeks, S. G. and Douek, D. C. (2006). "Microbial translocation is a cause of systemic immune activation in chronic HIV infection." Nat Med **12**(12): 1365-71.

Brennan, R. O. and Durack, D. T. (1981). "Gay compromise syndrome." Lancet **2**(8259): 1338-9.

Brenner, M., Munch, J., Schindler, M., Wildum, S., Stolte, N., Stahl-Hennig, C., Fuchs, D., Matz-Rensing, K., Franz, M., Heeney, J., Ten Haaft, P., Swigut, T., Hrecka, K., Skowronski, J. and Kirchhoff, F. (2006). "Importance of the N-distal AP-2 binding element in Nef for simian immunodeficiency virus replication and pathogenicity in rhesus macaques." J Virol **80**(9): 4469-81.

Briggs, J. A., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M. and Johnson, M. C. (2004). "The stoichiometry of Gag protein in HIV-1." Nat Struct Mol Biol **11**(7): 672-5.

Brockman, M. A., Schneidewind, A., Lahaie, M., Schmidt, A., Miura, T., Desouza, I., Ryvkin, F., Derdeyn, C. A., Allen, S., Hunter, E., Mulenga, J., Goepfert, P. A., Walker, B. D. and Allen, T. M. (2007). "Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A." J Virol **81**(22): 12608-18.

Brumme, Z. L., Brumme, C. J., Carlson, J., Streeck, H., John, M., Eichbaum, Q., Block, B. L., Baker, B., Kadie, C., Markowitz, M., Jessen, H., Kelleher, A. D., Rosenberg, E., Kaldor, J., Yuki, Y., Carrington, M., Allen, T. M., Mallal, S., Altfeld, M., Heckerman, D. and Walker, B. D. (2008). "Marked epitope and allele-specific differences in rates of mutation in HIV-1 Gag, Pol and Nef CTL epitopes in acute/early HIV-1 infection." J Virol.

Brumme, Z. L., Brumme, C. J., Heckerman, D., Korber, B. T., Daniels, M., Carlson, J., Kadie, C., Bhattacharya, T., Chui, C., Szinger, J., Mo, T., Hogg, R. S., Montaner, J. S., Frahm, N., Brander, C., Walker, B. D. and Harrigan, P. R. (2007). "Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1." PLoS Pathog **3**(7): e94.

Brumme, Z. L., Tao, I., Szeto, S., Brumme, C. J., Carlson, J. M., Chan, D., Kadie, C., Frahm, N., Brander, C., Walker, B., Heckerman, D. and Harrigan, P. R. (2008). "Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection." Aids **22**(11): 1277-86.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P. B., Lama, J. R., Marmor, M., Del Rio, C., McElrath, M. J., Casimiro, D. R., Gottesdiener, K. M., Chodakewitz, J. A., Corey, L. and Robertson, M. N. (2008). "Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial." Lancet **372**(9653): 1881-93.

Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P., Nabel, G. J., Sodroski, J., Wilson, I. A. and Wyatt, R. T. (2004). "HIV vaccine design and the neutralizing antibody problem." Nat Immunol **5**(3): 233-6.

Cano, P., Klitz, W., Mack, S. J., Maiers, M., Marsh, S. G., Noreen, H., Reed, E. F., Senitzer, D., Setterholm, M., Smith, A. and Fernandez-Vina, M. (2007). "Common and well-documented HLA alleles: report of the Ad-Hoc committee of the american society for histocompatiblity and immunogenetics." Hum Immunol **68**(5): 392-417.

Carlson, J. M., Brumme, Z. L., Rousseau, C. M., Brumme, C. J., Matthews, P., Kadie, C., Mullins, J. I., Walker, B. D., Harrigan, P. R., Goulder, P. J. and Heckerman, D. (2008). "Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag." PLoS Comput Biol **4**(11): e1000225.

Carrington, M., Martin, M. P. and van Bergen, J. (2008). "KIR-HLA intercourse in HIV disease." Trends Microbiol **16**(12): 620-7.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K. and O'Brien, S. J. (1999). "HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage." Science **283**(5408): 1748-52.

Casimiro, D. R., Wang, F., Schleif, W. A., Liang, X., Zhang, Z. Q., Tobery, T. W., Davies, M. E., McDermott, A. B., O'Connor, D. H., Fridman, A., Bagchi, A., Tussey, L. G., Bett, A. J., Finnefrock, A. C., Fu, T. M., Tang, A., Wilson, K. A., Chen, M., Perry, H. C., Heidecker, G. J., Freed, D. C., Carella, A., Punt, K. S., Sykes, K. J., Huang, L., Ausensi, V. I., Bachinsky, M., Sadasivan-Nair, U., Watkins, D. I., Emini, E. A. and Shiver, J. W. (2005). "Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with dna and recombinant adenoviral vaccine vectors expressing Gag." J Virol **79**(24): 15547-55.

Chandanayingyong, D., Stephens, H. A., Klaythong, R., Sirikong, M., Udee, S., Longta, P., Chantangpol, R., Bejrachandra, S. and Rungruang, E. (1997). "HLA-A, -B, -DRB1, -DQA1, and -DQB1 polymorphism in Thais." Hum Immunol **53**(2): 174-82.

Chen, H., Piechocka-Trocha, A., Miura, T., Brockman, M. A., Julg, B. D., Baker, B. M., Rothchild, A. C., Block, B. L., Schneidewind, A., Koibuchi, T., Pereyra, F., Allen, T. M. and Walker, B. D. (2009). "Differential neutralization of human immunodeficiency virus (HIV) replication in autologous CD4 T cells by HIV-specific cytotoxic T lymphocytes." J Virol **83**(7): 3138-49.

Chen, Z., Luckay, A., Sodora, D. L., Telfer, P., Reed, P., Gettie, A., Kanu, J. M., Sadek, R. F., Yee, J., Ho, D. D., Zhang, L. and Marx, P. A. (1997). "Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys." J Virol **71**(5): 3953-60.

Chomarat, P., Banchereau, J., Davoust, J. and Palucka, A. K. (2000). "IL-6 switches the differentiation of monocytes from dendritic cells to macrophages." Nat Immunol **1**(6): 510-4.

Chopera, D. R., Woodman, Z., Mlisana, K., Mlotshwa, M., Martin, D. P., Seoighe, C., Treurnicht, F., de Rosa, D. A., Hide, W., Karim, S. A., Gray, C. M. and Williamson, C. (2008). "Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage." PLoS Pathog **4**(3): e1000033.

Chun, T. W., Engel, D., Berrey, M. M., Shea, T., Corey, L. and Fauci, A. S. (1998). "Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection." Proc Natl Acad Sci U S A **95**(15): 8869-73.

Chung, C., Lee, W., Loffredo, J. T., Burwitz, B., Friedrich, T. C., Giraldo Vela, J. P., Napoe, G., Rakasz, E. G., Wilson, N. A., Allison, D. B. and Watkins, D. I. (2007). "Not all cytokine-producing CD8+ T cells suppress simian immunodeficiency virus replication." J Virol **81**(3): 1517-23.

Clapham, P. R. and McKnight, A. (2001). "HIV-1 receptors and cell tropism." Br Med Bull **58**: 43-59.

Clark, S. J. and Shaw, G. M. (1993). "The acute retroviral syndrome and the pathogenesis of HIV-1 infection." Semin Immunol **5**(3): 149-55.

Crawford, H., Lumm, W., Leslie, A., Schaefer, M., Boeras, D., Prado, J. G., Tang, J., Farmer, P., Ndung'u, T., Lakhi, S., Gilmour, J., Goepfert, P., Walker, B. D., Kaslow, R., Mulenga, J., Allen, S., Goulder, P. J. and Hunter, E. (2009). "Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients." J Exp Med **206**(4): 909-21.

Crawford, H., Prado, J. G., Leslie, A., Hue, S., Honeyborne, I., Reddy, S., van der Stok, M., Mncube, Z., Brander, C., Rousseau, C., Mullins, J. I., Kaslow, R., Goepfert, P., Allen, S., Hunter, E., Mulenga, J., Kiepiela, P., Walker, B. D. and Goulder, P. J. (2007). "Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection." J Virol **81**(15): 8346-51.

Daar, E. S., Moudgil, T., Meyer, R. D. and Ho, D. D. (1991). "Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection." N Engl J Med **324**(14): 961-4.

Daniel, M. D., Kirchhoff, F., Czajak, S. C., Sehgal, P. K. and Desrosiers, R. C. (1992). "Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene." Science **258**(5090): 1938-41.

Day, C. L., Kaufmann, D. E., Kiepiela, P., Brown, J. A., Moodley, E. S., Reddy, S., Mackey, E. W., Miller, J. D., Leslie, A. J., DePierres, C., Mncube, Z., Duraiswamy, J., Zhu, B., Eichbaum, Q., Altfeld, M., Wherry, E. J., Coovadia, H. M., Goulder, P. J., Klenerman, P., Ahmed, R., Freeman, G. J. and Walker, B. D. (2006). "PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression." Nature **443**(7109): 350-4.

Day, C. L., Kiepiela, P., Leslie, A. J., van der Stok, M., Nair, K., Ismail, N., Honeyborne, I., Crawford, H., Coovadia, H. M., Goulder, P. J., Walker, B. D. and Klenerman, P. (2007). "Proliferative capacity of epitope-specific CD8 T-cell responses is inversely related to viral load in chronic human immunodeficiency virus type 1 infection." J Virol **81**(1): 434-8.

de Oliveira, T., Salemi, M., Gordon, M., Vandamme, A. M., van Rensburg, E. J., Engelbrecht, S., Coovadia, H. M. and Cassol, S. (2004). "Mapping sites of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design?" Genetics **167**(3): 1047-58.

de Witte, L., Nabatov, A. and Geijtenbeek, T. B. (2008). "Distinct roles for DC-SIGN+-dendritic cells and Langerhans cells in HIV-1 transmission." Trends Mol Med **14**(1): 12-9.

Delport, W., Scheffler, K. and Seoighe, C. (2008). "Frequent toggling between alternative amino acids is driven by selection in HIV-1." PLoS Pathog **4**(12): e1000242.

Dorrell, L., Willcox, B. E., Jones, E. Y., Gillespie, G., Njai, H., Sabally, S., Jaye, A., DeGleria, K., Rostron, T., Lepin, E., McMichael, A., Whittle, H. and Rowland-Jones, S. (2001). "Cytotoxic T lymphocytes recognize structurally diverse, clade-specific and cross-reactive peptides in human immunodeficiency virus type-1 gag through HLA-B53." Eur J Immunol **31**(6): 1747-56.

Douek, D. (2007). "HIV disease progression: immune activation, microbes, and a leaky gut." Top HIV Med **15**(4): 114-7.

Draenert, R., Le Gall, S., Pfafferott, K. J., Leslie, A. J., Chetty, P., Brander, C., Holmes, E. C., Chang, S. C., Feeney, M. E., Addo, M. M., Ruiz, L., Ramduth, D., Jeena, P., Altfeld, M., Thomas, S., Tang, Y., Verrill, C. L., Dixon, C., Prado, J. G., Kiepiela, P., Martinez-Picado, J., Walker, B. D. and Goulder, P. J. (2004). "Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection." J Exp Med **199**(7): 905-15.

Duda, A., Lee-Turner, L., Fox, J., Robinson, N., Dustan, S., Kaye, S., Fryer, H., Carrington, M., McClure, M., McLean, A. R., Fidler, S., Weber, J., Phillips, R. E. and Frater, A. J. (2009). "HLA-associated clinical progression correlates with epitope reversion rates in early human immunodeficiency virus infection." J Virol **83**(3): 1228-39.

Eckert, K. A. and Kunkel, T. A. (1991). "DNA polymerase fidelity and the polymerase chain reaction." PCR Methods Appl **1**(1): 17-24.

Edwards, B. H., Bansal, A., Sabbaj, S., Bakari, J., Mulligan, M. J. and Goepfert, P. A. (2002). "Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma." J Virol **76**(5): 2298-305.

Emu, B., Sinclair, E., Hatano, H., Ferre, A., Shacklett, B., Martin, J. N., McCune, J. M. and Deeks, S. G. (2008). "HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control." J Virol **82**(11): 5398-407.

Escobar, H., Crockett, D. K., Reyes-Vargas, E., Baena, A., Rockwood, A. L., Jensen, P. E. and Delgado, J. C. (2008). "Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs." J Immunol **181**(7): 4874-82.

Eugenin, E. A., Gaskill, P. J. and Berman, J. W. (2009). "Tunneling nanotubes (TNT) are induced by HIV-infection of macrophages: a potential mechanism for intercellular HIV trafficking." Cell Immunol **254**(2): 142-8.

Fauci, A. S., Johnston, M. I., Dieffenbach, C. W., Burton, D. R., Hammer, S. M., Hoxie, J. A., Martin, M., Overbaugh, J., Watkins, D. I., Mahmoud, A. and Greene, W. C. (2008). "HIV vaccine research: the way forward." Science **321**(5888): 530-2.

Feeney, M. E., Tang, Y., Roosevelt, K. A., Leslie, A. J., McIntosh, K., Karthas, N., Walker, B. D. and Goulder, P. J. (2004). "Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child." J Virol **78**(16): 8927-30.

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., Wyniger, J., Descombes, P., Antonarakis, S. E., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A. and Goldstein, D. B. (2007). "A whole-genome association study of major determinants for host control of HIV-1." Science **317**(5840): 944-7.

Felsenstein, J. (1985). "Confidence limits on phylogenies: an approach using the bootstrap." Evolution **39**(4): 783-791.

Fernandez, C. S., Stratov, I., De Rose, R., Walsh, K., Dale, C. J., Smith, M. Z., Agy, M. B., Hu, S. L., Krebs, K., Watkins, D. I., O'Connor D, H., Davenport, M. P. and Kent, S. J. (2005). "Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost." J Virol **79**(9): 5721-31.

Flores-Villanueva, P. O., Hendel, H., Caillat-Zucman, S., Rappaport, J., Burgos-Tiburcio, A., Bertin-Maghit, S., Ruiz-Morales, J. A., Teran, M. E., Rodriguez-Tafur, J. and Zagury, J. F. (2003). "Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development." J Immunol **170**(4): 1925-9.

Flores-Villanueva, P. O., Yunis, E. J., Delgado, J. C., Vittinghoff, E., Buchbinder, S., Leung, J. Y., Uglialoro, A. M., Clavijo, O. P., Rosenberg, E. S., Kalams, S. A., Braun, J. D., Boswell, S. L., Walker, B. D. and Goldfeld, A. E. (2001). "Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity." Proc Natl Acad Sci U S A **98**(9): 5140-5.

Flynn, N. M., Forthal, D. N., Harro, C. D., Judson, F. N., Mayer, K. H. and Para, M. F. (2005). "Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection." J Infect Dis **191**(5): 654-65.

Foster, J. L. and Garcia, J. V. (2008). "HIV-1 Nef: at the crossroads." Retrovirology **5**: 84.

Frahm, N., Kaufmann, D. E., Yusim, K., Muldoon, M., Kesmir, C., Linde, C. H., Fischer, W., Allen, T. M., Li, B., McMahon, B. H., Faircloth, K. L., Hewitt, H. S., Mackey, E. W., Miura, T., Khatri, A., Wolinsky, S., McMichael, A., Funkhouser, R. K., Walker, B. D., Brander, C. and Korber, B. T. (2007). "Increased sequence diversity coverage improves detection of HIV-specific T cell responses." J Immunol **179**(10): 6638-50.

Frahm, N., Kiepiela, P., Adams, S., Linde, C. H., Hewitt, H. S., Sango, K., Feeney, M. E., Addo, M. M., Lichterfeld, M., Lahaie, M. P., Pae, E., Wurcel, A. G., Roach, T., St John, M. A., Altfeld, M., Marincola, F. M., Moore, C., Mallal, S., Carrington, M., Heckerman, D., Allen, T. M., Mullins, J. I., Korber, B. T., Goulder, P. J., Walker, B. D. and Brander, C. (2006). "Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes." Nat Immunol **7**(2): 173-8.

Frater, A. J., Brown, H., Oxenius, A., Gunthard, H. F., Hirschel, B., Robinson, N., Leslie, A. J., Payne, R., Crawford, H., Prendergast, A., Brander, C., Kiepiela, P., Walker, B. D., Goulder, P. J., McLean, A. and Phillips, R. E. (2007). "Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression." J Virol **81**(12): 6742-51.

Friedrich, T. C., Dodds, E. J., Yant, L. J., Vojnov, L., Rudersdorf, R., Cullen, C., Evans, D. T., Desrosiers, R. C., Mothe, B. R., Sidney, J., Sette, A., Kunstman, K., Wolinsky, S., Piatak, M., Lifson, J., Hughes, A. L., Wilson, N., O'Connor, D. H. and Watkins, D. I. (2004). "Reversion of CTL escape-variant immunodeficiency viruses in vivo." Nat Med **10**(3): 275-81.

Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I. and Hill, C. P. (1996). "Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid." Cell **87**(7): 1285-94.

Gamble, T. R., Yoo, S., Vajdos, F. F., von Schwedler, U. K., Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I. and Hill, C. P. (1997). "Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein." Science **278**(5339): 849-53.

Ganser-Pornillos, B. K., Yeager, M. and Sundquist, W. I. (2008). "The structural biology of HIV assembly." Curr Opin Struct Biol **18**(2): 203-17.

Gao, X., Bashirova, A., Iversen, A. K., Phair, J., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Altfeld, M., O'Brien, S. J. and Carrington, M. (2005). "AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis." Nat Med **11**(12): 1290-2.

Gao, X., Nelson, G. W., Karacki, P., Martin, M. P., Phair, J., Kaslow, R., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., O'Brien, S. J. and Carrington, M. (2001). "Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS." N Engl J Med **344**(22): 1668-75.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B. H., Bhattacharya, T. and Korber, B. (2002). "Diversity considerations in HIV-1 vaccine selection." Science **296**(5577): 2354-60.

Gatanaga, H., Das, D., Suzuki, Y., Yeh, D. D., Hussain, K. A., Ghosh, A. K. and Mitsuya, H. (2006). "Altered HIV-1 Gag protein interactions with cyclophilin A (CypA) on the acquisition of H219Q and H219P substitutions in the CypA binding loop." J Biol Chem **281**(2): 1241-50.

Gea-Banacloche, J. C., Migueles, S. A., Martino, L., Shupert, W. L., McNeil, A. C., Sabbaghian, M. S., Ehler, L., Prussin, C., Stevens, R., Lambert, L., Altman, J., Hallahan, C. W., de Quiros, J. C. and Connors, M. (2000). "Maintenance of large numbers of virus-specific CD8+ T cells in HIV-infected progressors and long-term nonprogressors." J Immunol **165**(2): 1082-92.

Geldmacher, C., Currier, J. R., Herrmann, E., Haule, A., Kuta, E., McCutchan, F., Njovu, L., Geis, S., Hoffmann, O., Maboko, L., Williamson, C., Birx, D., Meyerhans, A., Cox, J. and Hoelscher, M. (2007). "CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients." J Virol **81**(5): 2440-8.

Gilbert, M. T., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E. and Worobey, M. (2007). "The emergence of HIV/AIDS in the Americas and beyond." Proc Natl Acad Sci U S A **104**(47): 18566-70.

Gillespie, G. M., Stewart-Jones, G., Rengasamy, J., Beattie, T., Bwayo, J. J., Plummer, F. A., Kaul, R., McMichael, A. J., Easterbrook, P., Dong, T., Jones, E. Y. and Rowland-Jones, S. L. (2006). "Strong TCR conservation and altered T cell cross-reactivity characterize a B*57-restricted immune response in HIV-1 infection." J Immunol **177**(6): 3893-902.

Giraldo-Vela, J. P., Rudersdorf, R., Chung, C., Qi, Y., Wallace, L. T., Bimber, B., Borchardt, G. J., Fisk, D. L., Glidden, C. E., Loffredo, J. T., Piaskowski, S. M., Furlott, J. R., Morales-Martinez, J. P., Wilson, N. A., Rehrauer, W. M., Lifson, J. D., Carrington, M. and Watkins, D. I. (2008). "The major histocompatibility complex class II alleles Mamu-DRB1*1003 and -DRB1*0306 are enriched in a cohort of simian immunodeficiency virus-infected rhesus macaque elite controllers." J Virol **82**(2): 859-70.

Goepfert, P. A., Lumm, W., Farmer, P., Matthews, P., Prendergast, A., Carlson, J. M., Derdeyn, C. A., Tang, J., Kaslow, R. A., Bansal, A., Yusim, K., Heckerman, D., Mulenga, J., Allen, S., Goulder, P. J. and Hunter, E. (2008). "Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients." J Exp Med **205**(5): 1009-17.

Goila-Gaur, R. and Strebel, K. (2008). "HIV-1 Vif, APOBEC, and intrinsic immunity." Retrovirology **5**: 51.

Gorry, P. R., McPhee, D. A., Verity, E., Dyer, W. B., Wesselingh, S. L., Learmont, J., Sullivan, J. S., Roche, M., Zaunders, J. J., Gabuzda, D., Crowe, S. M., Mills, J., Lewin, S. R., Brew, B. J., Cunningham, A. L. and Churchill, M. J. (2007). "Pathogenicity and immunogenicity of attenuated, nef-deleted HIV-1 strains in vivo." Retrovirology **4**: 66.

Goulder, P. J., Brander, C., Tang, Y., Tremblay, C., Colbert, R. A., Addo, M. M., Rosenberg, E. S., Nguyen, T., Allen, R., Trocha, A., Altfeld, M., He, S., Bunce, M., Funkhouser, R., Pelton, S. I., Burchett, S. K., McIntosh, K., Korber, B. T. and Walker, B. D. (2001). "Evolution and transmission of stable CTL escape mutations in HIV infection." Nature **412**(6844): 334-8.

Goulder, P. J., Bunce, M., Krausa, P., McIntyre, K., Crowley, S., Morgan, B., Edwards, A., Giangrande, P., Phillips, R. E. and McMichael, A. J. (1996). "Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection." AIDS Res Hum Retroviruses **12**(18): 1691-8.

Goulder, P. J., Bunce, M., Luzzi, G., Phillips, R. E. and McMichael, A. J. (1997). "Potential underestimation of HLA-C-restricted cytotoxic T-lymphocyte responses." Aids **11**(15): 1884-6.

Goulder, P. J., Phillips, R. E., Colbert, R. A., McAdam, S., Ogg, G., Nowak, M. A., Giangrande, P., Luzzi, G., Morgan, B., Edwards, A., McMichael, A. J. and Rowland-Jones, S. (1997). "Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS." Nat Med **3**(2): 212-7.

Goulder, P. J., Sewell, A. K., Lalloo, D. G., Price, D. A., Whelan, J. A., Evans, J., Taylor, G. P., Luzzi, G., Giangrande, P., Phillips, R. E. and McMichael, A. J. (1997). "Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation." J Exp Med **185**(8): 1423-33.

Goulder, P. J. and Watkins, D. I. (2004). "HIV and SIV CTL escape: implications for vaccine design." Nat Rev Immunol **4**(8): 630-40.

Goulder, P. J. and Watkins, D. I. (2008). "Impact of MHC class I diversity on immune control of immunodeficiency virus replication." Nat Rev Immunol **8**(8): 619-30.

Gray, R. H., Wawer, M. J., Brookmeyer, R., Sewankambo, N. K., Serwadda, D., Wabwire-Mangen, F., Lutalo, T., Li, X., vanCott, T. and Quinn, T. C. (2001). "Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda." Lancet **357**(9263): 1149-53.

Greenier, J. L., Miller, C. J., Lu, D., Dailey, P. J., Lu, F. X., Kunstman, K. J., Wolinsky, S. M. and Marthas, M. L. (2001). "Route of simian immunodeficiency virus inoculation determines the complexity but not the identity of viral variant populations that infect rhesus macaques." J Virol **75**(8): 3753-65.

Grigorov, B., Decimo, D., Smagulova, F., Pechoux, C., Mougel, M., Muriaux, D. and Darlix, J. L. (2007). "Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers." Retrovirology **4**: 54.

Groot, F., Welsch, S. and Sattentau, Q. J. (2008). "Efficient HIV-1 transmission from macrophages to T cells across transient virological synapses." Blood **111**(9): 4660-3.

Harari, A., Cellerai, C., Enders, F. B., Kostler, J., Codarri, L., Tapia, G., Boyman, O., Castro, E., Gaudieri, S., James, I., John, M., Wagner, R., Mallal, S. and Pantaleo, G. (2007). "Skewed association of polyfunctional antigen-specific CD8 T cell populations with HLA-B genotype." Proc Natl Acad Sci U S A **104**(41): 16233-8.

Harari, A., Enders, F. B., Cellerai, C., Bart, P. A. and Pantaleo, G. (2009). "Distinct profiles of cytotoxic granules in memory CD8 T cells correlate with function, differentiation stage, and antigen exposure." J Virol **83**(7): 2862-71.

Harndahl, M., Justesen, S., Lamberth, K., Roder, G., Nielsen, M. and Buus, S. (2009). "Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays." J Biomol Screen **14**(2): 173-80.

Hartley, O., Klasse, P. J., Sattentau, Q. J. and Moore, J. P. (2005). "V3: HIV's switch-hitter." AIDS Res Hum Retroviruses **21**(2): 171-89.

Hasegawa, M., Kishino, H. and Yano, T. (1985). "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA." J Mol Evol **22**(2): 160-74.

Hatziioannou, T., Cowan, S., Von Schwedler, U. K., Sundquist, W. I. and Bieniasz, P. D. (2004). "Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid." J Virol **78**(11): 6005-12.

Hazenberg, M. D., Hamann, D., Schuitemaker, H. and Miedema, F. (2000). "T cell depletion in HIV-1 infection: how CD4+ T cells go out of stock." Nat Immunol **1**(4): 285-9.

Hearps, A. C. and Jans, D. A. (2007). "Regulating the functions of the HIV-1 matrix protein." AIDS Res Hum Retroviruses **23**(3): 341-6.

HIV Sequence Compendium 2009, Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

Hladik, F. and McElrath, M. J. (2008). "Setting the stage: host invasion by HIV." Nat Rev Immunol **8**(6): 447-57.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995). "Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection." Nature **373**(6510): 123-6.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A. and Brown, A. J. (1992). "Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient." Proc Natl Acad Sci U S A **89**(11): 4835-9.

Honeyborne, I., Prendergast, A., Pereyra, F., Leslie, A., Crawford, H., Payne, R., Reddy, S., Bishop, K., Moodley, E., Nair, K., van der Stok, M., McCarthy, N., Rousseau, C. M., Addo, M., Mullins, J. I., Brander, C., Kiepiela, P., Walker, B. D. and Goulder, P. J. (2007). "Control of human immunodeficiency virus type 1 is associated with HLA-B*13 and targeting of multiple gag-specific CD8+ T-cell epitopes." J Virol **81**(7): 3667-72.

Honeyborne, I., Rathod, A., Buchli, R., Ramduth, D., Moodley, E., Rathnavalu, P., Chetty, S., Day, C., Brander, C., Hildebrand, W., Walker, B. D., Kiepiela, P. and Goulder, P. J. (2006). "Motif inference reveals optimal CTL epitopes presented by HLA class I alleles highly prevalent in southern Africa." J Immunol **176**(8): 4699-705.

Hu, J., Gardner, M. B. and Miller, C. J. (2000). "Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells." J Virol **74**(13): 6087-95.

Hu, W. S. and Temin, H. M. (1990). "Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination." Proc Natl Acad Sci U S A **87**(4): 1556-60.

Iafrate, A. J., Carl, S., Bronson, S., Stahl-Hennig, C., Swigut, T., Skowronski, J. and Kirchhoff, F. (2000). "Disrupting surfaces of nef required for downregulation of CD4 and for enhancement of virion infectivity attenuates simian immunodeficiency virus replication in vivo." J Virol **74**(21): 9836-44.

Janeway, C. A., Travers, P., Walport, M. and Schlomchik, M. (2005). Immunobiology: The Immune System in Health and Disease. New York and London, Garland Science.

Jaskolski, M., Alexandratos, J. N., Bujacz, G. and Wlodawer, A. (2009). "Piecing together the structure of retroviral integrase, an important target in AIDS therapy." Febs J **276**(11): 2926-46.

Jin, X., Bauer, D. E., Tuttleton, S. E., Lewin, S., Gettie, A., Blanchard, J., Irwin, C. E., Safrit, J. T., Mittler, J., Weinberger, L., Kostrikis, L. G., Zhang, L., Perelson, A. S. and Ho, D. D. (1999). "Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques." J Exp Med **189**(6): 991-8.

Jin, X., Gao, X., Ramanathan, M., Jr., Deschenes, G. R., Nelson, G. W., O'Brien, S. J., Goedert, J. J., Ho, D. D., O'Brien, T. R. and Carrington, M. (2002). "Human immunodeficiency virus type 1 (HIV-1)-specific CD8+-T-cell responses for groups of HIV-1-infected individuals with different HLA-B*35 genotypes." J Virol **76**(24): 12603-10.

Johnston, M. I. and Fauci, A. S. (2007). "An HIV vaccine--evolving concepts." N Engl J Med **356**(20): 2073-81.

Kanki, P. J., Hamel, D. J., Sankale, J. L., Hsieh, C., Thior, I., Barin, F., Woodcock, S. A., Gueye-Ndiaye, A., Zhang, E., Montano, M., Siby, T., Marlink, R., I, N. D., Essex, M. E. and S, M. B. (1999). "Human immunodeficiency virus type 1 subtypes differ in disease progression." J Infect Dis **179**(1): 68-73.

Kaslow, R. A., Carrington, M., Apple, R., Park, L., Munoz, A., Saah, A. J., Goedert, J. J., Winkler, C., O'Brien, S. J., Rinaldo, C., Detels, R., Blattner, W., Phair, J., Erlich, H. and Mann, D. L. (1996). "Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection." Nat Med **2**(4): 405-11.

Katz, R. A. and Skalka, A. M. (1990). "Generation of diversity in retroviruses." Annu Rev Genet **24**: 409-45.

Kaufmann, S. H. (2007). "The contribution of immunology to the rational design of novel antibacterial vaccines." Nat Rev Microbiol **5**(7): 491-504.

Kawada, M., Igarashi, H., Takeda, A., Tsukamoto, T., Yamamoto, H., Dohki, S., Takiguchi, M. and Matano, T. (2006). "Involvement of multiple epitope-specific cytotoxic T-lymphocyte responses in vaccine-based control of simian immunodeficiency virus replication in rhesus macaques." J Virol **80**(4): 1949-58.

Kawada, M., Tsukamoto, T., Yamamoto, H., Iwamoto, N., Kurihara, K., Takeda, A., Moriya, C., Takeuchi, H., Akari, H. and Matano, T. (2008). "Gag-specific cytotoxic T-lymphocyte-based control of primary simian immunodeficiency virus replication in a vaccine trial." J Virol **82**(20): 10199-206.

Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga, H., Fujiwara, M., Hachiya, A., Koizumi, H., Kuse, N., Oka, S., Duda, A., Prendergast, A., Crawford, H., Leslie, A., Brumme, Z., Brumme, C., Allen, T., Brander, C., Kaslow, R., Tang, J., Hunter, E., Allen, S., Mulenga, J., Branch, S., Roach, T., John, M., Mallal, S., Ogwu, A., Shapiro, R., Prado, J. G., Fidler, S., Weber, J., Pybus, O. G., Klenerman, P., Ndung'u, T., Phillips, R., Heckerman, D., Harrigan, P. R., Walker, B. D., Takiguchi, M. and Goulder, P. (2009). "Adaptation of HIV-1 to human leukocyte antigen class I." Nature **458**(7238): 641-5.

Kearney, M., Maldarelli, F., Shao, W., Margolick, J. B., Daar, E. S., Mellors, J. W., Rao, V., Coffin, J. M. and Palmer, S. (2009). "Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals." J Virol **83**(6): 2715-27.

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J. L., Gao, F., Anderson, J. A., Ping, L. H., Swanstrom, R., Tomaras, G. D., Blattner, W. A., Goepfert, P. A., Kilby, J. M., Saag, M. S., Delwart, E. L., Busch, M. P., Cohen, M. S., Montefiori, D. C., Haynes, B. F., Gaschen, B., Athreya, G. S., Lee, H. Y., Wood, N., Seoighe, C., Perelson, A. S., Bhattacharya, T., Korber, B. T., Hahn, B. H. and Shaw, G. M. (2008). "Identification and characterization of transmitted and early founder virus

envelopes in primary HIV-1 infection." Proc Natl Acad Sci U S A **105**(21): 7552-7.

Kelleher, A. D., Long, C., Holmes, E. C., Allen, R. L., Wilson, J., Conlon, C., Workman, C., Shaunak, S., Olson, K., Goulder, P., Brander, C., Ogg, G., Sullivan, J. S., Dyer, W., Jones, I., McMichael, A. J., Rowland-Jones, S. and Phillips, R. E. (2001). "Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses." J Exp Med **193**(3): 375-86.

Kestler, H. W., 3rd, Ringler, D. J., Mori, K., Panicali, D. L., Sehgal, P. K., Daniel, M. D. and Desrosiers, R. C. (1991). "Importance of the nef gene for maintenance of high virus loads and for development of AIDS." Cell **65**(4): 651-62.

Kiepiela, P., Leslie, A. J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., Rathnavalu, P., Moore, C., Pfafferott, K. J., Hilton, L., Zimbwa, P., Moore, S., Allen, T., Brander, C., Addo, M. M., Altfeld, M., James, I., Mallal, S., Bunce, M., Barber, L. D., Szinger, J., Day, C., Klenerman, P., Mullins, J., Korber, B., Coovadia, H. M., Walker, B. D. and Goulder, P. J. (2004). "Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA." Nature **432**(7018): 769-75.

Kiepiela, P., Ngumbela, K., Thobakgale, C., Ramduth, D., Honeyborne, I., Moodley, E., Reddy, S., de Pierres, C., Mncube, Z., Mkhwanazi, N., Bishop, K., van der Stok, M., Nair, K., Khan, N., Crawford, H., Payne, R., Leslie, A., Prado, J., Prendergast, A., Frater, J., McCarthy, N., Brander, C., Learn, G. H., Nickle, D., Rousseau, C., Coovadia, H., Mullins, J. I., Heckerman, D., Walker, B. D. and Goulder, P. (2007). "CD8+ T-cell responses to different HIV proteins have discordant associations with viral load." Nat Med **13**(1): 46-53.

Klein, M. R., van Baalen, C. A., Holwerda, A. M., Kerkhof Garde, S. R., Bende, R. J., Keet, I. P., Eeftinck-Schattenkerk, J. K., Osterhaus, A. D., Schuitemaker, H. and Miedema, F. (1995). "Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics." J Exp Med **181**(4): 1365-72.

Kobayashi, M., Igarashi, H., Takeda, A., Kato, M. and Matano, T. (2005). "Reversion in vivo after inoculation of a molecular proviral DNA clone of simian immunodeficiency virus with a cytotoxic-T-lymphocyte escape mutation." J Virol **79**(17): 11529-32.

Koff, W. C., Johnson, P. R., Watkins, D. I., Burton, D. R., Lifson, J. D., Hasenkrug, K. J., McDermott, A. B., Schultz, A., Zamb, T. J., Boyle, R. and Desrosiers, R. C. (2006). "HIV vaccine design: insights from live attenuated SIV vaccines." Nat Immunol **7**(1): 19-23.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C. and Detours, V. (2001). "Evolutionary and immunological implications of contemporary HIV-1 variation." Br Med Bull **58**: 19-42.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. and Bhattacharya, T. (2000). "Timing the ancestor of the HIV-1 pandemic strains." Science **288**(5472): 1789-96.

Koup, R. A., Safrit, J. T., Cao, Y., Andrews, C. A., McLeod, G., Borkowsky, W., Farthing, C. and Ho, D. D. (1994). "Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome." J Virol **68**(7): 4650-5.

Lambotte, O., Ferrari, G., Moog, C., Yates, N. L., Liao, H. X., Parks, R. J., Hicks, C. B., Owzar, K., Tomaras, G. D., Montefiori, D. C., Haynes, B. F. and Delfraissy, J. F. (2009). "Heterogeneous neutralizing antibody and antibody-dependent cell cytotoxicity responses in HIV-1 elite controllers." Aids **23**(8): 897-906.

Lau, K. A., Wang, B. and Saksena, N. K. (2007). "Emerging trends of HIV epidemiology in Asia." AIDS Rev **9**(4): 218-29.

Le Gall, S., Stamegna, P. and Walker, B. D. (2007). "Portable flanking sequences modulate CTL epitope processing." J Clin Invest **117**(11): 3563-75.

Le Rouzic, E. and Benichou, S. (2005). "The Vpr protein from HIV-1: distinct roles along the viral life cycle." Retrovirology **2**: 11.

Learmont, J. C., Geczy, A. F., Mills, J., Ashton, L. J., Raynes-Greenow, C. H., Garsia, R. J., Dyer, W. B., McIntyre, L., Oelrichs, R. B., Rhodes, D. I., Deacon, N. J. and Sullivan, J. S. (1999). "Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort." N Engl J Med **340**(22): 1715-22.

Lee, J. K., Stewart-Jones, G., Dong, T., Harlos, K., Di Gleria, K., Dorrell, L., Douek, D. C., van der Merwe, P. A., Jones, E. Y. and McMichael, A. J. (2004). "T cell cross-reactivity and conformational changes during TCR engagement." J Exp Med **200**(11): 1455-66.

Lemey, P., Derdelinckx, I., Rambaut, A., Van Laethem, K., Dumont, S., Vermeulen, S., Van Wijngaerden, E. and Vandamme, A. M. (2005). "Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain." J Virol **79**(18): 11981-9.

Leslie, A., Kavanagh, D., Honeyborne, I., Pfafferott, K., Edwards, C., Pillay, T., Hilton, L., Thobakgale, C., Ramduth, D., Draenert, R., Le Gall, S., Luzzi, G., Edwards, A., Brander, C., Sewell, A. K., Moore, S., Mullins, J., Moore, C., Mallal, S., Bhardwaj, N., Yusim, K., Phillips, R., Klenerman, P., Korber, B., Kiepiela, P., Walker, B. and Goulder, P. (2005). "Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA." J Exp Med **201**(6): 891-902.

Leslie, A. J., Pfafferott, K. J., Chetty, P., Draenert, R., Addo, M. M., Feeney, M., Tang, Y., Holmes, E. C., Allen, T., Prado, J. G., Altfeld, M., Brander, C., Dixon, C., Ramduth, D., Jeena, P., Thomas, S. A., St John, A., Roach, T. A., Kupfer, B., Luzzi, G., Edwards, A., Taylor, G., Lyall, H., Tudor-Williams, G., Novelli, V., Martinez-Picado, J., Kiepiela, P., Walker, B. D. and Goulder, P. J. (2004). "HIV evolution: CTL escape mutation and reversion after transmission." Nat Med **10**(3): 282-9.

Lewis, M. J., Balamurugan, A., Ohno, A., Kilpatrick, S., Ng, H. L. and Yang, O. O. (2008). "Functional adaptation of Nef to the immune milieu of HIV-1 infection in vivo." J Immunol **180**(6): 4075-81.

Lewis, P. O. (1998). "A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data." Mol Biol Evol **15**(3): 277-83.

Li B, G. A., Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM. (2007). "Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution." Journal of Virology **81**(1): 193-201.

Li, H., Dou, J., Ding, L. and Spearman, P. (2007). "Myristoylation is required for human immunodeficiency virus type 1 Gag-Gag multimerization in mammalian cells." J Virol **81**(23): 12899-910.

Lichterfeld, M., Yu, X. G., Cohen, D., Addo, M. M., Malenfant, J., Perkins, B., Pae, E., Johnston, M. N., Strick, D., Allen, T. M., Rosenberg, E. S., Korber, B., Walker, B. D. and Altfeld, M. (2004). "HIV-1 Nef is preferentially recognized by CD8 T cells in primary HIV-1 infection despite a relatively high degree of genetic diversity." Aids **18**(10): 1383-92.

Little, S. J., Frost, S. D., Wong, J. K., Smith, D. M., Pond, S. L., Ignacio, C. C., Parkin, N. T., Petropoulos, C. J. and Richman, D. D. (2008). "Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection." J Virol **82**(11): 5510-8.

Liu, J., Ewald, B. A., Lynch, D. M., Denholtz, M., Abbink, P., Lemckert, A. A., Carville, A., Mansfield, K. G., Havenga, M. J., Goudsmit, J. and Barouch, D. H. (2008). "Magnitude and phenotype of cellular immune responses elicited by recombinant adenovirus vectors and heterologous prime-boost regimens in rhesus monkeys." J Virol **82**(10): 4844-52.

Liu, J., O'Brien, K. L., Lynch, D. M., Simmons, N. L., La Porte, A., Riggs, A. M., Abbink, P., Coffey, R. T., Grandpre, L. E., Seaman, M. S., Landucci, G., Forthal, D. N., Montefiori, D. C., Carville, A., Mansfield, K. G., Havenga, M. J., Pau, M. G., Goudsmit, J. and Barouch, D. H. (2009). "Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys." Nature **457**(7225): 87-91.

Liu, Y., McNevin, J., Zhao, H., Tebit, D. M., Troyer, R. M., McSweyn, M., Ghosh, A. K., Shriner, D., Arts, E. J., McElrath, M. J. and Mullins, J. I. (2007). "Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape." J Virol **81**(22): 12179-88.

Loffredo, J. T., Bean, A. T., Beal, D. R., Leon, E. J., May, G. E., Piaskowski, S. M., Furlott, J. R., Reed, J., Musani, S. K., Rakasz, E. G., Friedrich, T. C., Wilson, N. A., Allison, D. B. and Watkins, D. I. (2008). "Patterns of CD8+ immunodominance may influence the ability of Mamu-B*08-positive macaques to naturally control simian immunodeficiency virus SIVmac239 replication." J Virol **82**(4): 1723-38.

Loffredo, J. T., Rakasz, E. G., Giraldo, J. P., Spencer, S. P., Grafton, K. K., Martin, S. R., Napoe, G., Yant, L. J., Wilson, N. A. and Watkins, D. I. (2005). "Tat(28-35)SL8-specific CD8+ T lymphocytes are more effective than Gag(181-189)CM9-specific CD8+ T lymphocytes at suppressing simian immunodeficiency virus replication in a functional in vitro assay." J Virol **79**(23): 14986-91.

Lynch, R. M., Shen, T., Gnanakaran, S. and Derdeyn, C. A. (2009). "Appreciating HIV type 1 diversity: subtype differences in Env." AIDS Res Hum Retroviruses **25**(3): 237-48.

Mackelprang, R. D., John-Stewart, G., Carrington, M., Richardson, B., Rowland-Jones, S., Gao, X., Mbori-Ngacha, D., Mabuka, J., Lohman-Payne, B. and Farquhar, C. (2008). "Maternal HLA homozygosity and mother-child HLA concordance increase the risk of vertical transmission of HIV-1." J Infect Dis **197**(8): 1156-61.

Maddison, D. R. and Maddison, W. P. (2000). MacClade. Analysis of phylogeny and character evolution. 4th edition., Sinauer Associates, Sunderland Mass.

Malim, M. H. and Emerman, M. (2008). "HIV-1 accessory proteins--ensuring viral survival in a hostile environment." Cell Host Microbe **3**(6): 388-98.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. and Trono, D. (2003). "Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts." Nature **424**(6944): 99-103.

Marsh, S. G. E., Parham, P. and Barber, L. D. (2000). The HLA Facts Book, Academic Press, London NW1 7DX.

Martinez-Picado, J. and Martinez, M. A. (2008). "HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo." Virus Res **134**(1-2): 104-23.

Martinez-Picado, J., Prado, J. G., Fry, E. E., Pfafferott, K., Leslie, A., Chetty, S., Thobakgale, C., Honeyborne, I., Crawford, H., Matthews, P., Pillay, T., Rousseau, C., Mullins, J. I., Brander, C., Walker, B. D., Stuart, D. I., Kiepiela, P. and Goulder, P. (2006). "Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1." J Virol **80**(7): 3617-23.

Mascola, J. R. (2002). "Passive transfer studies to elucidate the role of antibody-mediated protection against HIV-1." Vaccine **20**(15): 1922-5.

Mascola, J. R., Snyder, S. W., Weislow, O. S., Belay, S. M., Belshe, R. B., Schwartz, D. H., Clements, M. L., Dolin, R., Graham, B. S., Gorse, G. J., Keefer, M. C., McElrath, M. J., Walker, M. C., Wagner, K. F., McNeil, J. G., McCutchan, F. E. and Burke, D. S. (1996). "Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group." J Infect Dis **173**(2): 340-8.

Mascola, J. R., Stiegler, G., VanCott, T. C., Katinger, H., Carpenter, C. B., Hanson, C. E., Beary, H., Hayes, D., Frankel, S. S., Birx, D. L. and Lewis, M. G. (2000). "Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies." Nat Med **6**(2): 207-10.

Masemola, A., Mashishi, T., Khoury, G., Mohube, P., Mokgotho, P., Vardas, E., Colvin, M., Zijenah, L., Katzenstein, D., Musonda, R., Allen, S., Kumwenda, N., Taha, T., Gray, G., McIntyre, J., Karim, S. A., Sheppard, H. W. and Gray, C. M. (2004). "Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8+ T cells: correlation with viral load." J Virol **78**(7): 3233-43.

Matano, T., Kobayashi, M., Igarashi, H., Takeda, A., Nakamura, H., Kano, M., Sugimoto, C., Mori, K., Iida, A., Hirata, T., Hasegawa, M., Yuasa, T., Miyazawa, M., Takahashi, Y., Yasunami, M., Kimura, A., O'Connor, D. H., Watkins, D. I. and Nagai, Y. (2004). "Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial." J Exp Med **199**(12): 1709-18.

Matthews, P. C., Leslie, A. J., Katzourakis, A., Crawford, H., Payne, R., Prendergast, A., Power, K., Kelleher, A. D., Klenerman, P., Carlson, J., Heckerman, D., Ndung'u, T., Walker, B. D., Allen, T. M., Pybus, O. G. and Goulder, P. J. (2009). "HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering." J Virol **83**(9): 4605-15.

Matthews, P. C., Prendergast, A., Leslie, A., Crawford, H., Payne, R., Rousseau, C., Rolland, M., Honeyborne, I., Carlson, J., Kadie, C., Brander, C., Bishop, K., Mlotshwa, N., Mullins, J. I., Coovadia, H., Ndung'u, T., Walker, B. D., Heckerman, D. and Goulder, P. J. (2008). "Central role of reverting mutations in HLA associations with human immunodeficiency virus set point." J Virol **82**(17): 8548-59.

McCutchan, F. E., Artenstein, A. W., Sanders-Buell, E., Salminen, M. O., Carr, J. K., Mascola, J. R., Yu, X. F., Nelson, K. E., Khamboonruang, C., Schmitt, D., Kieny, M. P., McNeil, J. G. and Burke, D. S. (1996). "Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa." J Virol **70**(6): 3331-8.

McElrath, M. J., De Rosa, S. C., Moodie, Z., Dubey, S., Kierstead, L., Janes, H., Defawe, O. D., Carter, D. K., Hural, J., Akondy, R., Buchbinder, S. P., Robertson, M. N., Mehrotra, D. V., Self, S. G., Corey, L., Shiver, J. W. and Casimiro, D. R. (2008). "HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis." Lancet **372**(9653): 1894-905.

McMichael, A. and Klenerman, P. (2002). "HIV/AIDS. HLA leaves its footprints on HIV." Science **296**(5572): 1410-1.

Mehra, N. K. and Kaur, G. (2003). "MHC-based vaccination approaches: progress and perspectives." Expert Rev Mol Med **5**(7): 1-17.

Mellors, J. W., Rinaldo, C. R., Jr., Gupta, P., White, R. M., Todd, J. A. and Kingsley, L. A. (1996). "Prognosis in HIV-1 infection predicted by the quantity of virus in plasma." Science **272**(5265): 1167-70.

Migueles, S. A. and Connors, M. (2001). "Frequency and function of HIV-specific CD8(+) T cells." Immunol Lett **79**(1-2): 141-50.

Migueles, S. A., Laborico, A. C., Imamichi, H., Shupert, W. L., Royce, C., McLaughlin, M., Ehler, L., Metcalf, J., Liu, S., Hallahan, C. W. and Connors, M. (2003). "The differential ability of HLA B*5701+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences." J Virol **77**(12): 6889-98.

Migueles, S. A., Laborico, A. C., Shupert, W. L., Sabbaghian, M. S., Rabin, R., Hallahan, C. W., Van Baarle, D., Kostense, S., Miedema, F., McLaughlin, M., Ehler, L., Metcalf, J., Liu, S. and Connors, M. (2002). "HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors." Nat Immunol **3**(11): 1061-8.

Migueles, S. A., Osborne, C. M., Royce, C., Compton, A. A., Joshi, R. P., Weeks, K. A., Rood, J. E., Berkley, A. M., Sacha, J. B., Cogliano-Shutta, N. A., Lloyd, M., Roby, G., Kwan, R., McLaughlin, M., Stallings, S., Rehm, C., O'Shea, M. A., Mican, J., Packard, B. Z., Komoriya, A., Palmer, S., Wiegand, A. P., Maldarelli, F., Coffin, J. M., Mellors, J. W., Hallahan, C. W., Follman, D. A. and Connors, M. (2008). "Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control." Immunity **29**(6): 1009-21.

Migueles, S. A., Sabbaghian, M. S., Shupert, W. L., Bettinotti, M. P., Marincola, F. M., Martino, L., Hallahan, C. W., Selig, S. M., Schwartz, D., Sullivan, J. and Connors, M. (2000). "HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors." Proc Natl Acad Sci U S A **97**(6): 2709-14.

Milicic, A., Price, D. A., Zimbwa, P., Booth, B. L., Brown, H. L., Easterbrook, P. J., Olsen, K., Robinson, N., Gileadi, U., Sewell, A. K., Cerundolo, V. and Phillips, R. E. (2005). "CD8+ T cell epitope-flanking mutations disrupt proteasomal processing of HIV-1 Nef." J Immunol **175**(7): 4618-26.

Miller, C. J. (1998). "Does viral tropism play a role in heterosexual transmission of HIV? Findings in the SIV-rhesus macaque model." AIDS Res Hum Retroviruses **14 Suppl 1**: S79-82.

Miura, T., Brockman, M. A., Schneidewind, A., Lobritz, M., Pereyra, F., Rathod, A., Block, B. L., Brumme, Z. L., Brumme, C. J., Baker, B., Rothchild, A. C., Li, B., Trocha, A., Cutrell, E., Frahm, N., Brander, C., Toth, I., Arts, E. J., Allen, T. M. and Walker, B. D. (2009). "HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphotye recognition." J Virol **83**(6): 2743-55.

Moir, S. and Fauci, A. S. (2009). "B cells in HIV infection and disease." Nat Rev Immunol **9**(4): 235-45.

Montefiori, D., Sattentau, Q., Flores, J., Esparza, J. and Mascola, J. (2007). "Antibody-based HIV-1 vaccines: recent developments and future directions." PLoS Med **4**(12): e348.

Montero, M., van Houten, N. E., Wang, X. and Scott, J. K. (2008). "The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: dominant site of antibody neutralization and target for vaccine design." Microbiol Mol Biol Rev **72**(1): 54-84, table of contents.

Moore, C. B., John, M., James, I. R., Christiansen, F. T., Witt, C. S. and Mallal, S. A. (2002). "Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level." Science **296**(5572): 1439-43.

Moore, J. P., Klasse, P. J., Dolan, M. J. and Ahuja, S. K. (2008). "AIDS/HIV. A STEP into darkness or light?" Science **320**(5877): 753-5.

Ngumbela, K. C., Day, C. L., Mncube, Z., Nair, K., Ramduth, D., Thobakgale, C., Moodley, E., Reddy, S., de Pierres, C., Mkhwanazi, N., Bishop, K., van der Stok, M., Ismail, N., Honeyborne, I., Crawford, H., Kavanagh, D. G., Rousseau, C., Nickle, D., Mullins, J., Heckerman, D., Korber, B., Coovadia, H., Kiepiela, P., Goulder, P. J. and Walker, B. D. (2008). "Targeting of a CD8 T cell env epitope presented by HLA-B*5802 is associated with markers of HIV disease progression and lack of selection pressure." AIDS Res Hum Retroviruses **24**(1): 72-82.

Nitayaphan, S., Pitisuttithum, P., Karnasuta, C., Eamsila, C., de Souza, M., Morgan, P., Polonis, V., Benenson, M., VanCott, T., Ratto-Kim, S., Kim, J., Thapinta, D., Garner, R., Bussaratid, V., Singharaj, P., el-Habib, R., Gurunathan, S., Heyward, W., Birx, D., McNeil, J. and Brown, A. E. (2004). "Safety and immunogenicity of an HIV subtype B and E prime-boost vaccine combination in HIV-negative Thai adults." J Infect Dis **190**(4): 702-6.

Nixon, D. F., Townsend, A. R., Elvin, J. G., Rizza, C. R., Gallwey, J. and McMichael, A. J. (1988). "HIV-1 gag-specific cytotoxic T lymphocytes defined with recombinant vaccinia virus and synthetic peptides." Nature **336**(6198): 484-7.

Nkengasong, J. N., Janssens, W., Heyndrickx, L., Fransen, K., Ndumbe, P. M., Motte, J., Leonaers, A., Ngolle, M., Ayuk, J., Piot, P. and et al. (1994). "Genotypic subtypes of HIV-1 in Cameroon." Aids **8**(10): 1405-12.

Nomaguchi, M., Fujita, M. and Adachi, A. (2008). "Role of HIV-1 Vpu protein for virus spread and pathogenesis." Microbes Infect **10**(9): 960-7.

O'Brien, S. J., Gao, X. and Carrington, M. (2001). "HLA and AIDS: a cautionary tale." Trends Mol Med **7**(9): 379-81.

O'Neill, E., Kuo, L. S., Krisko, J. F., Tomchick, D. R., Garcia, J. V. and Foster, J. L. (2006). "Dynamic evolution of the human immunodeficiency virus type 1 pathogenic factor, Nef." J Virol **80**(3): 1311-20.

Ogg, G. S., Jin, X., Bonhoeffer, S., Dunbar, P. R., Nowak, M. A., Monard, S., Segal, J. P., Cao, Y., Rowland-Jones, S. L., Cerundolo, V., Hurley, A., Markowitz, M., Ho, D. D., Nixon, D. F. and McMichael, A. J. (1998). "Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA." Science **279**(5359): 2103-6.

Overbaugh, J., Kreiss, J., Poss, M., Lewis, P., Mostad, S., John, G., Nduati, R., Mbori-Ngacha, D., Martin, H., Jr., Richardson, B., Jackson, S., Neilson, J., Long, E. M., Panteleeff, D., Welch, M., Rakwar, J., Jackson, D., Chohan, B., Lavreys, L., Mandaliya, K., Ndinya-Achola, J. and Bwayo, J. (1999). "Studies of human immunodeficiency virus type 1 mucosal viral shedding and transmission in Kenya." J Infect Dis **179 Suppl 3**: S401-4.

Parham, P. (1992). "Immunology. Deconstructing the MHC." Nature **360**(6402): 300-1.

Parham, P. (2005). "MHC class I molecules and KIRs in human history, health and survival." Nat Rev Immunol **5**(3): 201-14.

Parren, P. W., Marx, P. A., Hessell, A. J., Luckay, A., Harouse, J., Cheng-Mayer, C., Moore, J. P. and Burton, D. R. (2001). "Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization in vitro." J Virol **75**(17): 8340-7.

Payne, R. P., Matthews, P. C., Prado, J. G. and Goulder, P. J. (2009). "HLA-Mediated Control of HIV and HIV Adaptation to HLA." Adv Parasitol **68**: 1-20.

Peaper, D. R. and Cresswell, P. (2008). "Regulation of MHC class I assembly and peptide binding." Annu Rev Cell Dev Biol **24**: 343-68.

Peeters, M. and Sharp, P. M. (2000). "Genetic diversity of HIV-1: the moving target." Aids **14 Suppl 3**: S129-40.

Phillips, R. E., Rowland-Jones, S., Nixon, D. F., Gotch, F. M., Edwards, J. P., Ogunlesi, A. O., Elvin, J. G., Rothbard, J. A., Bangham, C. R., Rizza, C. R. and et al. (1991). "Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition." Nature **354**(6353): 453-9.

Philpott, S. M. (2003). "HIV-1 coreceptor usage, transmission, and disease progression." Curr HIV Res **1**(2): 217-27.

Pitisuttithum, P., Berman, P. W., Phonrat, B., Suntharasamai, P., Raktham, S., Srisuwanvilai, L. O., Hirunras, K., Kitayaporn, D., Kaewkangwal, J., Migasena, S., Sheppard, H. W., Li, E., Chernow, M., Peterson, M. L., Shibata, R., Heyward, W. L. and Francis, D. P. (2004). "Phase I/II study of a candidate vaccine designed against the B and E subtypes of HIV-1." J Acquir Immune Defic Syndr **37**(1): 1160-5.

Plotkin, S. A. (2009). "Sang Froid in a time of trouble: is a vaccine against HIV possible?" J Int AIDS Soc **12**(1): 2.

Pollakis, G., Abebe, A., Kliphuis, A., Chalaby, M. I., Bakker, M., Mengistu, Y., Brouwer, M., Goudsmit, J., Schuitemaker, H. and Paxton, W. A. (2004). "Phenotypic and genotypic comparisons of CCR5- and CXCR4-tropic human immunodeficiency virus type 1 biological clones isolated from subtype C-infected individuals." J Virol **78**(6): 2841-52.

Prendergast, A., Tudor-Williams, G., Jeena, P., Burchett, S. and Goulder, P. (2007). "International perspectives, progress, and future challenges of paediatric HIV infection." Lancet **370**(9581): 68-80.

Preston, B. D. (1997). "Reverse transcriptase fidelity and HIV-1 variation." Science **275**(5297): 228-9; author reply 230-1.

Price, D. A., West, S. M., Betts, M. R., Ruff, L. E., Brenchley, J. M., Ambrozak, D. R., Edghill-Smith, Y., Kuroda, M. J., Bogdan, D., Kunstman, K., Letvin, N. L., Franchini, G., Wolinsky, S. M., Koup, R. A. and Douek, D. C. (2004). "T cell receptor recognition motifs govern immune escape patterns in acute SIV infection." Immunity **21**(6): 793-803.

Priddy, F. H., Brown, D., Kublin, J., Monahan, K., Wright, D. P., Lalezari, J., Santiago, S., Marmor, M., Lally, M., Novak, R. M., Brown, S. J., Kulkarni, P., Dubey, S. A., Kierstead, L. S., Casimiro, D. R., Mogg, R., DiNubile, M. J., Shiver, J. W., Leavitt, R. Y., Robertson, M. N., Mehrotra, D. V. and Quirk, E. (2008). "Safety and immunogenicity of a replication-incompetent adenovirus type 5 HIV-1 clade B gag/pol/nef vaccine in healthy adults." Clin Infect Dis **46**(11): 1769-81.

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M. O., Lutalo, T. and Gray, R. H. (2000). "Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group." N Engl J Med **342**(13): 921-9.

Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M. and Holmes, E. C. (2001). "Human immunodeficiency virus. Phylogeny and the origin of HIV-1." Nature **410**(6832): 1047-8.

Resh, M. D. (2005). "Intracellular trafficking of HIV-1 Gag: how Gag interacts with cell membranes and makes viral particles." AIDS Rev **7**(2): 84-91.

Ridley, M. (1983). The Explanation of Organic Diversity: The Comparative Method and Adaptions for Mating, Oxford University Press.

Riviere, Y., McChesney, M. B., Porrot, F., Tanneau-Salvadori, F., Sansonetti, P., Lopez, O., Pialoux, G., Feuillie, V., Mollereau, M., Chamaret, S. and et al. (1995). "Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS." AIDS Res Hum Retroviruses **11**(8): 903-7.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000). "HIV-1 nomenclature proposal." Science **288**(5463): 55-6.

Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C. and Harrison, S. C. (1995). "The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1." Proc Natl Acad Sci U S A **92**(4): 1222-6.

Rollman, E., Smith, M. Z., Brooks, A. G., Purcell, D. F., Zuber, B., Ramshaw, I. A. and Kent, S. J. (2007). "Killing kinetics of simian immunodeficiency virus-specific CD8+ T cells: implications for HIV vaccine strategies." J Immunol **179**(7): 4571-9.

Rosenberg, E. S., Billingsley, J. M., Caliendo, A. M., Boswell, S. L., Sax, P. E., Kalams, S. A. and Walker, B. D. (1997). "Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia." Science **278**(5342): 1447-50.

Rousseau, C. M., Daniels, M. G., Carlson, J. M., Kadie, C., Crawford, H., Prendergast, A., Matthews, P., Payne, R., Rolland, M., Raugi, D. N., Maust, B. S., Learn, G. H., Nickle, D. C., Coovadia, H., Ndung'u, T., Frahm, N., Brander, C., Walker, B. D., Goulder, P. J., Bhattacharya, T., Heckerman, D. E., Korber, B. T. and Mullins, J. I. (2008). "HLA Class-I Driven Evolution of Human Immunodeficiency Virus Type 1 Subtype C Proteome: Immune Escape and Viral Load." J Virol.

Roux, K. H. and Taylor, K. A. (2007). "AIDS virus envelope spike structure." Curr Opin Struct Biol **17**(2): 244-52.

Russell, R. S., Roldan, A., Detorio, M., Hu, J., Wainberg, M. A. and Liang, C. (2003). "Effects of a single amino acid substitution within the p2 region of human immunodeficiency virus type 1 on packaging of spliced viral RNA." J Virol **77**(24): 12986-95.

Sacha, J. B., Chung, C., Rakasz, E. G., Spencer, S. P., Jonas, A. K., Bean, A. T., Lee, W., Burwitz, B. J., Stephany, J. J., Loffredo, J. T., Allison, D. B., Adnan, S., Hoji, A., Wilson, N. A., Friedrich, T. C., Lifson, J. D., Yang, O. O. and Watkins, D. I. (2007). "Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression." J Immunol **178**(5): 2746-54.

Sacha, J. B., Chung, C., Reed, J., Jonas, A. K., Bean, A. T., Spencer, S. P., Lee, W., Vojnov, L., Rudersdorf, R., Friedrich, T. C., Wilson, N. A., Lifson, J. D. and Watkins, D. I. (2007). "Pol-specific CD8+ T cells recognize simian immunodeficiency virus-infected cells prior to Nef-mediated major histocompatibility complex class I downregulation." J Virol **81**(21): 11703-12.

Sacha, J. B., Giraldo-Vela, J. P., Buechler, M. B., Martins, M. A., Maness, N. J., Chung, C., Wallace, L. T., Leon, E. J., Friedrich, T. C., Wilson, N. A., Hiraoka, A. and Watkins, D. I. (2009). "Gag- and Nef-specific CD4+ T cells recognize and inhibit SIV replication in infected macrophages early after infection." Proc Natl Acad Sci U S A **106**(24): 9791-6.

Salazar-Gonzalez, J. F., Bailes, E., Pham, K. T., Salazar, M. G., Guffey, M. B., Keele, B. F., Derdeyn, C. A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J. A., Swanstrom, R., Haynes, B. F., Athreya, G. S., Korber, B. T., Sharp, P. M., Shaw, G. M. and Hahn, B. H. (2008). "Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing." J Virol **82**(8): 3952-70.

Salter, R. D. and Parham, P. (1989). "Mutually exclusive public epitopes of HLA-A,B,C molecules." Hum Immunol **26**(2): 85-9.

Sarafianos, S. G., Marchand, B., Das, K., Himmel, D. M., Parniak, M. A., Hughes, S. H. and Arnold, E. (2009). "Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition." J Mol Biol **385**(3): 693-713.

Sattentau, Q. (2008). "Avoiding the void: cell-to-cell spread of human viruses." Nat Rev Microbiol **6**(11): 815-26.

Sattentau, Q. (2008). "HIV's gut feeling." Nat Immunol **9**(3): 225-7.

Schacker, T. W., Hughes, J. P., Shea, T., Coombs, R. W. and Corey, L. (1998). "Biological and virologic characteristics of primary HIV infection." Ann Intern Med **128**(8): 613-20.

Schellens, I. M., Borghans, J. A., Jansen, C. A., De Cuyper, I. M., Geskus, R. B., van Baarle, D. and Miedema, F. (2008). "Abundance of early functional HIV-specific CD8+ T cells does not predict AIDS-free survival time." PLoS ONE **3**(7): e2745.

Schmitt, A. O. and Herzel, H. (1997). "Estimating the entropy of DNA sequences." J Theor Biol **188**(3): 369-77.

Schmitz, J. E., Kuroda, M. J., Santra, S., Sasseville, V. G., Simon, M. A., Lifton, M. A., Racz, P., Tenner-Racz, K., Dalesandro, M., Scallon, B. J., Ghrayeb, J., Forman, M. A., Montefiori, D. C., Rieber, E. P., Letvin, N. L. and Reimann, K. A. (1999). "Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes." Science **283**(5403): 857-60.

Schneidewind, A., Brockman, M. A., Sidney, J., Wang, Y. E., Chen, H., Suscovich, T. J., Li, B., Adam, R. I., Allgaier, R. L., Mothe, B. R., Kuntzen, T., Oniangue-Ndza, C., Trocha, A., Yu, X. G., Brander, C., Sette, A., Walker, B. D. and Allen, T. M. (2008). "Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid." J Virol **82**(11): 5594-605.

Schneidewind, A., Brockman, M. A., Yang, R., Adam, R. I., Li, B., Le Gall, S., Rinaldo, C. R., Craggs, S. L., Allgaier, R. L., Power, K. A., Kuntzen, T., Tung, C. S., LaBute, M. X., Mueller, S. M., Harrer, T., McMichael, A. J., Goulder, P. J., Aiken, C., Brander, C., Kelleher, A. D. and Allen, T. M. (2007). "Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication." J Virol **81**(22): 12382-93.

Schneidewind, A., Tang, Y., Brockman, M. A., Ryland, E. G., Dunkley-Thompson, J., Steel-Duncan, J. C., St John, M. A., Conrad, J. A., Kalams, S. A., Noel, F., Allen, T. M., Christie, C. D. and Feeney, M. E. (2009). "Maternal transmission of HIV escape mutations subverts HLA-B57 immunodominance but facilitates viral control in the haploidentical infant." J Virol **83**(17): 8616-27.

Seki, S., Kawada, M., Takeda, A., Igarashi, H., Sata, T. and Matano, T. (2008). "Transmission of simian immunodeficiency virus carrying multiple cytotoxic T-lymphocyte escape mutations with diminished replicative ability can result in AIDS progression in rhesus macaques." J Virol **82**(10): 5093-8.

Serwold, T., Gaw, S. and Shastri, N. (2001). "ER aminopeptidases generate a unique pool of peptides for MHC class I molecules." Nat Immunol **2**(7): 644-51.

Serwold, T., Gonzalez, F., Kim, J., Jacob, R. and Shastri, N. (2002). "ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum." Nature **419**(6906): 480-3.

Sewell, A. K., Booth, B. L., Jr., Cerundolo, V., Phillips, R. E. and Price, D. A. (2002). "Differential processing of HLA A2-restricted HIV type 1 cytotoxic T lymphocyte epitopes." Viral Immunol **15**(1): 193-6.

Shankar, P., Sprang, H. and Lieberman, J. (1998). "Effective lysis of HIV-1-infected primary CD4+ T cells by a cytotoxic T-lymphocyte clone directed against a novel A2-restricted reverse-transcriptase epitope." J Acquir Immune Defic Syndr Hum Retrovirol **19**(2): 111-20.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." Bell Systems Technology **17**: 379-424.

Shastri, N., Cardinaud, S., Schwab, S. R., Serwold, T. and Kunisawa, J. (2005). "All the peptides that fit: the beginning, the middle, and the end of the MHC class I antigen-processing pathway." Immunol Rev **207**: 31-41.

Sheehy, A. M., Gaddis, N. C., Choi, J. D. and Malim, M. H. (2002). "Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein." Nature **418**(6898): 646-50.

Shiga, H., Shioda, T., Tomiyama, H., Takamiya, Y., Oka, S., Kimura, S., Yamaguchi, Y., Gojoubori, T., Rammensee, H. G., Miwa, K. and Takiguchi, M. (1996). "Identification of multiple HIV-1 cytotoxic T-cell epitopes presented by human leukocyte antigen B35 molecules." Aids **10**(10): 1075-83.

Shriner, D., Shankarappa, R., Jensen, M. A., Nickle, D. C., Mittler, J. E., Margolick, J. B. and Mullins, J. I. (2004). "Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection." Genetics **166**(3): 1155-64.

Shugars, D. C., Smith, M. S., Glueck, D. H., Nantermet, P. V., Seillier-Moiseiwitsch, F. and Swanstrom, R. (1993). "Analysis of human immunodeficiency virus type 1 nef gene sequences present in vivo." J Virol **67**(8): 4639-50.

Sidney, J., Peters, B., Frahm, N., Brander, C. and Sette, A. (2008). "HLA class I supertypes: a revised and updated classification." BMC Immunol **9**: 1.

Smith, D. B., McAllister, J., Casino, C. and Simmonds, P. (1997). "Virus 'quasispecies': making a mountain out of a molehill?" J Gen Virol **78 (Pt 7)**: 1511-9.

Smith, K. J., Reid, S. W., Stuart, D. I., McMichael, A. J., Jones, E. Y. and Bell, J. I. (1996). "An altered position of the alpha 2 helix of MHC class I is revealed by the crystal structure of HLA-B*3501." Immunity **4**(3): 203-13.

Smith, M. W., Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Lomb, D. A., Goedert, J. J., O'Brien, T. R., Jacobson, L. P., Kaslow, R., Buchbinder, S., Vittinghoff, E., Vlahov, D., Hoots, K., Hilgartner, M. W. and O'Brien, S. J. (1997). "Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study." Science **277**(5328): 959-65.

Sowinski, S., Jolly, C., Berninghausen, O., Purbhoo, M. A., Chauveau, A., Kohler, K., Oddos, S., Eissmann, P., Brodsky, F. M., Hopkins, C., Onfelt, B., Sattentau, Q. and Davis, D. M. (2008). "Membrane nanotubes physically connect T cells over long distances presenting a novel route for HIV-1 transmission." Nat Cell Biol **10**(2): 211-9.

Spira, A. I., Marx, P. A., Patterson, B. K., Mahoney, J., Koup, R. A., Wolinsky, S. M. and Ho, D. D. (1996). "Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques." J Exp Med **183**(1): 215-25.

Spira, S., Wainberg, M. A., Loemba, H., Turner, D. and Brenner, B. G. (2003). "Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance." J Antimicrob Chemother **51**(2): 229-40.

Steinbrook, R. (2007). "One step forward, two steps back--will there ever be an AIDS vaccine?" N Engl J Med **357**(26): 2653-5.

Stewart-Jones, G. B., di Gleria, K., Kollnberger, S., McMichael, A. J., Jones, E. Y. and Bowness, P. (2005). "Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705." Eur J Immunol **35**(2): 341-51.

Stewart-Jones, G. B., Gillespie, G., Overton, I. M., Kaul, R., Roche, P., McMichael, A. J., Rowland-Jones, S. and Jones, E. Y. (2005). "Structures of three HIV-1 HLA-B*5703-peptide complexes and identification of related HLAs potentially associated with long-term nonprogression." J Immunol **175**(4): 2459-68.

Storey, J. D. (2003). "The false positive discovery rate: a Bayesian interpretation and the q-value." Annals of Statistics **31**(6): 2013-2035.

Storey, J. D. and Tibshirani, R. (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci U S A **100**(16): 9440-5.

Streeck, H., Brumme, Z. L., Anastario, M., Cohen, K. W., Jolin, J. S., Meier, A., Brumme, C. J., Rosenberg, E. S., Alter, G., Allen, T. M., Walker, B. D. and Altfeld, M. (2008). "Antigen Load and Viral Sequence Diversification Determine the Functional Profile of HIV-1-Specific CD8(+) T Cells." PLoS Med **5**(5): e100.

Streeck, H., Lichterfeld, M., Alter, G., Meier, A., Teigen, N., Yassine-Diab, B., Sidhu, H. K., Little, S., Kelleher, A., Routy, J. P., Rosenberg, E. S., Sekaly, R. P., Walker, B. D. and Altfeld, M. (2007). "Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles." J Virol **81**(14): 7725-31.

Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P. and Sodroski, J. (2004). "The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys." Nature **427**(6977): 848-53.

Sun, Z. Y., Oh, K. J., Kim, M., Yu, J., Brusic, V., Song, L., Qiao, Z., Wang, J. H., Wagner, G. and Reinherz, E. L. (2008). "HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane." Immunity **28**(1): 52-63.

Sundquist, W. I. and Hill, C. P. (2007). "How to assemble a capsid." Cell **131**(1): 17-9.

Thobakgale, C. F., Prendergast, A., Crawford, H., Mkhwanazi, N., Ramduth, D., Reddy, S., Molina, C., Mncube, Z., Leslie, A., Prado, J., Chonco, F., Mphatshwe, W., Tudor-Williams, G., Jeena, P., Blanckenberg, N., Dong, K., Kiepiela, P., Coovadia, H., Ndung'u, T., Walker, B. D. and Goulder, P. J. (2009). "Impact of Hla in Mother and Child on Paediatric Hiv-1 Disease Progression." J Virol.

Thomson, M. M., de Parga, E. V., Vinogradova, A., Sierra, M., Yakovlev, A., Rakhmanova, A., Delgado, E., Casado, G., Munoz, M., Carmona, R., Vega, Y., Perez-Alvarez, L., Contreras, G., Medrano, L., Osmanov, S. and Najera, R. (2007). "New insights into the origin of the HIV type 1 subtype A epidemic in former soviet union's countries derived from sequence analyses of preepidemically transmitted viruses." AIDS Res Hum Retroviruses **23**(12): 1599-604.

Tindall, B., Cooper, D. A., Donovan, B., Barnes, T., Philpot, C. R., Gold, J. and Penny, R. (1988). "The Sydney AIDS Project: development of acquired immunodeficiency syndrome in a group of HIV seropositive homosexual men." Aust N Z J Med **18**(1): 8-15.

Tolstrup, M., Laursen, A. L., Gerstoft, J., Pedersen, F. S., Ostergaard, L. and Duch, M. (2006). "Cysteine 138 mutation in HIV-1 Nef from patients with delayed disease progression." Sex Health **3**(4): 281-6.

Tomiyama, H., Sakaguchi, T., Miwa, K., Oka, S., Iwamoto, A., Kaneko, Y. and Takiguchi, M. (1999). "Identification of multiple HIV-1 CTL epitopes presented by HLA-B*5101 molecules." Hum Immunol **60**(3): 177-86.

Tomiyama, H., Yamada, N., Komatsu, H., Hirayama, K. and Takiguchi, M. (2000). "A single CTL clone can recognize a naturally processed HIV-1 epitope presented by two different HLA class I molecules." Eur J Immunol **30**(9): 2521-30.

Towers, G. J. (2007). "The control of viral infection by tripartite motif proteins and cyclophilin A." Retrovirology **4**: 40.

Towers, G. J., Hatziioannou, T., Cowan, S., Goff, S. P., Luban, J. and Bieniasz, P. D. (2003). "Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors." Nat Med **9**(9): 1138-43.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T. B., Hraber, P. T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y. S., Kunstman, K., Wu, S., Phair, J., Erlich, H. and Wolinsky, S. (2003). "Advantage of rare HLA supertype in HIV disease progression." Nat Med **9**(7): 928-35.

Tsai, L., Trunova, N., Gettie, A., Mohri, H., Bohm, R., Saifuddin, M. and Cheng-Mayer, C. (2007). "Efficient repeated low-dose intravaginal infection with X4 and R5 SHIVs in rhesus macaque: implications for HIV-1 transmission in humans." Virology **362**(1): 207-16.

Turner, S. J., Doherty, P. C., McCluskey, J. and Rossjohn, J. (2006). "Structural determinants of T-cell receptor bias in immunity." Nat Rev Immunol **6**(12): 883-94.

Ueno, T., Idegami, Y., Motozono, C., Oka, S. and Takiguchi, M. (2007). "Altering effects of antigenic variations in HIV-1 on antiviral effectiveness of HIV-specific CTLs." J Immunol **178**(9): 5513-23.

Ueno, T., Tomiyama, H. and Takiguchi, M. (2002). "Single T cell receptor-mediated recognition of an identical HIV-derived peptide presented by multiple HLA class I molecules." J Immunol **169**(9): 4961-9.

UNAIDS (2008). Report on the global AIDS epidemic.

van Harmelen, J., Williamson, C., Kim, B., Morris, L., Carr, J., Karim, S. S. and McCutchan, F. (2001). "Characterization of full-length HIV type 1 subtype C sequences from South Africa." AIDS Res Hum Retroviruses **17**(16): 1527-31.

Veazey, R. S., DeMaria, M., Chalifoux, L. V., Shvetz, D. E., Pauley, D. R., Knight, H. L., Rosenzweig, M., Johnson, R. P., Desrosiers, R. C. and Lackner, A. A. (1998). "Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection." Science **280**(5362): 427-31.

Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B. and Delaporte, E. (2000). "Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa." J Virol **74**(22): 10498-507.

Wain, L. V., Bailes, E., Bibollet-Ruche, F., Decker, J. M., Keele, B. F., Van Heuverswyn, F., Li, Y., Takehisa, J., Ngole, E. M., Shaw, G. M., Peeters, M., Hahn, B. H. and Sharp, P. M. (2007). "Adaptation of HIV-1 to its human host." Mol Biol Evol **24**(8): 1853-60.

Walker, B. D. and Burton, D. R. (2008). "Toward an AIDS vaccine." Science **320**(5877): 760-4.

Wang, Y. E., Li, B., Carlson, J. M., Streeck, H., Gladden, A. D., Goodman, R., Schneidewind, A., Power, K. A., Toth, I., Frahm, N., Alter, G., Brander, C., Carrington, M., Walker, B. D., Altfeld, M., Heckerman, D. and Allen, T. M. (2009). "Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1." J Virol **83**(4): 1845-55.

Watkins, D. I., Burton, D. R., Kallas, E. G., Moore, J. P. and Koff, W. C. (2008). "Nonhuman primate models and the failure of the Merck HIV-1 vaccine in humans." Nat Med **14**(6): 617-21.

Wawer, M. J., Gray, R. H., Sewankambo, N. K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., Nalugoda, F., Wabwire-Mangen, F., Meehan, M. P. and Quinn, T. C. (2005). "Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda." J Infect Dis **191**(9): 1403-9.

Weinberg, J. B., Matthews, T. J., Cullen, B. R. and Malim, M. H. (1991). "Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes." J Exp Med **174**(6): 1477-82.

Whitney, J. B. and Ruprecht, R. M. (2004). "Live attenuated HIV vaccines: pitfalls and prospects." Curr Opin Infect Dis **17**(1): 17-26.

Wilk, T., Gross, I., Gowen, B. E., Rutten, T., de Haas, F., Welker, R., Krausslich, H. G., Boulanger, P. and Fuller, S. D. (2001). "Organization of immature human immunodeficiency virus type 1." J Virol **75**(2): 759-71.

Wille-Reece, U., Flynn, B. J., Lore, K., Koup, R. A., Miles, A. P., Saul, A., Kedl, R. M., Mattapallil, J. J., Weiss, W. R., Roederer, M. and Seder, R. A. (2006). "Toll-like receptor agonists influence the magnitude and quality of memory T cell responses after prime-boost immunization in nonhuman primates." J Exp Med **203**(5): 1249-58.

Wilson, N. A., Reed, J., Napoe, G. S., Piaskowski, S., Szymanski, A., Furlott, J., Gonzalez, E. J., Yant, L. J., Maness, N. J., May, G. E., Soma, T., Reynolds, M. R., Rakasz, E., Rudersdorf, R., McDermott, A. B., O'Connor, D. H., Friedrich, T. C., Allison, D. B., Patki, A., Picker, L. J., Burton, D. R., Lin, J., Huang, L., Patel, D., Heindecker, G., Fan, J., Citron, M., Horton, M., Wang, F., Liang, X., Shiver, J. W., Casimiro, D. R. and Watkins, D. I. (2006). "Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic simian immunodeficiency virus SIVmac239." J Virol **80**(12): 5875-85.

Wolthers, K. C. and Miedema, F. (1998). "Telomeres and HIV-1 infection: in search of exhaustion." Trends Microbiol **6**(4): 144-7.

Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J. J., Kabongo, J. M., Kalengayi, R. M., Van Marck, E., Gilbert, M. T. and Wolinsky, S. M. (2008). "Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960." Nature **455**(7213): 661-4.

Wyand, M. S., Manson, K. H., Garcia-Moll, M., Montefiori, D. and Desrosiers, R. C. (1996). "Vaccine protection by a triple deletion mutant of simian immunodeficiency virus." J Virol **70**(6): 3724-33.

Yan, J., Parekh, V. V., Mendez-Fernandez, Y., Olivares-Villagomez, D., Dragovic, S., Hill, T., Roopenian, D. C., Joyce, S. and Van Kaer, L. (2006). "In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules." J Exp Med **203**(3): 647-59.

Yang, O. O., Kalams, S. A., Trocha, A., Cao, H., Luster, A., Johnson, R. P. and Walker, B. D. (1997). "Suppression of human immunodeficiency virus type 1 replication by CD8+ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms." J Virol **71**(4): 3120-8.

Yang, O. O., Sarkis, P. T., Ali, A., Harlow, J. D., Brander, C., Kalams, S. A. and Walker, B. D. (2003). "Determinant of HIV-1 mutational escape from cytotoxic T lymphocytes." J Exp Med **197**(10): 1365-75.

Yang, O. O., Sarkis, P. T., Trocha, A., Kalams, S. A., Johnson, R. P. and Walker, B. D. (2003). "Impacts of avidity and specificity on the antiviral efficiency of HIV-1-specific CTL." J Immunol **171**(7): 3718-24.

Yang, Z. (2000). "Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A." J Mol Evol **51**(5): 423-32.

Yant, L. J., Friedrich, T. C., Johnson, R. C., May, G. E., Maness, N. J., Enz, A. M., Lifson, J. D., O'Connor, D. H., Carrington, M. and Watkins, D. I. (2006). "The high-frequency major histocompatibility complex class I allele Mamu-B*17 is associated with control of simian immunodeficiency virus SIVmac239 replication." J Virol **80**(10): 5074-7.

Yokomaku, Y., Miura, H., Tomiyama, H., Kawana-Tachikawa, A., Takiguchi, M., Kojima, A., Nagai, Y., Iwamoto, A., Matsuda, Z. and Ariyoshi, K. (2004). "Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection." J Virol **78**(3): 1324-32.

Zhou, T., Xu, L., Dey, B., Hessell, A. J., Van Ryk, D., Xiang, S. H., Yang, X., Zhang, M. Y., Zwick, M. B., Arthos, J., Burton, D. R., Dimitrov, D. S., Sodroski, J., Wyatt, R., Nabel, G. J. and Kwong, P. D. (2007). "Structural definition of a conserved neutralization epitope on HIV-1 gp120." Nature **445**(7129): 732-7.

Zimbwa, P., Milicic, A., Frater, J., Scriba, T. J., Willis, A., Goulder, P. J., Pillay, T., Gunthard, H., Weber, J. N., Zhang, H. T. and Phillips, R. E. (2007). "Precise identification of a human immunodeficiency virus type 1 antigen processing mutant." J Virol **81**(4): 2031-8.

Zuniga, R., Lucchetti, A., Galvan, P., Sanchez, S., Sanchez, C., Hernandez, A., Sanchez, H., Frahm, N., Linde, C. H., Hewitt, H. S., Hildebrand, W., Altfeld, M., Allen, T. M., Walker, B. D., Korber, B. T., Leitner, T., Sanchez, J. and Brander, C. (2006). "Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control." J Virol **80**(6): 3122-5.

APPENDIX I

HIV-1 sequences for *gag*, *pol* and *nef* genes. C-clade consensus and Durban consensus compared to B-clade reference strain HXB2. Durban consensus generated from taxa from a cohort of 710 treatment-naïve subjects. '.' indicates no change in consensus compared to HXB2. '-' indicates no amino acid at this position.

	10	20	30	40
HXB2 Gag p17	MGARASVLSG	GELDRWEKIR	LRPGGKKKYK	LKHIVWASRE
C clade consensus	......I.R.	.K..K.....	.......H.M	...L......
Durban consensus	......I.R.	EK..K.....	.......H.M	...L......

	50	60	70	80
HXB2 Gag p17	LERFAVNPGL	LETSEGCRQI	LGQLQPSLQT	GSEELRSLYN
C clade consensus	.....L....	.......K..	IK....A...	.T........
Durban consensus	.....L....	.......K..	IK....A...	.T...K..F.

	90	100	110	120
HXB2 Gag p17	TVATLYCVHQ	RIEIKDTKEA	LDKIEEEQNK	SKKKAQQAAA
C clade consensus	.........E	K..VR.....	....	.QQ.T...K.
Durban consensus	.........E	K..VR.....	....	.QQ.T..K..

	130	
HXB2 Gag p17	DTGHSNQVSQ	NY
C clade consensus	AD.---K...	..
Durban consensus	.GAD.GK...	..

	10	20	30	40
HXB2 Gag p24	PIVQNIQGQM	VHQAISPRTL	NAWVKVVEEK	AFSPEVIPMF
C clade consensus	.....L....	....	......I...	....
Durban consensus	.....L....	....	......I...	....

	50	60	70	80
HXB2 Gag p24	SALSEGATPQ	DLNTMLNTVG	GHQAAMQMLK	ETINEEAAEW
C clade consensus	T.........	....	....	D.........
Durban consensus	T.........	....	....	D.........

	90	100	110	120
HXB2 Gag p24	DRVHPVHAGP	IAPGQMREPR	GSDIAGTTST	LQEQIGWMTN
C clade consensus	..L.......	....	....	.....A...S
Durban consensus	..L.......	....	....	.....A...S

	130	140	150	160
HXB2 Gag p24	NPPIPVGEIY	KRWIILGLNK	IVRMYSPTSI	LDIRQGPKEP
C clade consensus	.......D..	....	.......V..	...K......
Durban consensus	.......D..	....	.......V..	....

	170	180	190	200
HXB2 Gag p24	FRDYVDRFYK	TLRAEQASQE	VKNWMTETLL	VQNANPDCKT
C clade consensus	........F.	.......T.D	......D...	....
Durban consensus	........F.	.......T.D	......D...	....

	210	220	230	
HXB2 Gag p24	ILKALGPAAT	LEEMMTACQG	VGGPGHKARV	L
C clade consensus	..R....G..	....	....S.....	.
Durban consensus	..R....G..	....	....	.

```
                          10         20         30         40
HXB2 Gag p15         AEAMSQVTNS ATIMMQRGNF RNQRKIVKCF NCGKEGHTAR
C clade consensus    ......AN.T -N.....S.. KGPKR..... .......I..
Durban consensus     ......AN.T SN.....S.. KGPKR..... .......I..

                          50         60         70         80
HXB2 Gag p15         NCRAPRKKGC WKCGKEGHQM KDCTERQANF LGKIWPSYKG
C clade consensus    .......... .......... .......... .......H..
Durban consensus     .......... .......... .......... .......H..

                          90         100        110        120
HXB2 Gag p15         RPGNFLQSRP EPTAPPEESF RSGVETTTPP QKQEPIDKEL
C clade consensus    .......N.. ......A... .F-E...PA. -....K.R.-
Durban consensus     .......N.. ......A... .F.E...PA. L....K.R.P

                          130
HXB2 Gag p15         YPLTSLRSLF GNDPSSQ
C clade consensus    -.....K... .S..L..
Durban consensus     ......K... .S..L..


                          10         20         30         40
HXB2 Protease        FFREDLAFLQ GKAREFSSEQ TRANSPTRRE LQVWGRDNNS
C clade consensus    ....N...P. .E....P... .......S.. ...--.GD.P
Durban consensus     ....N...P. QG....P... .......S.. .EREQ.REPA

                          50         60         70         80
HXB2 Protease        PSEAGADRQG TVSFNFPQVT LWQRPLVTIK IGGQLKEALL
C clade consensus    R.....E... --TL....I. .......S.. V...I.....
Durban consensus     TANSPTS-ES RSEL....I. .......S.. V...I.....

                          90         100        110        120
HXB2 Protease        DTGADDTVLE EMSLPGRWKP KMIGGIGGFI KVRQYDQILI
C clade consensus    .......... .IN...K... .......... ..........
Durban consensus     .......... .IN...K... .......... ..........

                          130        140        150
HXB2 Protease        EICGHKAIGT VLVGPTPVNI IGRNLLTQIG CTLNF
C clade consensus    ....K..... .......... ....M...L. .....
Durban consensus     ....K..... .......... ....M...L. .....


                          10         20         30         40
HXB2 RT              PISPIETVPV KLKPGMDGPK VKQWPLTEEK IKALVEICTE
C clade consensus    .......... .......... .......... ....TA..E.
Durban consensus     .......... .......... .......... ....TA..E.

                          50         60         70         80
HXB2 RT              MEKEGKISKI GPENPYNTPV FAIKKKDSTK WRKLVDFREL
C clade consensus    .......T.. .......... .......... ..........
Durban consensus     .......T.. .........I .......... ..........

                          90         100        110        120
HXB2 RT              NKRTQDFWEV QLGIPHPAGL KKKKSVTVLD VGDAYFSVPL
C clade consensus    .......... .......... .......... ..........
Durban consensus     .......... .......... .......... ..........

                          130        140        150        160
HXB2 RT              DEDFRKYTAF TIPSINNETP GIRYQYNVLP QGWKGSPAIF
C clade consensus    ..G....... .......... .......... ..........
Durban consensus     ..G....... .......... .......... ..........
```

```
                          170        180        190        200
HXB2 RT                   QSSMTKILEP FRKQNPDIVI YQYMDDLYVG SDLEIGQHRT
C clade consensus         .......... ..A...E... .......... .........A
Durban consensus          .......... ..AK..E... .......... .........A

                          210        220        230        240
HXB2 RT                   KIEELRQHLL RWGLTTPDKK HQKEPPFLWM GYELHPDKWT
C clade consensus         ......E... K..F...... .......... ..........
Durban consensus          ......E... K..F...... .......... ..........

                          250        260        270        280
HXB2 RT                   VQPIVLPEKD SWTVNDIQKL VGKLNWASQI YPGIKVRQLC
C clade consensus         ....Q..... .......... .......... ..........
Durban consensus          ....Q..... .......... .......... ..........

                          290        300        310        320
HXB2 RT                   KLLRGTKALT EVIPLTEEAE LELAENREIL KEPVHGVYYD
C clade consensus         .....A.... DIV....... .......... ..........
Durban consensus          .....A.... DIV....... .......... ..........

                          330        340        350        360
HXB2 RT                   PSKDLIAEIQ KQGQGQWTYQ IYQEPFKNLK TGKYARMRGA
C clade consensus         .......... ...HD..... .......... .....K..T.
Durban consensus          .......... ...HD..... .......... .....K..T.

                          370        380        390        400
HXB2 RT                   HTNDVKQLTE AVQKITTESI VIWGKTPKFK LPIQKETWET
C clade consensus         .......... .....AM... .........R ..........
Durban consensus          .......... .....AL... .........R ..........

                          410        420        430        440
HXB2 RT                   WWTEYWQATW IPEWEFVNTP PLVKLWYQLE KEPIVGAETF
C clade consensus         ...D...... .......... .......... ....A.....
Durban consensus          ...D...... .......... .......... ....A.....

                          450        460        470        480
HXB2 RT                   YVDGAANRET KLGKAGYVTN RGRQKVVTLT DTTNQKTELQ
C clade consensus         .......... .I.......D .....I.S.. E.........
Durban consensus          .......... .I.......D .....I.S.. E.........

                          490        500        510        520
HXB2 RT                   AIYLALQDSG LEVNIVTDSQ YALGIIQAQP DQSESELVNQ
C clade consensus         ..Q....... S......... .......... .K........
Durban consensus          ..Q....... S......... .......... .K........

                        1 530        540        550        560
HXB2 RT                   IIEQLIKKEK VYLAWVPAHK GIGGNEQVDK LVSAGIRKVL
C clade consensus         .........R ...S...... .......... ...S......
Durban consensus          .........R ...S...... .......... ...S......


                          10         20         30         40
HXB2 Integrase            FLDGIDKAQD EHEKYHSNWR AMASDFNLPP VVAKEIVASC
C clade consensus         .........E .......... ....E..... I.........
Durban consensus          .........E .......... ....E..... I.........

                          50         60         70         80
HXB2 Integrase            DKCQLKGEAM HGQVDCSPGI WQLDCTHLEG KVILVAVHVA
C clade consensus         .........I .......... .......... .I........
Durban consensus          .........I .......... .......... ..........
```

```
                          90         100        110        120
HXB2 Integrase     SGYIEAEVIP AETGQETAYF LLKLAGRWPV KTIHTDNGSN
C clade consensus  .......... .........Y I......... .V........
Durban consensus   .......... .........Y I......... .V........

                         130        140        150        160
HXB2 Integrase     FTGATVRAAC WWAGIKQEFG IPYNPQSQGV VESMNKELKK
C clade consensus  ..S.A.K... .....Q.... .......... ..........
Durban consensus   ..S.A.K... .....Q.... .......... ..........

                         170        180        190        200
HXB2 Integrase     IIGQVRDQAE HLKTAVQMAV FIHNFKRKGG IGGYSAGERI
C clade consensus  .......... .......... .......... ..........
Durban consensus   .......... .......... .......... ..........

                         210        220        230        240
HXB2 Integrase     VDIIATDIQT KELQKQITKI QNFRVYYRDS RNPLWKGPAK
C clade consensus  I......... .......I.. .......... .D.I......
Durban consensus   I......... .......I.. .......... .D.I......

                         250        260        270        280
HXB2 Integrase     LLWKGEGAVV IQDNSDIKVV PRRKAKIIRD YGKQMAGDDC
C clade consensus  .......... .......... ........K. .......A..
Durban consensus   .......... .......... ....V...K. .......A..

HXB2 Integrase     VASRQDED
C clade consensus  ..G.....
Durban consensus   ..G.....
```

```
                          10         20         30         40
HXB2 Nef           MGGKWSKSSV IGWPTVRERM RRAEPAADRV GAASRDLEKH
C clade consensus  .........I V...A....I ..T....EG. ....Q..D..
Durban consensus   .........I V...A....I ..T....E.. ....Q..D..

                          50         60         70         80
HXB2 Nef           GAITSSNTAA TNAACAWLEA QEEEEVGFPV TPQVPLRPMT
C clade consensus  ..L......T N..D...... .......... R.........
Durban consensus   ..L......T N..D....Q. ....G..... R.........

                          90         100        110        120
HXB2 Nef           YKAAVDLSHF LKEKGGLEGL IHSQRRQDIL DLWIYHTQGY
C clade consensus  ....F...F. .......... .Y.KK..E.. ...V......
Durban consensus   ....F...F. .......... .Y.KK..E.. ...V.....F

                         130        140        150        160
HXB2 Nef           FPDQNYTPGP GVRYPLTFGW CYKLVPVEPD KIEEANKGEN
C clade consensus  .......... .......... .F.....D.R EV....E...
Durban consensus   .......... .......... .F.....D.R EV........

                         170        180        190        200
HXB2 Nef           TSLLHPVSLH GMDDPEREVL EWRFDSRLAF HHVARELHPE
C clade consensus  NC....M.Q. ..E.ED.... K.K...H..R R.M.......
Durban consensus   NC....M.Q. ..E.ED.... K.K...H..R R.L.......

HXB2 Nef           YFKNC
C clade consensus  .Y.D.
Durban consensus   .Y.D.
```

APPENDIX II

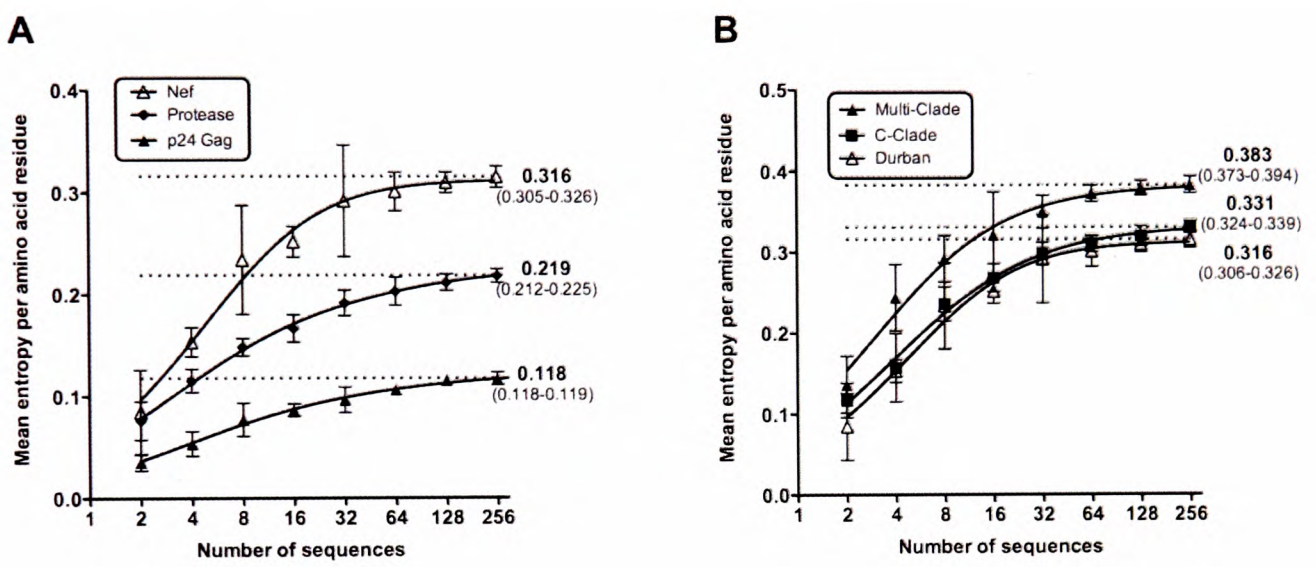
Pairs of HLA alleles transmitted in linkage disequilibrium in 710 HIV-infected subjects in Durban, South Africa. p values computed by Fisher's Exact Test; cut-off for statistical significance at $p<10^5$ according to Bonferroni correction.

HLA-B allele	HLA-A allele	p value	HLA-B allele	HLA-C allele	p value	HLA-B allele	HLA-C allele	p value
B0702	A2402	1.29E-29	B0702	C0102	8.60E-06	B4101	C17	8.06E-11
B0702	A3201	5.78E-06	B0702	C0401	5.86E-07	B4102	C17	1.41E-08
B0702	A6802	2.86E-09	B0702	C0602	1.45E-12	B4201	C0202	9.73E-07
B0801	A0301	8.20E-08	B0702	C0702	1.35E-141	B4201	C0210	2.44E-06
B0801	A3002	6.38E-06	B0702	C17	3.89E-09	B4201	C0304	9.91E-08
B1302	A2902	1.62E-17	B0705	C0702	3.51E-08	B4201	C0401	1.67E-12
B1401	A6802	7.14E-06	B0801	C0401	2.73E-09	B4201	C0602	1.42E-24
B1503	A4301	2.48E-15	B0801	C0602	9.56E-18	B4201	C0701	1.62E-19
B1503	A74	5.69E-25	B0801	C0701	9.10E-61	B4201	C0702	4.16E-09
B1510	A6802	9.20E-34	B0801	C0702	2.37E-10	B4201	C1601	9.72E-07
B1801	A8001	8.81E-25	B0801	C17	7.60E-12	B4201	C17	1.10E-307
B3501	A74	4.23E-23	B0801	C18	9.95E-06	B4201	C18	6.63E-06
B3910	A3002	5.00E-07	B1302	C0602	1.00E-33	B4202	C17	5.26E-28
B3910	A6601	1.11E-06	B1401	C0801	1.43E-09	B4403	C0303	1.64E-07
B4201	A3001	6.71E-60	B1401	C0802	1.49E-37	B4403	C0304	5.01E-06
B4201	A6602	1.53E-06	B1401	C0804	3.32E-10	B4403	C0401	1.00E-52
B4202	A3001	1.86E-21	B1402	C0802	1.49E-51	B4403	C0602	5.23E-19
B4403	A2902	8.45E-27	B1503	C0202	1.71E-25	B4403	C0701	5.47E-19
B4403	A3402	2.66E-40	B1503	C0210	4.51E-124	B4403	C17	4.47E-12
B4501	A0201	1.33E-25	B1503	C0304	4.24E-06	B4501	C0401	6.51E-06
B5101	A2601	8.67E-08	B1503	C0602	7.65E-21	B4501	C0701	1.97E-08
B5301	A3303	7.17E-11	B1503	C0701	2.93E-16	B4501	C1601	6.60E-88
B5301	A3601	5.84E-08	B1503	C0702	1.44E-08	B4501	C1604	1.35E-06
B5703	A0202	3.66E-06	B1503	C1601	8.18E-06	B4501	C17	2.17E-07
B5703	A74	1.73E-14	B1503	C17	2.78E-12	B4901	C0701	3.55E-19
B5801	A0205	7.38E-48	B1503	C18	3.62E-07	B5301	C0401	1.09E-41
B5802	A0301	1.91E-07	B1510	C0304	1.13E-101	B5702	C18	3.59E-11
B5802	A4301	2.86E-06	B1510	C0602	3.04E-23	B5703	C0701	9.12E-31
B5802	A6601	3.96E-37	B1510	C0701	3.87E-15	B5801	C0302	2.62E-18
B5802	A6801	1.07E-26	B1510	C0702	2.56E-07	B5801	C0401	5.61E-06
B5802	A6802	9.39E-06	B1510	C0801	2.17E-13	B5801	C0701	2.77E-34
B5802	A74	3.46E-06	B1510	C0804	1.72E-12	B5801	C17	2.38E-08
B8101	A0101	9.50E-76	B1510	C1601	1.01E-15	B5802	C0202	2.49E-10
			B1510	C17	2.37E-16	B5802	C0210	1.48E-09
			B1510	C18	1.74E-07	B5802	C0304	1.69E-10
			B1516	C1402	4.14E-08	B5802	C0401	4.19E-19
			B1516	C1601	2.56E-07	B5802	C0701	3.57E-25
			B1801	C0202	5.78E-21	B5802	C0702	2.19E-12
			B1801	C0501	1.20E-16	B5802	C1601	7.57E-11
			B1801	C0602	8.78E-07	B5802	C17	1.77E-22
			B1801	C0704	1.25E-38	B5802	C18	1.44E-10
			B3501	C0401	3.60E-42	B8101	C0401	1.55E-08
			B3910	C1203	4.48E-70	B8101	C0602	9.59E-11
			B3910	C1505	7.74E-08	B8101	C0701	5.04E-07
						B8101	C17	1.85E-08
						B8101	C18	5.62E-86

APPENDIX III

Relationship between number of sequences analysed and mean Shannon Entropy Score per amino acid residue.

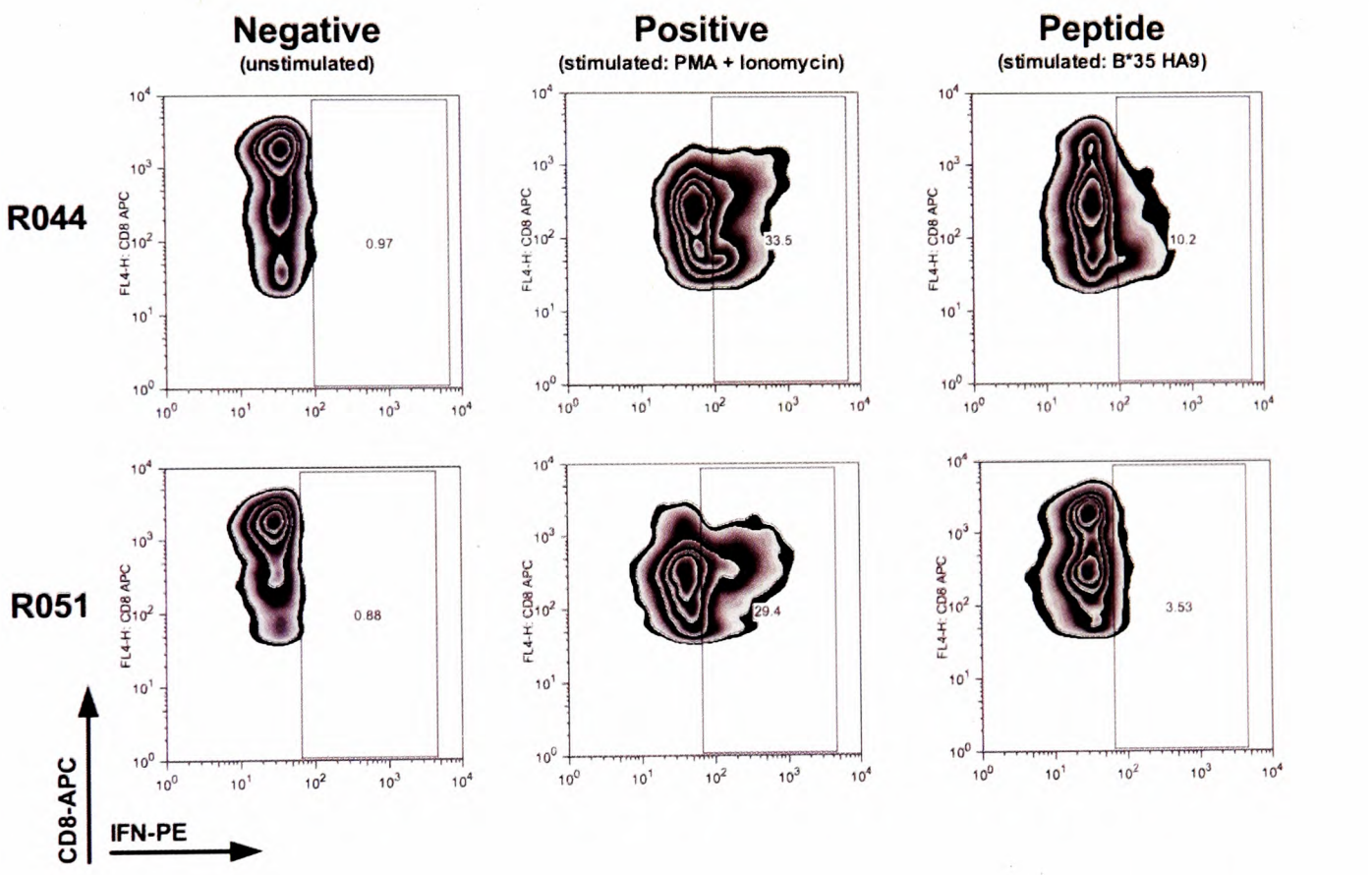
A: Three proteins from Durban cohort; B: Nef protein from three different cohorts. Mean ± 95% confidence intervals plotted for five repetitions of entropy calculations from randomised datasets. Nef C-clade and multi-clade taxa downloaded from Los Alamos HIV databases at www.hiv.lanl.gov.



APPENDIX IV

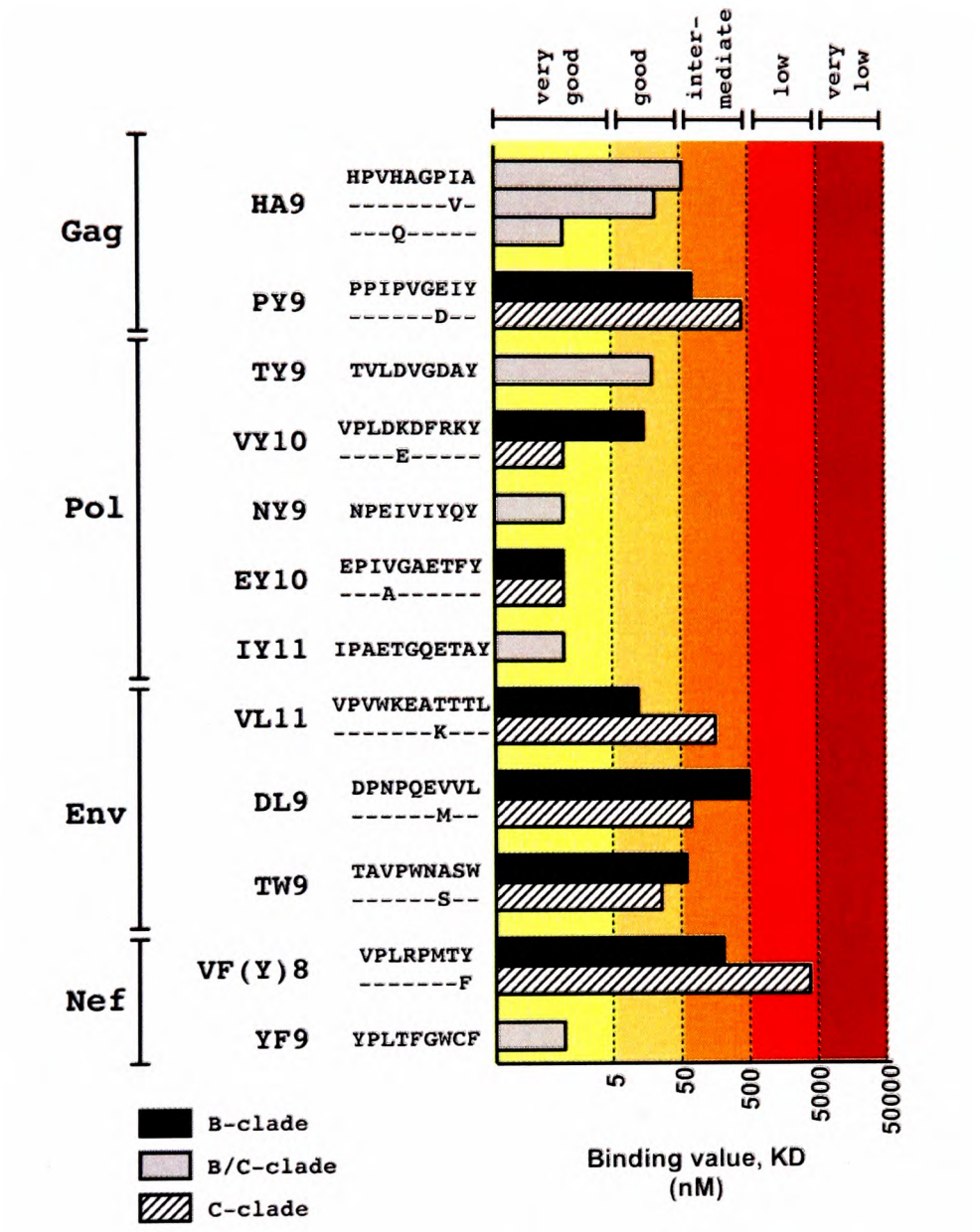
Intracellular staining of HLA-B*35 HA9 T cell lines derived from HIV-infected subjects R044 and R051 to identify IFN-γ production.

Staining was undertaken following no stimulation, stimulation with PMA/Ionomycin or stimulation with HLA-B*35 optimal peptide HA9 (HPVHAGPIA). 2.0-3.0x10^5 events collected for each plot. Gated on live $CD8^+$ lymphocyte population; gating for IFN-γ was set to achieve <1% of cells positive in the absence of stimulation.



APPENDIX V

Binding of HLA-B*35 optimal epitopes to the HLA-B*35 allele. Assay performed using luminescent oxygen channelling immunoassay (LOCI) technology.



APPENDIX VI

Supplementary methods for Chapter 7

$CD4^+$ T cell enrichment using negative selection by MACS

Marked variability in the success of $CD4^+$ T cell enrichment was observed among cells from HIV positive and negative study subjects (Fig. VI.I). Part of the explanation for this difference may be the altered CD4:CD8 ratio characteristically seen in HIV. Accelerated cell death in HIV-infected cells may also have contributed to poor enrichment in certain HIV study subjects; however, enrichment was also not significantly correlated to baseline CD4 count or viral load (p=0.1 and p=0.7 respectively, linear regression; data not shown).

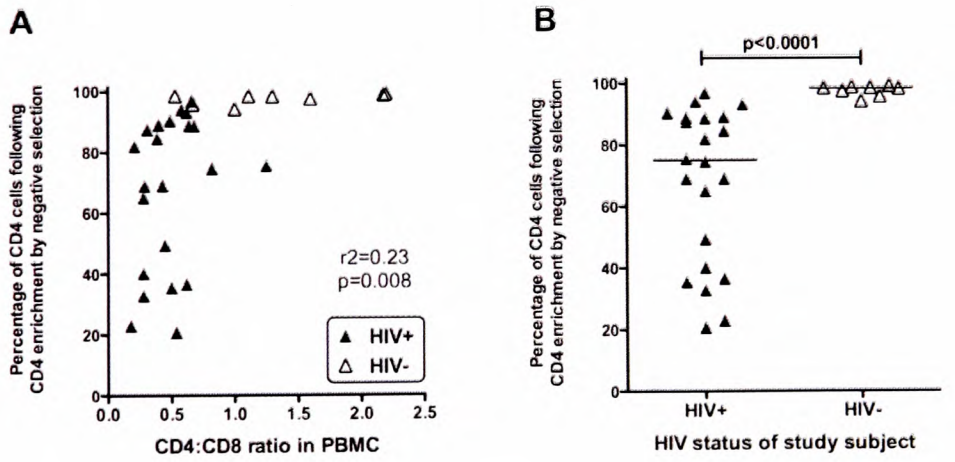


Figure VI.I: $CD4^+$ T cell enrichment by negative selection by MACS in subjects with and without HIV infection. Data generated by FACS analysis. A: Percentage of $CD4^+$ T cells in CD4-enriched fraction plotted against CD4:CD8 ratio in PBMCs (p=0.008, linear regression). B: Percentage of $CD4^+$ T cells in CD4-enriched fraction according to HIV status (p<0.0001, Mann Whitney test).

In order to optimise cell numbers and reduce this variability in the success of $CD4^+$ T cell enrichment, a final protocol was established whereby the $CD8^+$ T cell fraction was first enriched by positive selection from bulk PBMCs, and the remaining

unlabelled cells were used as the $CD4^+$ fraction of target cells. In order to account for the lowered CD4:CD8 ratio characteristically seen in HIV-infected subjects, the manufacturer's protocol for $CD8^+$ selection was optimized by adding extra CD8 beads (original recommendation 20ul beads in 100ul buffer per 10 million PBMC, increased to 30ul beads in the same volume).

Optimising depletion of $CD8^+$ T cells from CD4-enriched fraction

Following positive selection of $CD8^+$ T cells from PBMCs, the remaining unlabelled cells still contained $CD8^+$ T cells (characteristically 1-5% of total cell number). Two strategies were adopted to improve the success of $CD8^+$ depletion:

i. Following collection of the $CD4^+$ fraction, repeat depletion of $CD8^+$ T cells was undertaken using Dynabeads (Invitrogen, UK) (Fig. VI.II).

ii. A Miltenyi *MS* column was replaced with an *LD* column. These columns are optimized specifically for depletion, containing a larger surface area over which labelled cells may be retained, and successfully reduced the $CD8^+$ T cell content of the CD4-enriched fraction to <1% (Fig. VI.III).

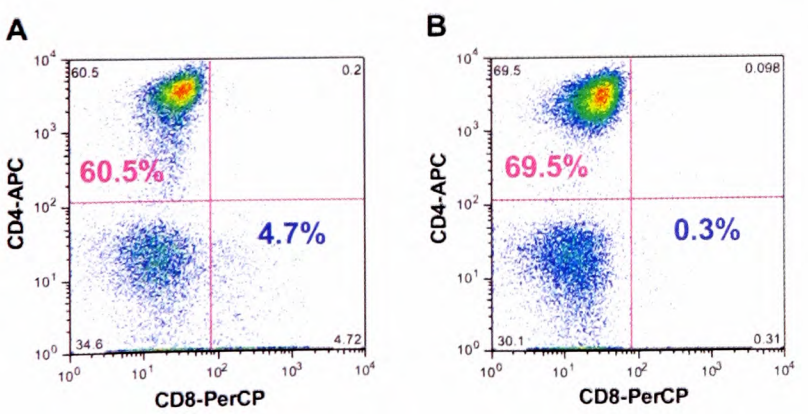


Figure VI.II: FACS plots to show $CD4^+$ T cell enrichment from PBMCs in an HIV-infected donor, subject N007, before and after repeat $CD8^+$ T cell depletion. A: CD4-enriched fraction following depletion of $CD8^+$ T cells by positive selection. B: Further enrichment of $CD4^+$ fraction using CD8-dynabeads.

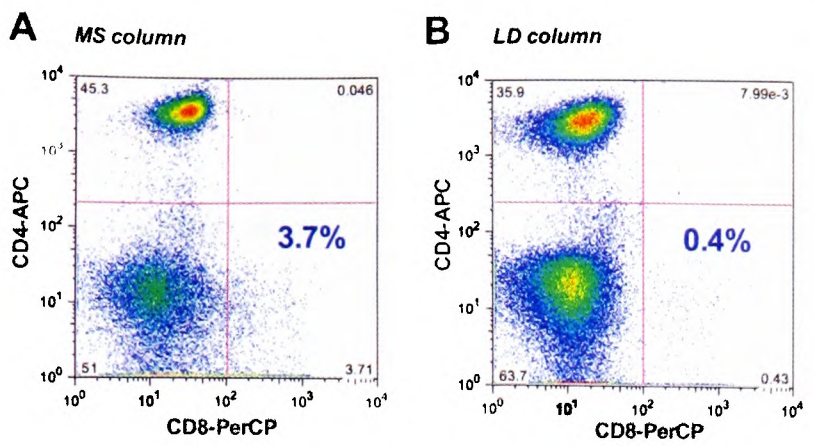


Figure VI.III: Typical FACS plots to show $CD4^+$ T cell fraction achieved by depletion of $CD8^+$ T cells by positive selection from PBMCs in HIV-infected donors using MS or LD columns. A: MS column to enrich $CD4^+$ cells by positive selection of $CD8^+$ population (subject N079). B: LD column to enrich $CD4^+$ T cells (subject LD011).

Reduction in number of cells required to perform assay

Initial experiments were performed with $CD4^+$ target cells resuspended at $1x10^6$ cells/ml and placed at $2x10^6$ cells/well in a 24 well plate. Subsequently, this was reduced to $1x10^5$ cells/well in a 96 well plate, in order to reduce the total number of cells required and facilitate repetitions. The only limitation of this approach is that a smaller volume of supernatant is available for subsequent p24 ELISA assays, altering the lower threshold for detection (see section 7.2.6).

Duration of cell culture prior to starting viral suppression assay

One of the aims of this experimental approach was to measure viral suppression as close to *ex vivo* as possible. Following enrichment, cells were left in culture for a maximum of 48-72 hours to allow activation and recovery from the enrichment process.

GFP-production at 7 days can be used to measure viral suppression

GFP was transcribed in-frame with Nef in this NL4-3 construct, and should thus be detectable in productively infected cells from 12-24 hours post-infection. However, despite the logarithmic increase seen in p24 antigen production, only small numbers

of the total cell population (median <1%) were seen to be GFP-positive even at seven days post-infection (Fig. VI.IV).

GFP production was significantly suppressed by $CD8^+$ T cells (p=0.03 paired t-test at E:T ratio 1:1), suggesting this assay could be a useful adjunct to measuring p24 antigen production. However, this level of statistical significance was only achieved at seven days. Based on the low levels of GFP production, lack of statistical significance at earlier time points, and the increased number of cells required to perform FACS analysis, further experiments measuring GFP were not undertaken.

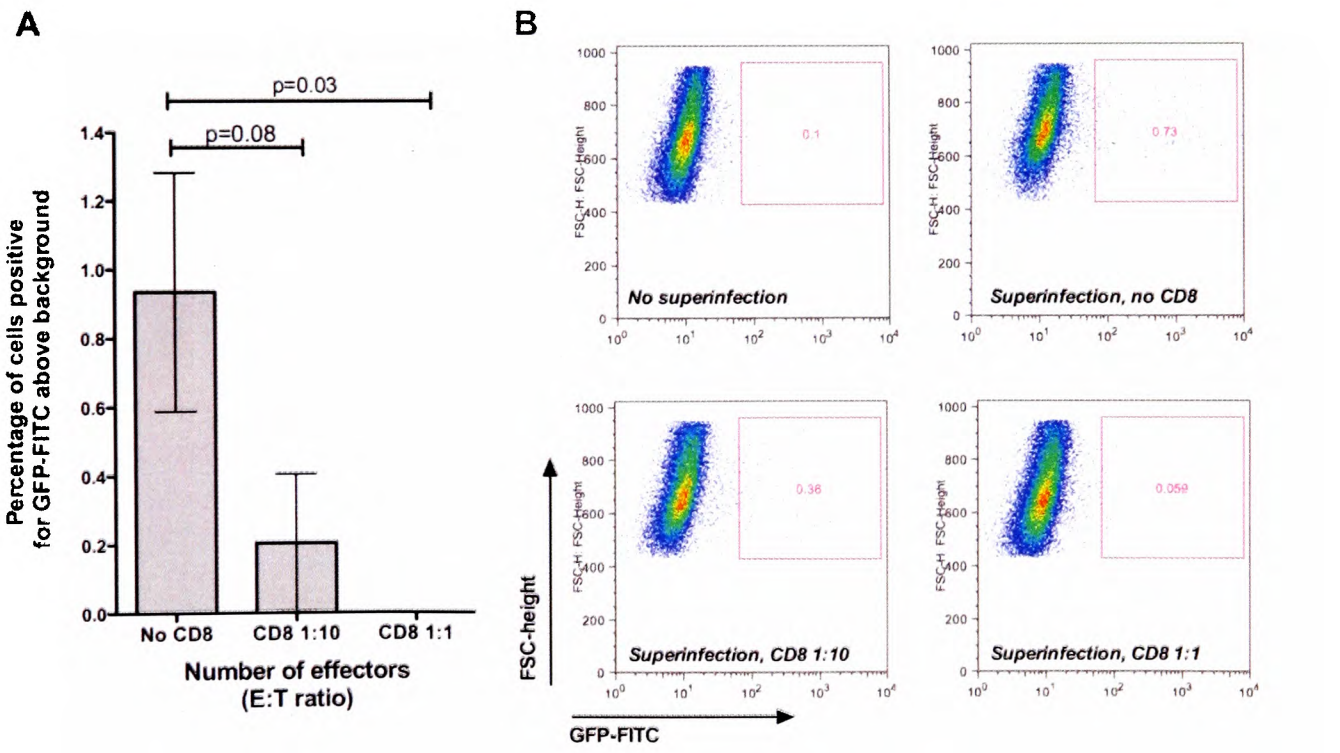


Figure VI.IV: Measurement of GFP at Day 7 following superinfection of CD4-enriched target cells in the presence and absence of $CD8^+$ effectors in 5 subjects. A: GFP-detection above background using mean value from triplicates in each patient; p values by paired t-test. B: Characteristic FACS data showing GFP-production in the absence of effector cells, reduced in the presence of effector cells at E:T ratio of 1:10, and to below background at E:T ratio of 1:1.