

https://dspace.org/

Host factors and broadly neutralizing antibodies in South African women infected with HIV-1 subtype C

Scheepers, Cathrine

2016-02-22
http://hdl.handle.net/10539/19680

Downloaded from a DSpace repository. https://dspace.org/

HOST FACTORS AND BROADLY NEUTRALIZING ANTIBODIES IN SOUTH AFRICAN WOMEN INFECTED WITH HIV-1 SUBTYPE C

Cathrine Scheepers

A thesis submitted to the Faculty of Health Sciences, School of Pathology,
University of the Witwatersrand in fulfilment of the degree of Doctor of Philosophy (PhD)

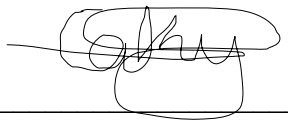
Johannesburg,

September 2015

PLAGIARISM DECLARATION

I, Cathrine Scheepers (Student number: 0112972N), am a student registered for Doctor of Philosophy in the year 2015. I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- The report generated by "Turnitin" for this thesis can be found in Appendix A.

Signature:  Date: 3 September 2015

ABSTRACT

Broadly neutralizing antibodies are capable of neutralizing a large number of HIV-1 strains and have shown to be protective against infection in non-human primate models. These antibodies are likely to play an important role in an effective vaccine against HIV. Eliciting them by vaccination has thus far been unsuccessful and their unusual features such as long CDHR3 lengths and high levels of somatic hypermutation make this a particularly challenging task. Approximately 15-30% of chronically HIV-1 infected individuals develop these types of antibodies but an understanding of the underlying mechanisms is limited. The aim of this study, therefore, was to investigate host factors associated with the development of broadly neutralizing antibodies in HIV-1 subtype C infected individuals. In particular we analysed genetic variation of the genes encoding the variable region of antibodies, the evolution of an HIV-1 specific antibody lineage and glycan-binding profiles of serum antibodies.

The human heavy chain variable region genes (IGHV) are the largest and most variable of all human immunoglobulin genes and encode the major antigen-binding region. These genes are divided into seven subgroups, each subgroup contains numerous genes and alleles. Using genomic DNA from 28 HIV-1 subtype C infected individuals we performed next generation sequencing using both Illumina MiSeq and Roche 454 technologies. Included were 13 individuals who developed broadly neutralizing antibodies, 13 who did not despite chronic HIV-1 infection and two intermediate neutralizers. We found no genetic differences in the IGHV genes between these two groups. However, we

identified 85 novel alleles and 38 alleles that had previously only been observed in rearranged antibody sequences. Of these alleles, eight were used by functional antibodies, two of which were HIV-1 specific. This study highlights the importance of a fully comprehensive database for inferring germline gene usage and the unmutated common ancestors of antibody lineages. In addition it showed that everyone has the same genetic potential of developing broadly neutralizing antibodies, which has positive implications for vaccine development.

A number of studies have demonstrated the importance of strain-specific antibodies in the development of broadly neutralizing antibodies. In addition to being the forerunners of broadly neutralizing antibodies, strain-specific antibodies can help shape the viral populations that elicit different broadly neutralizing antibody lineages. We therefore studied the evolution of a strain-specific HIV antibody lineage (CAP88-CH06) in an individual who failed to develop neutralization breadth even after 5 years of infection, to understand why some strain-specific antibodies remain limited to autologous viruses. CAP88-CH06 was previously isolated as an IgA1 antibody using the IGHV4-39 gene with 8.8% divergence from donor CAP88 at 34 weeks post-infection, which mapped to the C3-V4 region of gp120. IgA and IgG antibodies using the same germline IGHV were sequenced on an Illumina MiSeq from 5 to 121 weeks post-infection. IgA sequences identical to that of the fully matured antibody with 8.8% divergence were detected from early infection and throughout until after 2 years, well after viral escape. This was consistent with plasma neutralization of the C3-V4 region within CAP88. However, very little evidence of evolution was seen within the IgA sequences. A group of related IgG sequences were also identified between 11 and 34 weeks but not at other time-points. Interestingly, within the 11 week transcripts we identified an identical sequence as both an IgA and IgG isotype, which likely gave rise to these transient IgG antibodies. The lack of

neutralization breadth in this individual could therefore be the result of both limited evolution of the IgA isotype and well as the disappearance of the IgG isotype.

The HIV envelope is surrounded by glycans, known as the glycan shield. These glycans contribute towards the structural integrity of the envelope and serve as protection against immune responses to conserved regions. However, glycans often form targets for broadly neutralizing antibodies. Thus we studied the glycan-binding profiles of HIV-negative and HIV-positive individuals (including 12 individuals who develop broadly neutralizing antibodies and 13 who did not despite chronic infection) to determine whether glycan-binding was specific to individuals who develop broadly neutralizing antibodies. Longitudinal samples were taken yearly for three years from all 47 individuals and their serum IgG levels were tested on glycan microarrays. We observed fluctuations in glycan-binding over time within the HIV-negative individuals and these were used to establish baseline values. The HIV-positive individuals were found to have elevated levels of antibodies targeting high mannose N-linked glycans, Tn-peptides and glycolipids during infection. Binding to Tn-peptides and glycolipids were elevated throughout infection, whereas high mannose N-linked glycans were elevated from 2-3 years post-infection. We observed no differences in these glycans between the individuals who developed broadly neutralizing antibodies compared to those who did not despite chronic HIV infection. This data suggests that the elevated levels of glycan-binding serum antibodies were a consequence of infection rather than specific to broadly neutralizing antibodies. Since glycan-binding antibodies against Tn-peptides and glycolipids were detected earlier than high mannose N-linked glycans and antibodies targeting these glycans were elevated during infection, they might warrant further investigation with respect to immunogen design.

Collectively this study has contributed to a greater understanding of the role of various host factors in the development of broadly neutralizing antibodies to HIV. This includes showing that there were no differences in the IGHV genes between individuals who did and did not develop broadly neutralizing antibodies as well as providing a wealth of new data on human antibody genes that will have benefits beyond the field of HIV. Furthermore our study has reinforced the essential role of somatic hypermutation in developing neutralization breadth and the need for further co-evolution studies on strain-specific lineages to understand this roadblock. Finally our study using glycan arrays has highlighted that glycan-binding antibodies are induced in all HIV-infected individuals even though only a minority go on to develop broadly neutralizing antibodies. Overall these data suggest that all humans have the ability to develop broadly neutralizing antibodies but a vaccine capable of eliciting such protective responses remains a major hurdle.

ACKNOWLEDGMENTS

I would like to acknowledge the numerous funding agencies involved in this study:

The Poliomyelitis Research Foundation for both project and bursary funding (grant numbers: 11/15 and 11/45), the National Research Foundation for bursary funding (grant number: 79531), the WITS Health Sciences Faculty Research Committee for project funding (grant number: 00125484612015121105000000000000000004884) the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) through the Fogarty International Centre, National Institutes of Health (NIH) (grant number: 5 D43 TW000231) for funding my training at Columbia University in New York City, New York, USA and at the National Cancer Institute (NCI) part of the NIH, Frederick, Maryland, USA. Some of the 454 GS Junior sequencing runs were sponsored by Roche as part of a competition that we won.

I would like to acknowledge Centre for the AIDS Programme of Research in South Africa (CAPRISA) in KwaZulu-Natal, for the collection of samples used throughout this study as well as to the participants in the various CAPRISA cohorts. The NIH AIDS Reagent Program (National Institute of Allergy and Infectious Diseases part of the US National Institutes of Health, (NIH AIDS Reagent, 2013)) and the IAVI Center for Neutralizing antibodies (Scripps Research Institute, La Jolla, CA) from whom bNAbs were obtained.

I would like to acknowledge various people who contributed throughout my PhD: Firstly to Professor Lynn Morris who has been an amazing supervisor and great inspiration. It has been an honour working for you. Professor Penny Moore who has been a

fantastic leader and provided a lot of guidance. Dr Bronwen Lambson for her daily input into my projects and supervision throughout. To those involved in the germline IGHV repertoire study: Professor Simon Travers, Dr Ram Shrestha and Imogen Wright from the South African National Bioinformatics Institute (SANBI) and the University of the Western Cape and Dr Katherine Jackson from Stanford University in California, USA for their assistance in the data analysis of the next generation sequencing. Mark Goosen, Dshanta Naicker and Arshad Ismail from the National Institute for Communicable Diseases (NICD), Sandringham, Johannesburg, South Africa for their assistance in the next generation sequencing on the 454 GS Junior and Illumina MiSeq. To those involved in the strain-specific anti-HIV antibody evolution during infection study: Professor Lawrence Shapiro, Dr Zizhang Sheng and Dr Chaim Schramm from Columbia University, Department of Biochemistry and Molecular Biophysics, New York City, New York, USA for their assistance in the antibody sequencing data analysis. Arshad Ismail once again for assistance in the next generation sequencing on the Illumina MiSeq. To those involved in the induction of glycan-binding antibodies during infection study: Dr Jeff Gildersleeve, Dr Christopher Campbell, Dr Saddam Muthana, Sudipa Chowdhury and Shea Wright from the NCI, Frederick, Maryland, USA for their help with the glycan arrays. Dr Mashudu Madzivhandila and Tandile Hermanus from the NICD, for neutralization data. Dr Nono Mkhize and Dr Elin Gray for mAb sequences.

I would also like to thank my amazing husband, Jannie Scheepers, for all the encouragement and support throughout this process. To my family for their support and most importantly to God for giving me the perseverance to continue this process to completion.

PREFACE

The following publications and conference presentations arose from the work presented in this thesis.

1. PUBLICATIONS

Scheepers C, Shrestha R.K, Lambson B.E, Jackson K.J.L, Wright I.A, Naicker D, Goosen M, Berrie L, Ismail A. Garrett N, Abdool Karim Q, Abdool Karim S.S, Moore P.L, Travers S.A and Morris L. (2015) **Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline immunoglobulin gene repertoire.** *Journal of Immunology,* 194 (9) : 4371-4378

Scheepers C, Sheng Z, Schramm C, Ismail A, Lambson B.E, Garrett N, Abdool Karim S.S, Moore P.L, Shapiro L and Morris L. **Limited evolution of a potent strain-specific antibody lineage in an HIV-infected individual who fails to develop neutralization breadth**
Manuscript in preparation

Scheepers C, Chowdhury S, Wright S, Campbell C, Garrett N, Abdool Karim Q, Abdool Karim S.S, Moore P.L, Gildersleeve J.C. and Morris L. **Induction of glycan-binding antibodies during HIV-1 infection**
Manuscript in preparation

2. CONFERENCE PRESENTATIONS

Scheepers C, Naicker D, Sheng Z, Schramm C, Ismail A, Abdool Karim S.S, Lambson B.E, Moore P.L, Shapiro L and Morris L. (2014) **Strain-specific anti-HIV antibody evolution during acute infection and viral escape**. HIV R4P Conference, Cape Town, South Africa

Scheepers C, Naicker D, Schramm C, Sheng Z, Ismail A, Abdool Karim S.S, Lambson B.E, Moore P.L, Shapiro L and Morris L. (2014) **Strain-specific anti-HIV antibody evolution during acute infection and viral escape.** WITS Health Sciences Research Day, Johannesburg, South Africa

Mitchell (Scheepers) C, Shrestha R.K, Wright I.A, Goosen M, Berrie L, Abdool Karim S.S, Lambson B.L, Travers S.A and Morris L. (2013) **454 sequencing of germline IGHV genes in HIV-infected South African women.** SA AIDS Conference, Durban, South Africa

Mitchell (Scheepers) C, Shrestha R.K, Wright I.A, Goosen M, Berrie L, Abdool Karim S.S., Lambson B.L, Travers S.A and Morris L (2013) **454 sequencing of germline IGHV genes in HIV-infected South African women**. Keystone HIV Vaccines (X2) Conference, Colorado, USA

Mitchell (Scheepers) C, Campbell C, Muthana S, Chowdhury S, Lambson B.E, Moore P.L, Morris L, Gildersleeve J and the CAPRISA 002 cohort. (2012) **Induction of glycan-binding antibodies in HIV infection**. WITS Health Sciences Research Day, Johannesburg, South Africa.

ETHICS

Both the parent study involving the CAPRISA cohorts and this particular study were granted ethics clearance under the following certificate numbers: M080470 for the parent study and M111104 for this study. The ethics certificate for this study can be found in the Appendix.

LIST OF FIGURES

Section A

Figure 1: Estimated prevalence of HIV in adults and children worldwide in 2012.....3

Figure 2: Distribution of HIV-1 clades worldwide.....6

Figure 3: HIV-1 genome and viral particle structure.....7

Figure 4: Life cycle of HIV-1.....8

Figure 5: HIV-1 envelope trimer structure before and after fusing with CD4+ T-cells.....11

Figure 6: Reconstruction of the Env trimer structure bound by PGV04.....12

Figure 7: Glycan shield on the HIV-1 envelope trimer.....13

Figure 8: HIV viral load and CD4 count during infection.....15

Figure 9: B-cell dysfunction and subpopulations associated with HIV-1 infection.....16

Figure 10: Anti-HIV-1 antibody response during infection.....17

Figure 11: Broadly neutralizing antibody targets on the HIV-1 envelope trimer.....19

Figure 12: Germline heavy and light chain immunoglobulin gene loci.....22

Figure 13: Early B-cell development and VDJ recombination.....26

Figure 14: Somatic hypermutation and class-switch recombination in the germinal centre of lymph nodes.....28

Figure 15: Process of class-switch recombination.....29

Section B

Figure 1.1: Novel germline IGHV alleles in 28 South African women.....64

Figure 1.2: Number of single nucleotide changes (SNPs) in the novel and non-IMGT germline IGHV alleles.....65
Figure 1.3: Number of germline IGHV alleles for each individual.....65
Figure 1.4: Phylogenetic tree of all germline IGHV alleles observed in 28 South African individuals.....66
Figure 1.5: Novel IGHV alleles used by isolated mAbs and some bNAbs.....67
Figure 1.6: Comparison of germline IGHV genes in the repertoires of BCN and non-BCN individuals.....70
Figure 1.7: Comparison of the germline IGHV alleles in the repertoires of BCN and non-BCN individuals.....71
Figure 1.8: Comparison of bNAb germline IGHV alleles between BCN and non-BCN individuals.....72
Figure 2.1: Plasma neutralization of the C3-V4 region of CAP88 Env during HIV-1 infection.....93
Figure 2.2: Identity-divergence plots for CAP88-CH06 IgA and IgG plasmid controls.....95
Figure 2.3: Identity-divergence plots of CAP88 cDNA sequences over 2 years of infection.....97
Figure 2.4: Phylogenetic tree using clonally related IgA sequences.....101
Figure 2.5: Phylogenetic tree using clonally related IgG sequences.....102
Figure 2.6: Phylogenetic tree of all clonally related sequences to CAP88-CH06.....103
Figure 2.7: Alignment of CAP88-CH06 clonally related sequences.....105
Figure 3.1: bNAb glycan-binding profiles.....127
Figure 3.2: IgG antibody profiles within individuals.....128

Figure 3.3: Fluctuations in anti-glycan IgG binding in HIV-negative longitudinal serum.....130
Figure 3.4: Difference in log2 binding for PBS and gp120-incubated sera.....134
Figure 3.5: Longitudinal binding profiles of glycans that are elevated with the duration of HIV-1 infection.....140
Figure 3.6: Longitudinal binding of glycans that were elevated throughout HIV-1 infection.....141
Figure 3.7: Glycans with significantly lower binding in the BCN individuals compared to the non-BCN individuals.....143

Appendix

Supplementary Figure 1.1: Nucleotide sequence alignment for IGHV3-11*5mm.....173
Supplementary Figure 1.2: Nucleotide sequence alignment for IGHV4-59*8mm1.....173
Supplementary Figure 1.3: Nucleotide sequence alignment for IGHV4-61*2mm.....174
Supplementary Figure 1.4: Nucleotide sequence alignment for IGHV6-1*1m8.....174
Supplementary Figure 1.5: Germline IGHV gene usage for CAP255-75D.....175
Supplementary Figure 1.6: Germline IGHV gene usage for CAP255-37C and CAP255-18F.....175
Supplementary Figure 1.7: Germline IGHV gene usage for CAP255-53E.....176
Supplementary Figure 1.8: Germline IGHV gene usage for CAP177-2D, 13F, 60B, 6C and 13G.....177
Supplementary Figure 1.9: Germline IGHV gene usage for CAP177-23-3G, 1C and 5G.....178
Supplementary Figure 1.10: Germline IGHV gene usage for CAP255-91B and 71E.....179
Supplementary Figure 1.11: Germline IGHV gene usage for 2G12.....179
Supplementary Figure 1.12: Germline IGHV gene usage for CH103.....180

Supplementary Figure 1.13: Germline IGHV gene usage for PGTs121-123 180
Supplementary Figure 1.14: Germline IGHV gene usage for PGTs125-128 181
Supplementary Figure 1.15: Germline IGHV gene usage for PGTs135-137 182
Supplementary Figure 3.1: Longitudinal binding of glycans with elevated binding during HIV-1 infection.....206

LIST OF TABLES

Section A

Table 1: Comparison of non-broadly- and broadly neutralizing antibody features.....18

Table 2: Germline immunoglobulin gene diversity.....22

Section B

Table 2.1: Number of reads obtained from the pooled MiSeq run containing all seven time points and plasmid controls.....94

Table 2.2: CAP88-CH06 clonally related IgA and IgG sequences.....99

Table 3.1: Glycans with increased binding in HIV-positive compared to HIV-negative samples.....132

Table 3.2: List of glycans with elevated levels during HIV-1 infection compared to matched pre-infection samples.....134

Table 3.3: List of glycans significantly affected by the gp120 competition assay.....136

Table 3.4: List of glycans that were elevated during HIV infection.....138

Table 3.5: List of glycans significantly lower in BCN vs. non-BCN individuals.....142

Appendix

Supplementary Table 1.1: Viral loads and neutralization breadth from all CAPRISA participants.....170

Supplementary Table 1.2: Germline IGHV primers for 454 sequencing.....171

Supplementary Table 1.3: Germline IGHV primers for Illumina sequencing.....172

Supplementary Table 3.1: Plasma neutralization breadth, viral loads and duration of infection for HIV-positive samples.....191
Supplementary Table 3.2: List of all glycan components spotted onto the glycan arrays.....192
Supplementary Table 3.3: List of individuals used for the gp120 competition assay.....204
Supplementary Table 3.4: p-values and fold-changes for glycans affected by the gp120 competition assay.....204

LIST OF ABBREVIATIONS

aa:	Amino Acids
ADCC:	Antibody-Dependent Cellular-Cytotoxicity
AIDS:	Acquired Immune Deficiency Syndrome
ART:	Antiretroviral Therapy
ARV:	Antiretroviral
BCN:	Broad Cross-Neutralization
BCR:	B-cell Receptor
bNAb(s):	Broadly Neutralizing Antibody(s)
BSA:	Bovine Serum Albumin
C:	Constant
CA:	Capsid Protein
CAPRISA:	Centre for the AIDS Programme of Research in South Africa
CATNAP:	Compile Analyse and Tally NAb Panels Database
CCR5:	C-C motif Chemokine Receptor 5
cDNA:	Complementary DNA
CDR:	Complementarity determining Region
CDRH:	Complementarity determining Region on the Heavy Chain
CD4bs:	CD4 Binding Site
CSR:	Class-Switch Recombination
D:	Diversity
dH_2O:	Distilled Water
DNA:	Deoxyribonucleic Acid
FBS:	Fetal Bovine Serum
FR:	Framework Region
gDNA:	Genomic DNA

gp41:	Glycoprotein 41 on the HIV-1 envelope
gp120:	Glycoprotein 120 on the HIV-1 envelope
HIV:	Human Immunodeficiency Virus
HIV-1:	Human Immunodeficiency Virus Type 1
HIV-2:	Human Immunodeficiency Virus Type 2
HSA:	Human Serum Albumin
IAVI:	International AIDS Vaccine Initiative
IGHD:	Immunoglobulin Heavy Chain Diversity Gene
IGHJ:	Immunoglobulin Heavy Chain Joining Gene
IGHV:	Immunoglobulin Heavy Chain Variable Gene
IgA:	Immunoglobulin Alpha (A)
IgD:	Immunoglobulin Delta (D)
IgE:	Immunoglobulin Epsilon (E)
IgG:	Immunoglobulin Gamma (G)
IgM:	Immunoglobulin Mu (M)
IgPdb:	Immunoglobulin Polymorphism Database
IMGT:	Immunoglobulin Genetics Database
IN:	Integrase
J:	Joining
mAb(s):	Monoclonal Antibody(s)
MA:	Matrix Protein
MPER:	Membrane Proximal External Region
nAb:	Neutralizing Antibody
NC:	Nucleocapsid Protein
NCBI:	National Center for Bioinformatics Information
NCI:	National Cancer Institute
NGS:	Next Generation Sequencing
NICD:	National Institute for Communicable Diseases
NIH:	National Institutes of Health
NRF:	National Research Foundation

NON-BCN: Non-Broad Cross-Neutralization

PBMCs: Peripheral Blood Mononuclear Cells

PBS: Phosphate Buffered Saline

PBST: Phosphate Buffered Saline with Triton X

PCR: Polymerase Chain Reaction

PR: Protease

PRF: Poliomyelitis Research Foundation

RFU: Relative Fluorescence Unit

RNA: Ribonucleic Acid

RNAse: Ribonuclease

RT: Reverse Transcriptase

SANBI: South African National Bioinformatics Institute

SIV: Simian Immunodeficiency Virus

SHM: Somatic Hypermutation

SNP(s): Single Nucleotide Polymorphism(s)

SU: Surface Protein

TM: Transmembrane Protein

UCA: Unmutated Common Ancestor

UCSC: University of California, Santa Cruz

V: Variable

VH: Heavy Chain Variable Region

WHO: World Health Organization

W.P.I: Weeks Post-Infection

TABLE OF CONTENTS

PLAGIARISM DECLARATION ..... **II**

ABSTRACT ..... **III**

ACKNOWLEDGMENTS ..... **VII**

PREFACE ..... **IX**

1. PUBLICATIONS IX

2. CONFERENCE PRESENTATIONS X

ETHICS ..... **XI**

LIST OF FIGURES ..... **XII**

LIST OF TABLES ..... **XVI**

LIST OF ABBREVIATIONS ..... **XVIII**

TABLE OF CONTENTS ..... **XXI**

SECTION A **1**

1. INTRODUCTION 2

1.1. HIV Pandemic and Interventions *2*

1.2. HIV Diversity, Composition and Life Cycle *4*

1.2.1. HIV Diversity 4

1.2.2. HIV-1 Genetic Composition 6

1.2.3. HIV-1 Life Cycle 7

1.3. HIV-1 Envelope Structure *10*

1.4. HIV-1 Pathogenesis *13*

1.5. Antibody Responses to HIV-1 ..... *17*

1.6. Broadly-Neutralizing Antibody Targets *19*

1.7. Glycan Targets and Microarrays *21*

1.8. Antibody Development *22*

1.8.1. Germline Immunoglobulin Genes 22

1.8.2. Early B-cell Development and V(D)J Recombination 23

1.8.3. Antibody Maturation 26

1.8.4. Antibody Constant Region Effector Functions 29

1.9. Antibody Repertoire Analysis and Next Generation Sequencing *31*

1.10. Current Status of Vaccine Development *32*

2. STUDY OBJECTIVES 35

3. METHODS SUMMARY 37

3.1. Samples *37*

3.2. Nucleic Acid extraction and cDNA Synthesis *38*

3.3. Primer Design *38*

3.4. Library Construction for Next Generation Sequencing *39*

3.5. Next Generation Sequencing *40*

3.6. Sequence Data Analysis *40*

3.7. Glycan Microarrays *41*

3.8. Glycan Microarray Data Analysis *41*

3.9. Statistical Analysis *42*

4. REFERENCES 43

SECTION B: RESULTS 48

1. ABILITY TO DEVELOP BROADLY NEUTRALIZING HIV-1 ANTIBODIES IS NOT RESTRICTED BY THE GERMLINE IMMUNOGLOBULIN GENE REPERTOIRE 49

1.1. ABSTRACT *50*

1.2. INTRODUCTION *51*

1.3. METHODS AND MATERIALS *54*

1.3.1. Samples 54

1.3.2. DNA Extraction 54
1.3.3. Primer Design and Amplicon Library Preparation 55
1.3.4. Next Generation Sequencing 56
1.3.5. Sequence Data Analysis 57
1.3.6. Statistical Analysis 59
1.4. RESULTS *60*
1.4.1. Comparison of NGS technologies 60
1.4.2. Novel Germline IGHV Alleles in South African Women 60
1.4.3. Novel Germline IGHV Allele Usage in bNAbs 62
1.4.4. Comparison of Germline IGHV Repertoires in BCN and non-BCN Individuals 68
1.5. DISCUSSION *73*
1.6. CONCLUSIONS *77*
1.7. REFERENCES *78*
2. LIMITED EVOLUTION OF A POTENT STRAIN-SPECIFIC ANTIBODY LINEAGE IN AN HIV-INFECTED INDIVIDUAL WHO FAILS TO DEVELOP NEUTRALIZATION BREADTH 81
2.1. ABSTRACT *82*
2.2. INTRODUCTION *83*
2.3. METHODS AND MATERIALS *86*
2.3.1. Samples 86
2.3.2. Total Nucleic Acid Extraction and cDNA Synthesis 86
2.3.3. Primer Design 87
2.3.4. Library Construction and Illumina Sequencing 87
2.3.5. Antibody Sequence Data Analysis and Antibodyomics Pipeline 89
2.4. RESULTS *92*
2.4.1. CAP88-CH06 Neutralization and Sequencing 92
2.4.2. CAP88-CH06 Plasmid Controls 94
2.4.3. CAP88-CH06 Clonally Related Antibodies 95
2.4.4. CAP88-CH06 Phylogeny 100

2.4.5. CAP88-CH06 Lineage Evolution 104
2.5. DISCUSSION 106
2.6. CONCLUSIONS 111
2.7. REFERENCES 112
3. INDUCTION OF SERUM GLYCAN-BINDING ANTIBODIES DURING HIV-1 INFECTION 114
3.1. ABSTRACT 115
3.2. INTRODUCTION 116
3.3. METHODS AND MATERIALS 120
3.3.1. Serum Samples 120
3.3.2. Array Fabrication and Binding 121
3.3.3. Image Analysis and Data Processing 123
3.3.4. gp120 Competition Assay 123
3.3.5. Statistical and Data Analysis 124
3.4. RESULTS 126
3.4.1. Detection of bNAb Binding on Glycan Arrays 126
3.4.2. Natural Fluctuations in Anti-Glycan IgG Antibodies in HIV-Negative Individuals 127
3.4.3. Induction of IgG Anti-Glycan Antibodies During HIV infection 130
3.4.4. BCN vs. non-BCN Glycan-Binding Profiles 140
3.5. DISCUSSION 145
3.6. CONCLUSIONS 151
3.7. REFERENCES 152
GENERAL CONCLUSIONS 155
FULL REFERENCE LIST 159
APPENDIX 168

SECTION A

1. INTRODUCTION

1.1. HIV Pandemic and Interventions

The HIV pandemic continues to be one of the worlds major public health problems, with approximately 35.3 million people (3.3 million of whom are children) at the end of 2012 living with HIV (UNAIDS, 2013). The greatest burden of disease is in sub-Saharan Africa, accounting for 71% (25 million) of the global infections (Figure 1). Globally, there were approximately 2.3 million new infections (with ~6,300 infections occurring daily) and 1.6 million AIDS-related deaths in 2012 (UNAIDS, 2013). Although these numbers are still extremely high, they represent a decrease of 33% in adult HIV prevalence rates and 52% in new infections in children since 2001. AIDS-related deaths have also decreased by 30% since 2005 (UNAIDS, 2013).

The decreases in number of infections and AIDS-related deaths can be attributed to the increased rollout of antiretroviral therapy (ART) and the use of ART in the prevention of mother-to-child transmission (PMTCT) (Barre-Sinoussi*, et al.*, 2013). At the end of 2013, ~12.9 million people living with HIV were receiving ART, of whom ~11.9 million are from low- to middle-income countries (WHO, 2014). While this represents the bulk of people on ART it accounts only for 36% (32.6 million) of the individuals living with HIV in these areas (WHO, 2014), highlighting that much still needs to be done. Other prevention interventions such as male circumcision, pre-exposure prophylaxis (PrEP) and condom use have also had some impact on the reduced infection rate.

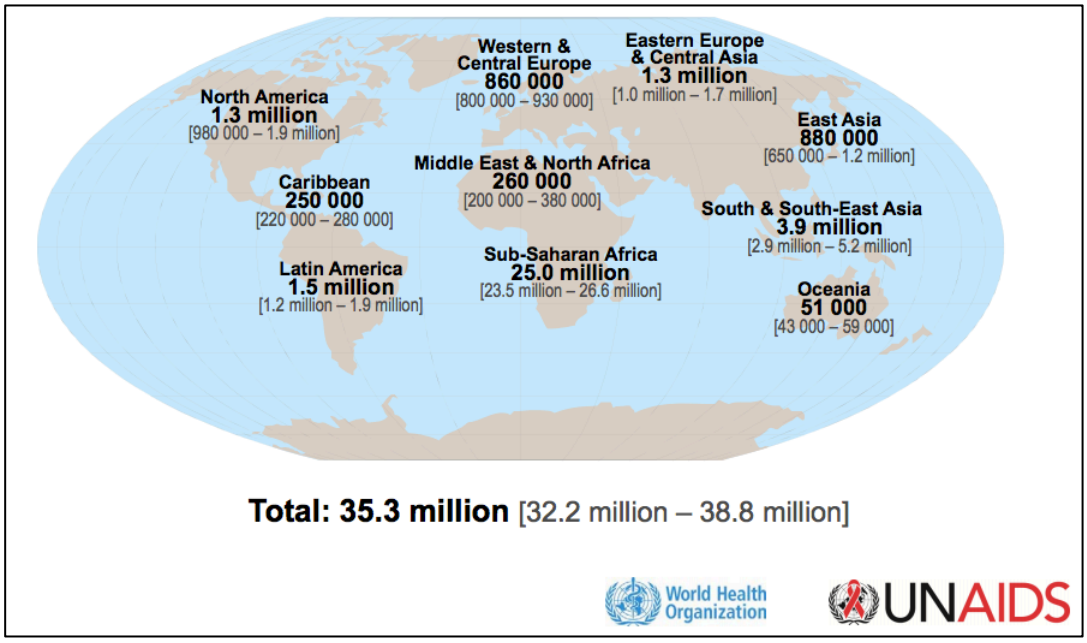


Figure 1: Estimated prevalence of HIV in adults and children worldwide in 2012

Shown are the numbers of adults and children estimated to be living with HIV worldwide, at the end of 2012. The figures in brackets represent the range of estimated values. Figure taken from (UNAIDS, 2013).

Although many of these preventative interventions are helping to reduce the transmission rate, the HIV epidemic continues to grow, particularly in sub-Saharan Africa among high-risk populations. In addition, although ART delays disease progression it is not able to completely eradicate the virus from an infected individual nor can it restore damage caused by the infection, such as immune exhaustion and gut damage (Barre-Sinoussi, *et al.*, 2013). Furthermore many individuals eligible for ART choose not to take the treatment or are non-compliant, due to the adverse effects associated with the ART, or socio-economic challenges (Enriquez, *et al.*, 2011). Nonadherence to ART is associated with non-suppression of viral load, which can lead to drug resistance and treatment failure.

Since adherence is a large problem associated with ART use and since ART does not eliminate the virus from an infected individual or completely reverse the pathology associated with HIV infection, many studies are focused on finding a vaccine for HIV. A vaccine would ideally prevent further HIV infections, but is still lacking despite years of

research in this area (discussed in more detail in Section 1.8). However there is also active research in the area of a cure for HIV. The two main strategies in this field of research are 1) eradication, in which the virus is completely eliminated from the body and therapy is no longer required; 2) functional cure in which the virus is still present but at very low levels and is controlled without requiring standard treatments (Fauci, *et al.*, 2013). Latently infected resting CD4 T-cells, which contain the integrated viral genome, have been identified as part of the HIV reservoir and is one of the sources of the viral re-emergence when ART is discontinued (Chun, *et al.*, 1999). It has also been shown that early initiation of ART may prevent the establishment of reservoirs enabling patients to discontinue treatment after viral load suppression without viral rebound (Fauci, *et al.*, 2013). Early treatment initiation is often impossible, especially in resource limited settings, thus other strategies such as the activation and elimination of latently infected cells, immuno-toxic therapy of the HIV reservoir, gene therapy and stem-cell transplantation are being researched (Cohen, 2013).

1.2. HIV Diversity, Composition and Life Cycle

1.2.1. HIV Diversity

There are two types of the Human Immunodeficiency Virus, HIV-1 and HIV-2. HIV-2 is less virulent than HIV-1, with typically lower viral loads and is largely restricted to West Africa (Sharp, *et al.*, 2011). Since HIV-1 is responsible for the pandemic seen throughout the world, the majority of the research is focused on this virus.

HIV-1 originated as a primate lentivirus, Simian Immunodeficiency Virus (SIV), in Chimpanzees (SIVcpz) but crossed the species barrier to infect humans (Sharp, *et al.*, 2011). HIV-1 is divided into four distinct lineages known as M, N, O and P, each of which

resulted from four separate cross-species transmission events. Group M was the first to be discovered and accounts for the majority of the pandemic viruses. The remaining groups are less prevalent with group O accounting for less than 1% of infections, group N identified in 13 cases and group P found in two individuals; all three groups (N, O and P) originated in West Africa (Sharp*, et al.*, 2011).

Group M is further subdivided into different subtypes/clades A, B, C, D, F, G, H, J, K and circulating recombinant forms (CRFs) and are usually restricted to certain parts of the world, shown in Figure 2 (Taylor*, et al.*, 2008). The most prevalent subtypes worldwide are A, B and C, with subtype C accounting for almost half of the infections (Buonaguro*, et al.*, 2007). Subtype A is found in central and eastern African countries (including Kenya, Uganda, Tanzania and Rwanda) and eastern European countries. Subtype B is found in northern America, western and central Europe and Australia while subtype C is found in southern Africa (including South Africa) and India.

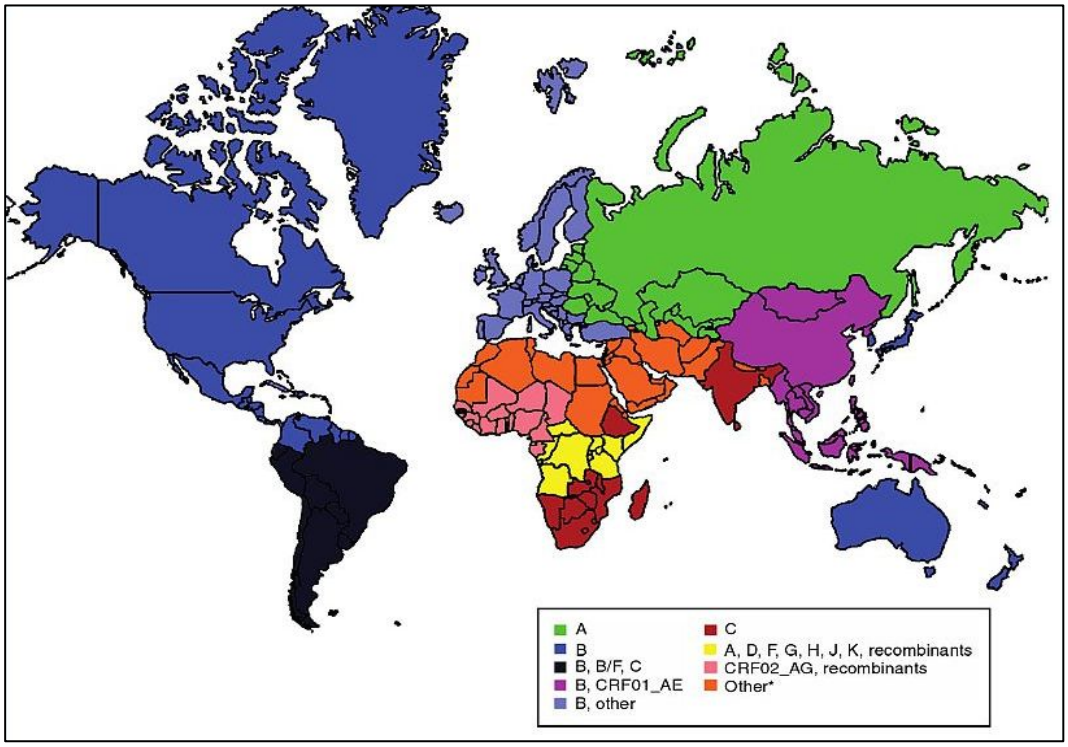


Figure 2: Distribution of HIV-1 clades worldwide

Shown are the different clades spread throughout the world, coloured according to clade. This figure was taken from (Medscape, 2009).

1.2.2. HIV-1 Genetic Composition

HIV-1 is a retrovirus with nine open reading frames in an RNA genome, which encode 15 proteins, shown in Figure 3 (Frankel*, et al.*, 1998). Three of the open reading frames (*gag*, e*nv* and *pol*) encode structural proteins, with g*ag* comprising a Matrix (MA), Capsid (CA), Nucleocapsid (NC) and p6 proteins which make up the viral capsid (Frankel*, et al.*, 1998). *Env* encodes the surface (SU) gp120 glycoprotein and transmembrane (TM) gp41 glycoprotein of the viral spike, while *pol* encodes viral polymerases: Protease (PR), Reverse Transcriptase (RT), Ribonuclease (RNase) and Integrase (IN) necessary for viral replication (Frankel*, et al.*, 1998). The remaining open reading frames encode accessory proteins, three (Vif, Vpr and Nef) of which are found in the viral particle. Vif is involved in viral infectivity, Vpr is essential for viral nucleoprotein transport into the host nucleus

and Nef is involved in host receptor degradation and cell apoptosis (Frankel*, et al.*, 1998). The remaining proteins (Rev, Tat and Vpu) are not found in the viral particle but are essential for gene regulation and viral assembly (Frankel*, et al.*, 1998).

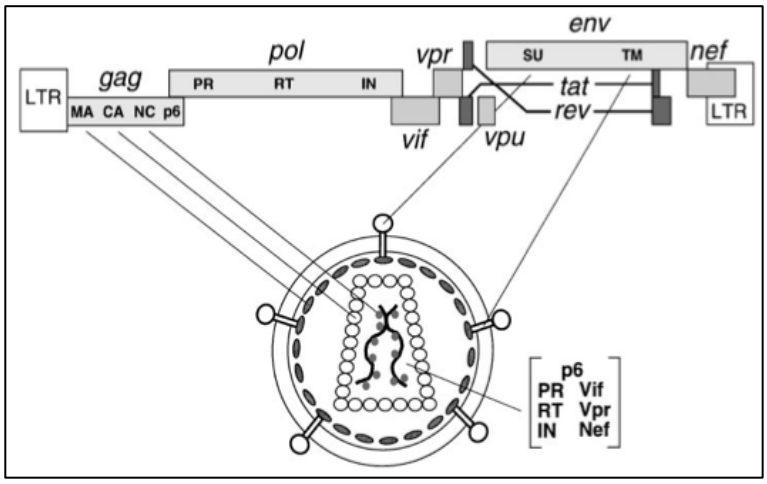


Figure 3: HIV-1 genome and viral particle structure Shown are the 15 proteins encoded in the HIV-1 RNA genome as well as the viral particle comprising structural proteins (viral envelope spikes SU and TM, MA, CA, and NC), enzymes (PR, RT and IN), accessory proteins (Vif, Vpr and Nef) and two single stranded RNA molecules. Figure taken from (Frankel*, et al.*, 1998).

1.2.3. HIV-1 Life Cycle

Like many viruses, HIV-1 uses host cells for viral replication, targeting immune cells expressing CD4 such as dendritic cells, macrophages and $CD4^+$ T-cells. The envelope spikes on the virus, comprising trimeric gp120 and gp41 glycoproteins (shown in Figure 3), are responsible for binding and initiating entry into $CD4^+$ T-cells. Contact is first made between the CD4-receptor on T-cells and the CD4-binding site on gp120, (Figure 4). This interaction causes the virus to undergo conformational changes opening up a secondary binding site for a cell surface co-receptor, either CCR5 or CXCR4, following which the fusion peptide on gp41 is inserted into the host cell (Engelman*, et al.*, 2012; Barre-Sinoussi*, et al.*, 2013). The viral capsid then opens up to release the viral RNA, enzymes and accessory proteins into the host cytoplasm. Once in the cytoplasm the single stranded viral RNA molecules released from the capsid are reverse transcribed into double-stranded DNA using the viral RT. The DNA viral genome is transported into the cell nucleus as a

pre-integrated complex (PIC) and inserted into the host genome with the aid of viral proteins Vpr and IN and the host cell protein lens epithelium-derived growth factor (LEDGF) (Frankel, *et al.*, 1998; Engelman, *et al.*, 2012; Barre-Sinoussi, *et al.*, 2013).

Transcription of the integrated viral genome is facilitated by the viral Tat and host proteins: RNA polymerase II (RNA Pol II) and positive transcription elongation factor b (P-TEFb) (Engelman, *et al.*, 2012). The transcribed mRNA molecules are transported out of the host cell nucleus into the cytoplasm, in some cases facilitated by the viral protein Nef and host protein CRM1. Polyproteins Gag and Pol along with accessory viral proteins Vif, Vpr and Nef are translated in the host cytoplasm and assembled along with the entire viral genome length RNA molecules.

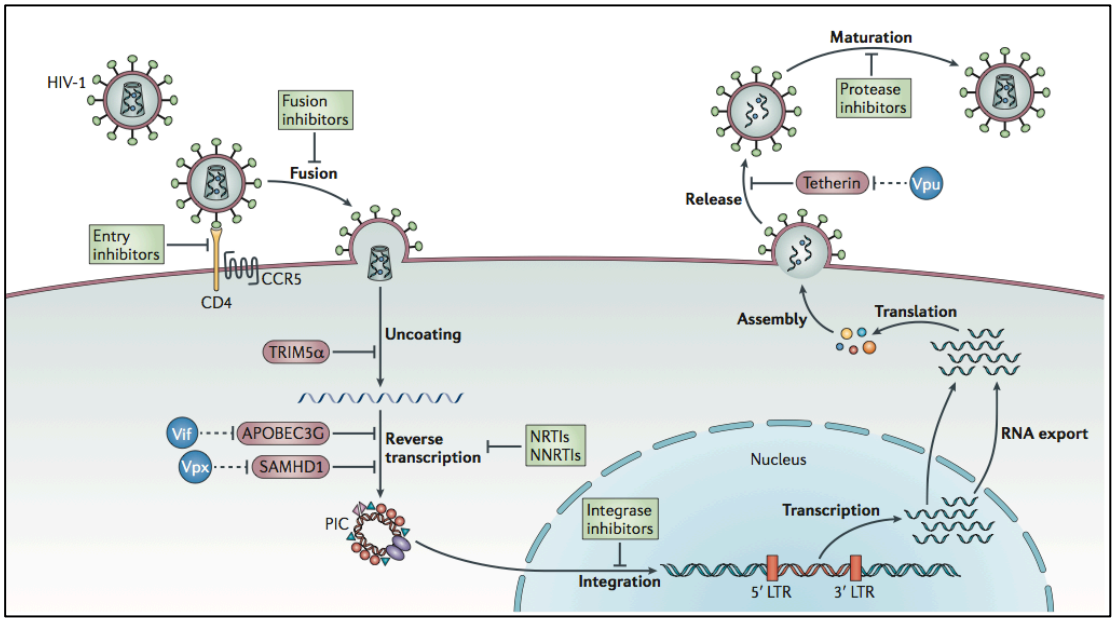


Figure 4: Life cycle of HIV-1

Shown is a simplification of the life cycle of HIV-1 in CD4 T-cells. The green boxes represent anti-retroviral drug regimens that target specific aspects of the life cycle. The brown boxes represent host proteins targeting the different steps in the viral life cycle. Figure taken from (Barre-Sinoussi, *et al.*, 2013).

The Env protein is synthesised as a trimeric gp160 precursor polyprotein and glycans are added to its surface in the endoplasmic reticulum of the infected host cell

(Wyatt, *et al.*, 1998). The gp160 trimeric protein is then transported to the Golgi apparatus where it is cleaved by the host enzyme furin to gp120 and gp41 (Bonomelli, *et al.*, 2011). The glycans, surrounding the complex, undergo modifications, which result in inefficiently trimmed glycans. The trimeric gp120 and gp41 complexes are then transported to the host cell membrane in preparation for budding (Wyatt, *et al.*, 1998).

The assembled viral proteins and RNA molecules bud out of the host cell surface embedded with HIV envelope glycoprotein trimers (gp120 and gp41). These are released with the aid of host proteins Endosomal Sorting Complex Required for Transport (ESCRT) I and II and Actin Intergrating Protein 1 (AIP1) (Engelman, *et al.*, 2012). In the final step of the HIV-1 life cycle, PR is used to cleave the Gag and Pol polyproteins into the structural components of the viral capsid: MA, CA, NC and viral enzymes PR, RT and IN, respectively.

Numerous host responses (highlighted in brown boxes in Figure 4) and anti-retroviral drug regimens (highlighted in green boxes in Figure 4) target various aspects of the HIV-1 life cycle that interfere with viral replication. Some individuals have a natural protection against HIV infection as a result of an inherited CCR5 deletion resulting a non-functional CCR5 receptor on their $CD4^+$ T-cells and thus the inability of CCR5 dependent viruses to attach to $CD4^+$ T-cells and gain entry (Liu, *et al.*, 1996; Samson, *et al.*, 1996). Once the virus has gained entry into the host cell, TRIM5α, a host restriction factor, prevents the uncoating of the virus, while APOBEC3G and SAMHD1 prevent the reverse transcription of the viral RNA genome to DNA (Barre-Sinoussi, *et al.*, 2013). These host enzymes, however, are counteracted by the viral proteins Vpu, Vif and Vpx (Barre-Sinoussi, *et al.*, 2013). Tetherin, a host type II transmembrane protein, prevents viral particles from budding out of the host cell, but can be thwarted by Vpu (Engelman, *et al.*, 2012; Barre-Sinoussi, *et al.*, 2013).

Similarly antiretroviral drug regimens such as entry inhibitors prevent the virus from entering the host cell and include both CCR5 inhibitors, which block the binding of the virus to the CCR5 co-receptor, and fusion inhibitors which prevent the fusion of the virus into the host cell (Arts*, et al.*, 2012). Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs) and Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) are classes of drugs that inhibit the transcription of the viral RNA genome to DNA, while Integrase inhibitors prevent the integration of the viral genome into the host genome (Arts*, et al.*, 2012). The Protease inhibitor class of anti-viral drugs prevents PR from cleaving the Gag and Pol polyproteins into the viral capsid and enzyme proteins and thus the formation of a mature viral particle (Arts*, et al.*, 2012).

1.3. HIV-1 Envelope Structure

Since HIV gains entry into the host cell via the envelope trimeric spike, which is also the main target of the antibody response (discussed in Sections 1.5 and 1.6), many studies have focused on elucidating the structure of the trimer. As previously mentioned the envelope of HIV is composed of trimeric gp120 and gp41 glycoproteins, which are initially synthesised as a gp160 precursor protein and cleaved by furin in the endoplasmic reticulum and further processed with the addition of glycans in the Golgi complex (Wyatt*, et al.*, 1998; Bonomelli*, et al.*, 2011).

The trimeric structure undergoes various conformational changes prior, during and after fusion of the virus into host cells. The pre-fusion state of the envelope that includes a detailed view of the gp41 subunit was recently elucidated using a BG505 SOSIP.664 protein bound to PGT122 and 35O22 antibodies (Pancera*, et al.*, 2014). In the pre-fusion state four helices (α6 to α9) from gp41 enclose the amino- and carboxy- termini of gp120 forming a collar, which is clasped by the insertion of a methionine residue on the α6 and a

triple-tryptophan motif on α8 and α9 (Pancera, *et al.*, 2014). Upon binding of the trimeric structure to host cell receptors the clasp is released between gp41 and gp120 allowing the trimer to open up and gp41 to extend and penetrate the host cell membrane and promote cell fusion (Figure 5).

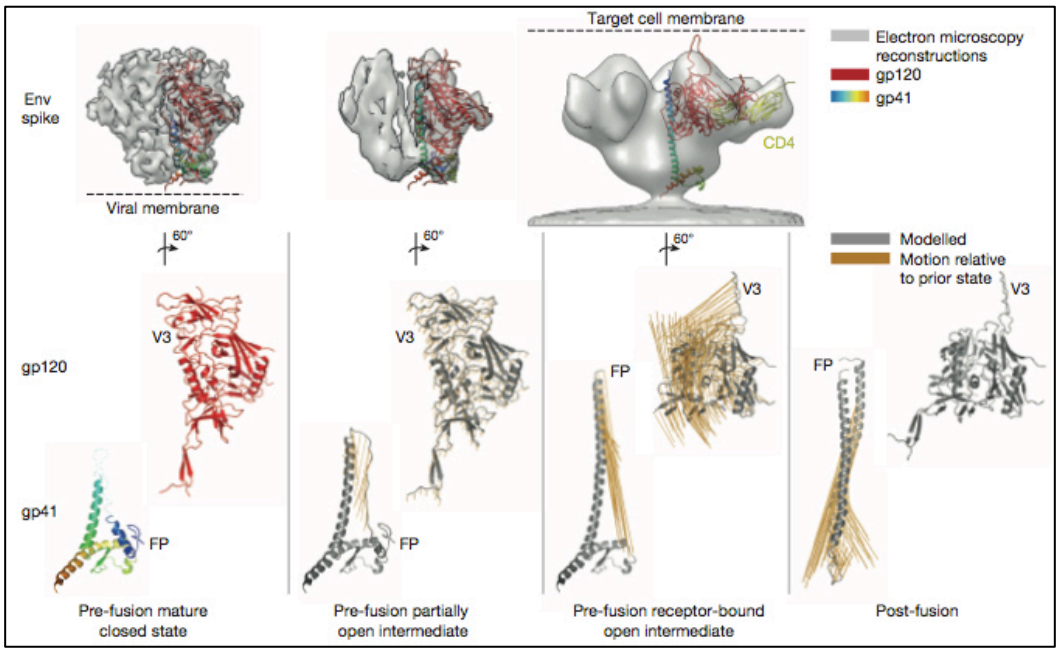


Figure 5: HIV-1 envelope trimer structure before and after fusing with CD4$^+$ T-cells
Shown are the HIV-1 envelope trimer (grey), gp120 (red) and gp41 (rainbow colours) before and after fusing with CD4 T-cells. Figure taken from (Pancera, *et al.*, 2014).

Prior to the elucidation of the gp41 subunit of the trimer, the pre-fusion state of gp120 was resolved using three different structures. Two of which used the BG505.SOSIP.664 trimer bound either by a CD4 binding site antibody (PGV04) or PGT122, which targets the N332 supersite (discussed in more detail in Section 1.6) (Julien, *et al.*, 2013; Lyumkis, *et al.*, 2013). The third used a KNH1144.SOSIP trimer bound by VRC03, also a CD4 binding site antibody (Bartesaghi, *et al.*, 2013). All three structures show that the gp120 trimer is held together at the apex by the V1, V2 and V3 loops and that the core of gp120 and the interaction between gp41 and gp120 is stabilized by the α helices of gp41 (Figure 6).

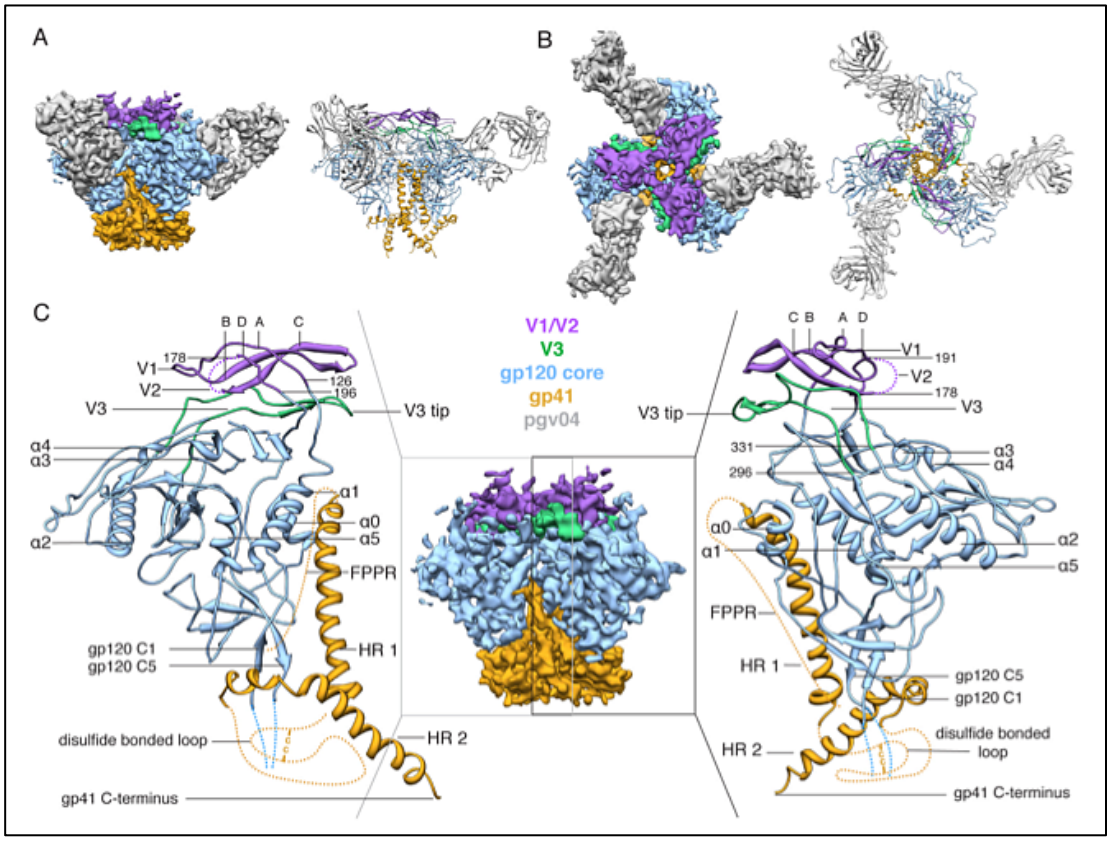


Figure 6: Reconstruction of the Env trimer structure bound by PGV04

Shown is a reconstruction of the crystal structure of the BG505.SOSIP trimer bound by PGV04 CD4bs bNAb (shown in grey). **A)** and **B)** Show the side and top views of the trimer structure. **C)** Shows the side view of the trimer with and without the Fab density. The V1V2 region of gp120 is represented in purple, the V3 region represented in green, the gp120 core in blue and gp41 in gold. This figure is taken from (Lyumkis, *et al.*, 2013).

Both gp120 and gp41 are heavy glycosylated with between 25-30 N-linked glycosylation sites and these glycans account for half of the molecular mass of the trimer, known as the glycan shield (Figure 7, (Wang, 2013; Pancera, *et al.*, 2014)). The glycan shield on trimeric gp120 is mostly comprised of densely packed high mannose, particularly Manα1-2Man terminating glycans such as $Man_{6\text{-}9}GlcNAc_2$, and complex N-linked glycans although some O-linked glycans, such as Tn-peptides, do occur (Zhu, *et al.*, 2000; Doores, *et al.*, 2010b; Go, *et al.*, 2013). Due to the high density of the glycans on the Env trimer they often lack the typical glycosylation processing and tend to be inefficiently trimmed. In addition to protecting immunogenic sites on the virus, these glycans are also essential for

glycoprotein folding and are thus often conserved (Bonomelli, *et al.*, 2011; Pancera, *et al.*, 2014).

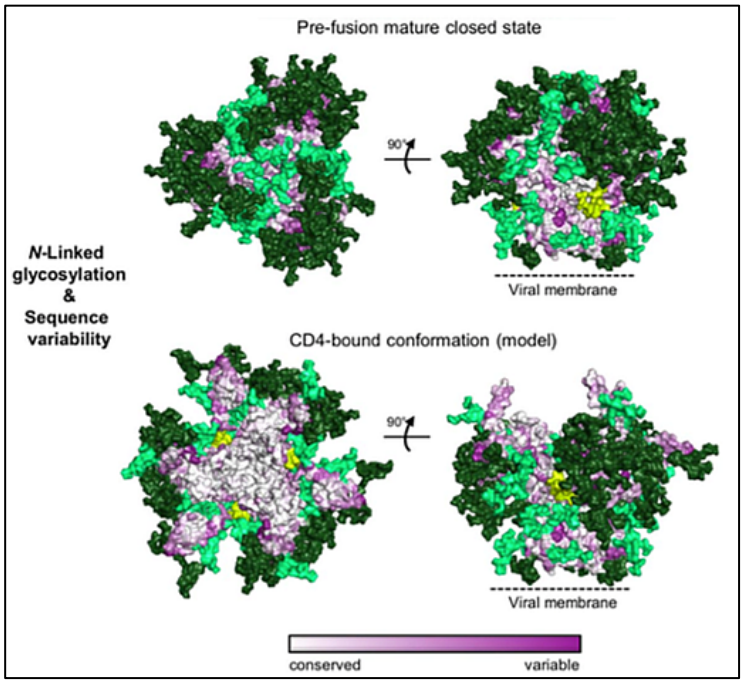


Figure 7: Glycan shield on the HIV-1 envelope trimer
Shown are the conserved (light green) and variable (dark green) glycans and sequence variability (pink) on the HIV-1 Env trimer in both the pre-fusion and CD4-bound conformations. Figure taken from (Pancera, *et al.*, 2014).

1.4. HIV-1 Pathogenesis

HIV is transmitted through bodily fluids such as blood, semen, pre-seminal fluid, vaginal fluids, rectal fluids, breast milk and sharing of contaminated needles (intravenous drug use). It targets cells expressing CD4 such as macrophages, dendritic cells and $CD4^+$ T-cells. Macrophages and dendritic cells are usually the first cells to be infected at the point of viral entry, followed by $CD4^+$ T-cells, which carry the burden of infection (Parham, 2009).

Disease progression is directly proportional to the level of virus and $CD4^+$ T-cells in plasma. The onset of infection is associated with high viral loads in the peripheral blood (shown in red in Figure 8) and rapid decline of $CD4^+$ T-cells (shown in blue in Figure 8) and is often but not always associated with flu-like symptoms (An, *et al.*, 2010). This rapid increase in viral load and decrease in $CD4^+$ T-cells is known as the acute phase and can

last approximately 6-12 weeks. During the acute phase the immune system responds by producing antibodies against the virus (described in more detail in Section 1.5) as well as a cytotoxic T-cell response, which results in a decrease in viral load and increase in $CD4^+$ T-cells.

Although these immune responses decrease the viral load in plasma they are unable to clear the virus and patients enter into a chronic phase of infection, which is associated with a steady state of viral replication (known as a set-point) and a gradual decline in $CD4^+$ T-cells if ARVs are not introduced (An*, et al.*, 2010). The reason why viral load is not completely suppressed is partly explained through the highly error prone rate of the viral reverse transcriptase (described in Section 1.2.3) which introduces nucleotide polymorphisms into the viral genome. Thus while infection is usually triggered by a single virus the accumulation of polymorphisms during infection gives rise to different viral strains or quasi-species within an infected individual (Parham, 2009).

During chronic HIV infection most patients are in a state of clinical latency in which no symptoms of the viral infection are apparent. Chronic phase can last years before the onset of AIDS, associated with a $CD4^+$ T-cell count below 200 cells/mm^3 (healthy individuals have $CD4^+$ T-cell counts above 1000 cells/mm^3) and an increased viral load (An*, et al.*, 2010). Some individuals, termed long-term non-progressors never progress to AIDS and others do not become infected despite exposure to the virus in some cases because of a mutation in their CCR5 co-receptor (discussed in Section 1.2.3). Subtype C infection has been associated with rapid disease progression with $CD4^+$ T-cells below 350 cells/mm^3 within 2 years of infection (Mlisana*, et al.*, 2014). The majority of individuals without ARV treatment, however, will progress to AIDS. The current recommendation for the initiation of ARV treatment in South Africa is when an individuals' $CD4^+$ T-cell count reaches below 500 cells/mm^3. In addition, all HIV-positive pregnant women are initiated

onto treatment regardless of $CD4^+$ T-cell count because of the significant efficacy of reducing infections from mother to child (South African Department of Health: www.doh.gov.za).

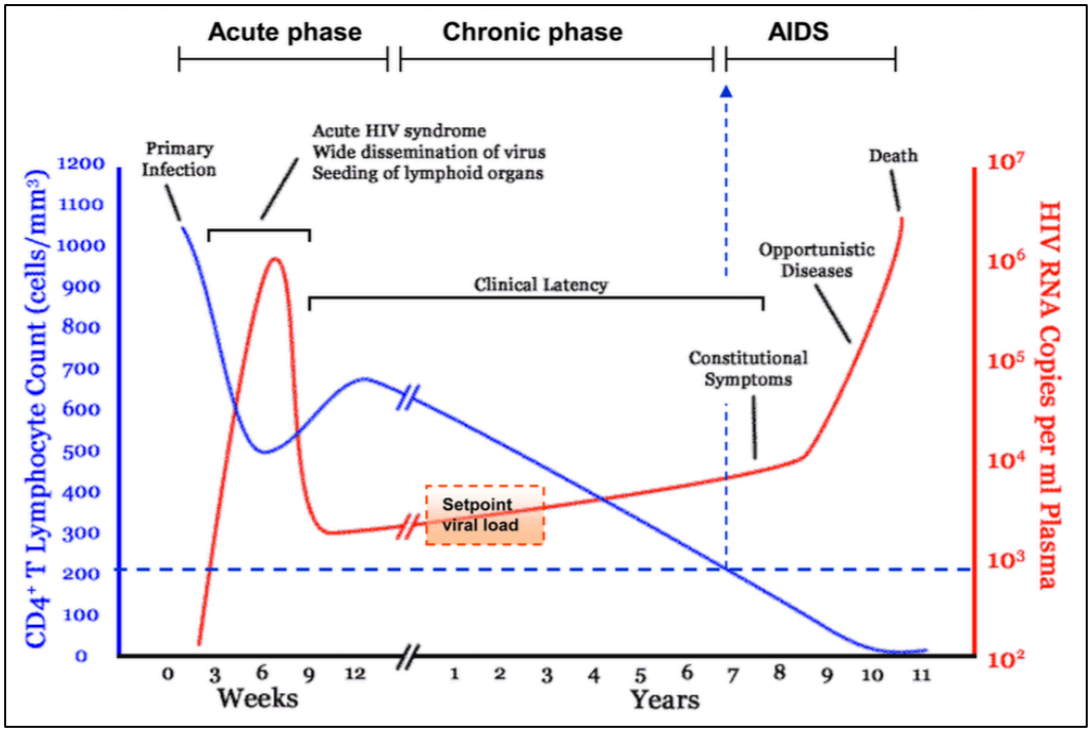


Figure 8: HIV viral load and CD4 count during infection

Shown are the typical viral loads (red) and $CD4^+$ T-cell count (blue) observed in an individual during primary infection, clinical latency and AIDS. Figure taken from (An, *et al.*, 2010).

Due to the declining number of $CD4^+$ T-cells the patient is unable to mount an effective immune response to HIV as well as to other potential pathogens that one encounters daily. Thus AIDS is associated with increased susceptibility to opportunistic infections including parasitic, bacterial, the most common of which is TB (*Mycobacterium tuberculosis*), fungal and other viral infections. In addition cancers associated with different viruses such as Kaposi's sarcoma, Non-Hodgkin's lymphoma and EBV-positive Burkitt's lymphoma are increased in AIDS patients (Parham, 2009). Typically an opportunistic infection overwhelms the patient leading to death.

In addition to the loss of $CD4^+$ T-cells during HIV infection a wide range of B-cell dysfunctions have also been reported. One of the defects associated with HIV infection is the presence of B-cell populations not normally observed in healthy individuals, such as immature transitional B-cells, activated mature B-cells, exhausted tissue-like memory B-cells and short-lived plasmablasts (Figure 9, (Moir, *et al.*, 2009). In contrast resting memory B-cells are decreased in HIV-infected individuals compared to HIV-negative individuals. The hyperactivation of B-cells due to chronic HIV leads to hypergammaglobulinaemia (over expression of IgG antibodies), increased polyclonal B-cell activation, increased differentiation into plasmablasts, increased autoreactivity and B-cell malignancies. The hyperactivated B-cells are not HIV-specific and tend to overwhelm the HIV-specific response.

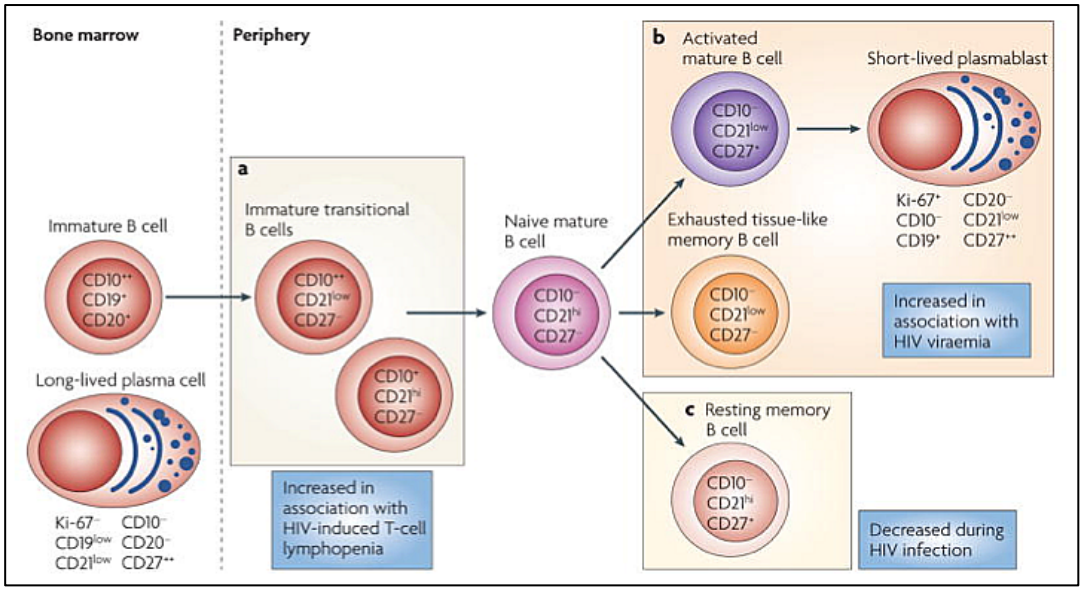


Figure 9: B-cell dysfunction and subpopulations associated with HIV-1 infection

Shown are different B-cell populations that are affected by HIV infection. Figure taken from (Moir, *et al.*, 2009).

1.5. Antibody Responses to HIV-1

During HIV-1 infection antibodies targeting various regions of HIV can be detected. Only antibodies targeting the envelope (gp120 and gp41 glycoproteins) are able to bind and neutralize HIV and will be discussed here. Binding antibody responses are first detected from ~12 days post-infection which bind gp41, however these responses have no impact on viral load (Figure 10, (Tomaras, *et al.*, 2008)). Autologous strain-specific neutralizing antibodies are detected from 4-14 weeks after infection and persist throughout infection and typically target gp120 (Gray, *et al.*, 2007; Euler, *et al.*, 2012; West, *et al.*, 2014). Although neutralization by these antibodies is limited to a single virus they exert enough pressure on the virus to drive viral escape. In approximately 15-30% of HIV-1 infected individuals broadly neutralizing antibodies (bNAbs) develop, which are able to neutralize numerous subtypes of the virus and are detected from 2-3 years post-infection (Gray, *et al.*, 2011b; West, *et al.*, 2014; Moore, *et al.*, 2015).

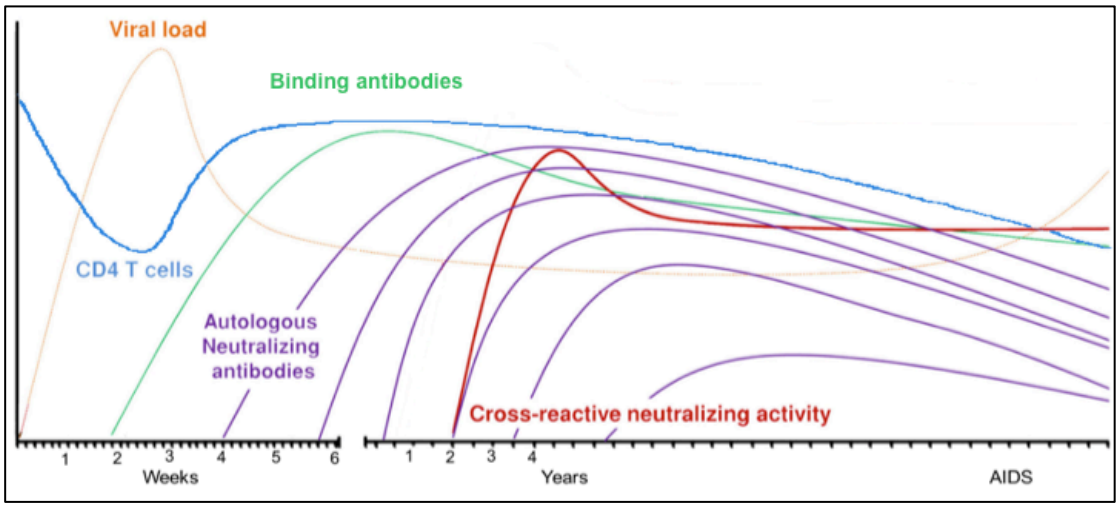


Figure 10: Anti-HIV-1 antibody response during infection

Shown is the $CD4^+$ T-cell count (blue), viral load (orange), non-neutralizing antibody responses (green), autologous neutralizing antibody responses (purple) and broadly neutralizing antibody responses (red) during HIV-1 infection. Figure adapted from (Euler, *et al.*, 2012).

BNAbs are the main focus of vaccine development since such antibodies have been shown to be protective in non-human primates (Moldt*, et al.*, 2012; Barouch*, et al.*, 2013; Ko*, et al.*, 2014; Saunders*, et al.*, 2015). However, thus far envelope vaccines have been unable to elicit such antibodies (discussed in more detail in Section 1.8), partly because the mechanisms underlying their emergence remain unknown. BNAbs have unusual features in that they tend to have unusually long CDRH3 regions, high levels of somatic hypermutation and are frequently polyreactive (Table 1, (West*, et al.*, 2014)).

Table 1: Comparison of non-broadly- and broadly neutralizing antibody features

	Non-bNAbs	bNAbs
Neutralization breadth	1 - 10%	40 - 90%
Level of somatic hypermutation	1 - 12%	12 - 32%
CDRH3 length	~16 amino acids	26 - 39 amino acids
Detected during infection	weeks - months	≥ 2 years

Data taken from (West*, et al.*, 2014; Zolla-Pazner, 2014).

To date over 100 bNAbs targeting 5 epitopes on the HIV-1 envelope have been isolated from a number of HIV-1 infected individuals. Typically these are IgG antibodies usually IgG1 although some have been isolated as IgG3 (CATNAP, 2014). In general most bNAbs neutralize >70% of viruses with 10E8, an MPER directed antibody, showing the highest breadth of 98% (Huang*, et al.*, 2012). BNAbs show a wide range of potencies with IC_{80} values of 0.0001 to 100 μg/ml (Doria-Rose*, et al.*, 2015). Despite their ability to potently neutralize numerous HIV strains and show protection against infection, these antibodies are not able to slow disease progression or reduce viraemia since viral escape against these targets readily occurs (Gray*, et al.*, 2011b; Moore*, et al.*, 2015).

1.6. Broadly-Neutralizing Antibody Targets

There are five main bNAb epitopes on the HIV envelope (Figure 11). These five sites are defined as the V2 site, N332 supersite, CD4-binding site, gp120-gp41 interface and the membrane proximal external region (MPER) (Wibmer, *et al.*, 2015). The V2 (highlighted in red in Figure 11) is at the apex of the trimer involving all three gp120 glycoproteins, which are protected by densely packed glycans (McLellan, *et al.*, 2011; Bartesaghi, *et al.*, 2013; Julien, *et al.*, 2013; Lyumkis, *et al.*, 2013). Neutralization of this region usually involves antibodies with very long CDRH3s, which are able to get around the variable loops and/or penetrate the glycan shield (Walker, *et al.*, 2009; McLellan, *et al.*, 2011; Doria-Rose, *et al.*, 2014). Due to the presence of many glycans at this site neutralization is often dependent on glycan-binding, such as PG9 and PG16, which bind glycans at the N160 region (Walker, *et al.*, 2009; McLellan, *et al.*, 2011). However neutralization at this site does occur without glycan-binding as seen in CAP256.VRC26 (Doria-Rose, *et al.*, 2014).

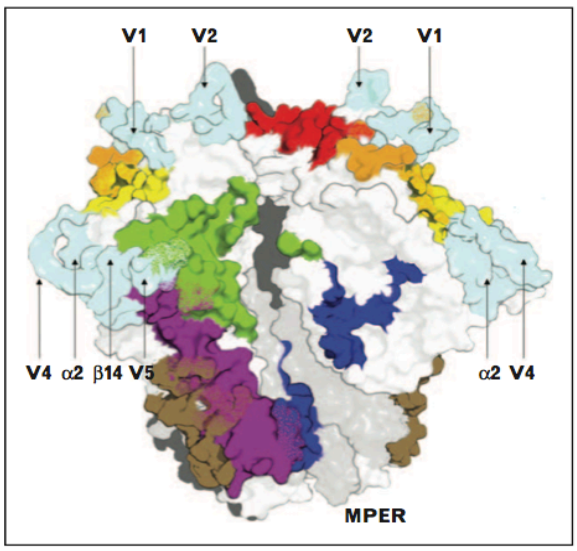


Figure 11: Broadly neutralizing antibody targets on the HIV-1 envelope trimer

Shown is the gp120 region of the HIV-1 Env trimer (cyan) with the variable loops labelled (V1-V5), gp41 region (grey) and the five main bNAb targets: V2 region (red), V3 and V4 regions (orange and yellow), CD4-binding site (CD4bs) (green) and the gp41-gp120 interface (brown, purple and blue for 8ANC195, 35022 and PGT151, respectively). The region where the MPER joins the trimer is shown. Figure taken from (Wibmer, *et al.*, 2015).

The N332 supersite involves many glycans covering the V2, V3 and V4 regions (red, orange and yellow in Figure 11). The key glycans involved in the N332 supersite are at positions: N137, N156, N295, N301, N332, N339, N386 and N392 (Kong*, et al.*, 2013). Many of the bNAbs targeting this region also have long CDRH3s, such as PGT121 and PGT128 which target glycans at N301 and N332 (Walker*, et al.*, 2011). 2G12, the first bNAb identified as having a glycan dependent specificity, binds glycans at N295, N332, N339 and N392 (Calarese*, et al.*, 2003). Other examples of bNAbs that target the N332 supersite are PGT130 and PGT135 (Walker*, et al.*, 2011; Kong*, et al.*, 2013; Doores*, et al.*, 2015).

The CD4-binding site (highlighted in green in Figure 11), which is the site at which the virus contacts $CD4^+$ T-cells to gain entry into the cell, usually does not involve glycans (Wibmer*, et al.*, 2015). bNAbs targeting this site tend to use the germline heavy chain variable genes IGHV1-2 and IGHV1-46, although some use IGHV1-3 and IGHV4-59 (West*, et al.*, 2012). Examples of bNAbs targeting the CD4-binding site are VRC01-03, NIH45-46, b12, 8ANC131 and CH103.

The gp120-gp41 interface (highlighted in purple, brown and blue in Figure 11) is a newly identified target. This site is found just below the CD4-binding site and involves numerous glycans at positions: N88, N230, N234, N241, N276, N611, N625 and N637 (Scharf*, et al.*, 2014). Three bNAbs associated with this epitope are 8ANC195 (epitope highlighted in purple in Figure 11), 35O22 (epitope highlighted in brown in Figure 11) and PGT151 (epitope highlighted in blue in Figure 11), (Falkowska*, et al.*, 2014; Huang*, et al.*, 2014).

The MPER (highlighted in grey in Figure 11) is a α-helical linear region within gp41 that links the transmembrane domain to the ectodomain (Wibmer*, et al.*, 2015).

Examples of bNAbs that bind the MPER are 2F5, Z13e1, 4E10 and 10E8 (Buchacher, *et al.*, 1994; Zwick, *et al.*, 2001; Binley, *et al.*, 2004; Huang, *et al.*, 2012).

1.7. Glycan Targets and Microarrays

Since many of the bNAbs targets involve glycans on the HIV envelope studies have focused on the types of glycans that are bound by these antibodies. As previously mentioned the HIV envelope is comprised of densely packed N-linked high mannose glycans such as $Man_{6-9}GlcNAc_2$, as well as complex N-linked glycans and O-linked glycans, such as Tn-peptides (Doores, *et al.*, 2010b; Go, *et al.*, 2013). Glycan-binding bNAbs 2G12, PGT125-131 and PGT135, which target the N332 supersite, bind high mannose N-linked glycans such as Man_{7-9} (Scanlan, *et al.*, 2002; Walker, *et al.*, 2011). PGT121-123, also targeting the N332 supersite, bind complex type glycans, such as α2-6 sialyated (A2), monoantennary *N*-glycan (N2) and complex-type asialyl biantennary glycan (NA2) (Mouquet, *et al.*, 2012). PG9 and PG16, which target the V2 site, bind both high mannose and complex sialylated N-linked glycans (Amin, *et al.*, 2013; Shivatare, *et al.*, 2013).

Glycan microarrays, which contain various glycans spotted onto glass slides, have been particularly useful in determining the types of glycans bound by bNAbs. The glycan specificities of PGTs 125-131 and PGT135 were determined using the mammalian printed glycan array version 5.0, created by the Consortium for Functional Glycomics (Walker, *et al.*, 2011; CFG, 2014), while the glycan specificities of PGT121-123 and PG9 and PG16 were determined using custom glycan arrays (Mouquet, *et al.*, 2012; Shivatare, *et al.*, 2013). Glycan arrays have also been useful in determining glycan responses in serum of non-human primates after Ad5hr-SIV vaccination and SIV infection as well as potential

HIV vaccine responses in rabbits (Astronomo, *et al.*, 2008; Dunlop, *et al.*, 2010; Campbell, *et al.*, 2013).

1.8. Antibody Development

1.8.1. Germline Immunoglobulin Genes

Human antibodies are made up of two identical heavy and light chains which are encoded at three different immunoglobulin gene loci: one for the heavy chain and two for the different light chains (κ and λ) (Lefranc, *et al.*, 2009). The heavy chain is encoded on chromosome 14 at 14q32.33, the κ light chain is encoded at 2p11.2 on chromosome 2 and the λ light chain is encoded at 22q11.2 on chromosome 22. Each human immunoglobulin gene locus contains a variable (V), joining (J) and constant (C) region, with the heavy chain having an additional diversity (D) region (Figure 12).

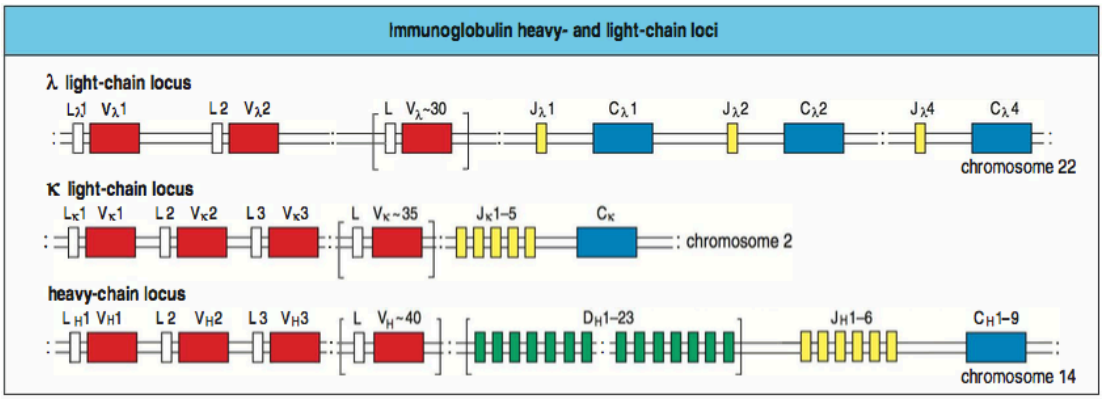


Figure 12: Germline heavy and light chain immunoglobulin gene loci

Shown are the three loci encoding the different human immunoglobulin gene segments that make up the heavy and light (λ and κ) chains. Figure taken from (Parham, 2009).

The human germline immunoglobulin gene loci are highly variable with each gene region (V, D, J and C) having multiple functional genes, open-reading frames (ORF) and pseudogenes with only functional genes being involved in antibody production. Most of these genes have multiple alleles, including functional and non-functional alleles, which

differ by either a single nucleotide polymorphism (SNP) or by multiple SNPs, which can be either synonymous or nonsynonymous, or frameshift mutations caused by indels that contribute diversity to the immunoglobulin gene repertoire (Table 2). Furthermore, whole IGHV genes have been reported as duplicated or deleted from the germline repertoire of some individuals resulting in varied gene copy numbers (Watson*, et al.*, 2012; Watson*, et al.*, 2013).

Table 2: Germline immunoglobulin gene diversity

Loci	V			D			J		C	
	Sub-groups	Genes	Alleles	Sub-groups	Genes	Alleles	Genes	Alleles	Genes	Alleles
λ	11	33	70				7	7	5	12
κ	7	42	70				5	9	1	5
Heavy	7	48	239	7	23	30	6	13	5	49

Numbers represent only human functional genes and alleles. Information obtained from (IMGT, 2014).

The five different human heavy chain constant region genes (μ, δ, γ, α and ε, which encode IgM, IgD, IgG, IgA and IgE, respectively) are responsible for determining the isotype of the antibody (Lefranc*, et al.*, 2009). The IgG and IgA isotypes are further divided into different subtypes IgG1-IgG4 and IgA1 and IgA2. All of the genes in both the heavy and light immunoglobulin gene loci are termed the germline immunoglobulin gene repertoire. The large number of genes and high level of variability in the germline repertoire is necessary to enable immune system to respond to a wide range of antigenic stimuli. Diversity is further achieved through somatic recombination of the variable, diversity, joining and constant regions of the heavy and light chains.

1.8.2. Early B-cell Development and V(D)J Recombination

B-cells are responsible for antibody development and secretion. During the embryonic stage B-cells arise in the fetal liver and in the bone marrow in adults (Tobon*, et al.*, 2013).

B-cells start off as hematopoietic precursor cells that become progenitor B-cells (pro-B-cells, Figure 13) involving somatic recombination of a single variable (V), diversity (D) and joining (J) regions of the heavy chain immunoglobulin genes found on chromosome 14 (Pieper*, et al.*, 2013; Tobon*, et al.*, 2013).

The VDJ recombination of the heavy chain follows two steps, the first step involves the joining of a single D gene to a J gene which is then followed by the joining of the DJ segment to a single V gene (Janeway*, et al.*, 2001). This process occurs at the recombination signal sequences (RSS) on either side of the different V, D and J segments and is mediated by the recombination activating gene proteins 1 and 2 (RAG1/2) (Van Gent*, et al.*, 2001). The DNA of the intervening V, D and J segments are excised and the recombined ends joined with the aid of several proteins, KU70, KU80, DNA protein kinase (DNA-PK_{CS}), DNA ligase IV and XRCC4. During the ligation phase of the recombination, terminal deoxynucleotidyl transferase adds nucleotides between the recombined segments, creating additional diversity in the VDJ exon (Market*, et al.*, 2003).

Following the recombination of the heavy chain in the pro-B-cell, another round of somatic recombination occurs in the light chain using the same process as that of the heavy chain, resulting in a precursor B-cell (pre-B-cell, Figure 13 (Pieper*, et al.*, 2013; Tobon*, et al.*, 2013)). An immature B-cell expresses a membrane bound IgM antibody (Figure 13B, (Pieper*, et al.*, 2013; Tobon*, et al.*, 2013)). The V(D)J region of an antibody is collectively known as the variable region and is the region that binds antigens. This region is further subdivided into three complementarity determining regions (CDR, CDRH (for heavy chains) and CDRL (for light chains)), which are flanked by four framework regions (FR) (Parham, 2009). The complementarity determining regions are highly variable while the framework regions tend to be more conserved. The first two complementarity determining regions are encoded entirely by the variable region genes, while the CDR3 is created

through the recombination of the variable, diversity and joining regions and thus includes part of the variable region, and all of the diversity (heavy chain only) and part of the joining regions. The CDR3 on the heavy chain (CDRH3) is the largest and most variable, often showing high levels of SHM and a wide range in length. It is also often involved in antigen binding (Wu*, et al.*, 1993).

Before leaving the bone marrow the immature B-cell goes through a checkpoint testing for autoreactivity (Gonzalez*, et al.*, 2011). If the B-cell receptor (BCR, created through V(D)J recombination) binds a self-antigen the B-cell is either deleted through apoptosis, or the BCR is edited by the assembly of a different light chain to the heavy chain or silenced in a constant state of unresponsiveness, known as anergy (Cambier*, et al.*, 2007; Gonzalez*, et al.*, 2011). If the B-cell does not bind a self-antigen it migrates out of the bone marrow into the spleen where it develops into a mature antigen naive B-cell expressing both IgM and IgD cell surface antigens (Tobon*, et al.*, 2013). In the spleen mature B-cells undergo a second round of autoreactivity testing (Gonzalez*, et al.*, 2011).

The rearrangement of the heavy and light chain V(D)J segments and the addition of nucleotides between the recombined segments creates a vast repertoire of B-cells (approximately 10^{13}) each expressing a unique membrane bound BCR enabling the adaptive immune response to recognise a wide range of antigens (Pieper*, et al.*, 2013). The affinity of the BCR to the antigen is further enhanced through somatic hypermutation (SHM), a process known as affinity maturation. In addition, different immune responses to a particular antigen are enabled through class-switch recombination (CSR or class-switching) of the constant regions.

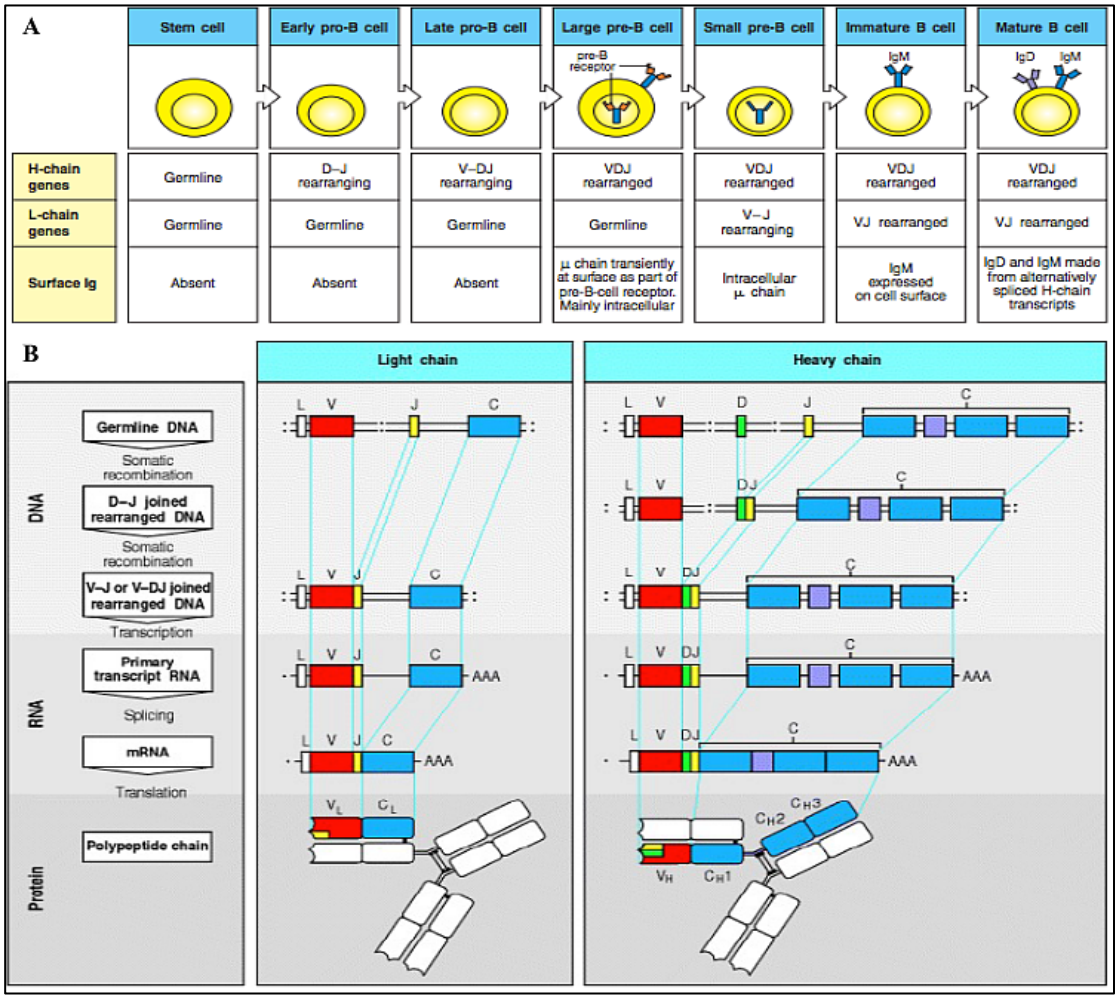


Figure 13: Early B-cell development and VDJ recombination

A) Early B-cell development. Shown are the different stages of B-cell development within the bone marrow. Figure taken from (Parham, 2005). B) VDJ recombination. Shown are the steps involved in somatic recombination of the V, (D) and J segments in the heavy and light chain immunoglobulin genes during B-cell development. Figure taken from (Janeway, *et al.*, 2001).

1.8.3. Antibody Maturation

Mature B-cells expressing cell surface IgM and IgD become activated following engagement with T helper follicular (T_{FH}) cells (Pone, *et al.*, 2010). In the case of HIV infection the loss of CD4 T helper has a profound effect on B cell activation which impacts on the quality of the antibody response (Binley, *et al.*, 2000). The activated B-cell then differentiates into short lived plasmablasts secreting IgM and/or low levels of IgD or proliferate and undergo affinity maturation and CSR (Pone, *et al.*, 2010). Affinity

maturation occurs in the dark zone of the germinal centre of lymph nodes through SHM, in this state the mature B-cells are known as centroblasts (Figure 14, (Klein*, et al.*, 2008)). SHM is a process by which single nucleotide polymorphisms are introduced into the variable region of the antibody structure (the part created through V(D)J recombination in both heavy and light chains) mediated by activation-induced deaminase (AID) (Klein*, et al.*, 2008). Further antibody diversity can be achieved through a process known as gene conversion. Gene conversion involves a non-reciprocal homologous recombination of an upstream variable gene to the recombined VDJ segment from one chromosome to another (Lavinder*, et al.*, 2014). This process, like SHM, is also mediated by AID.

Following SHM the centroblasts enter the light zone of the germinal centre as a centrocyte and undergo a process of selection, which is mediated by T_H-cells and follicular dendritic cells (Figure 14, (Klein*, et al.*, 2008). BCRs with increased affinity for the antigen will be preferentially selected by T_H-cells and receive a survival signal from follicular dendritic cells (Heesters*, et al.*, 2014). However, should the introduction of single nucleotide polymorphisms during SHM decrease the affinity of the BCR for the antigen, the centrocyte will undergo apoptosis. The high affinity centrocytes are then selected by T_H-cells, while centrocytes that are not selected by T-cells undergo apoptosis (Heesters*, et al.*, 2014). Approximately 90% of the selected centrocytes return to the dark zone of the germinal centre and continue SHM and the rest differentiate into memory B-cells or plasma cells, some of which undergo CSR (Klein*, et al.*, 2008; Wang*, et al.*, 2015b).

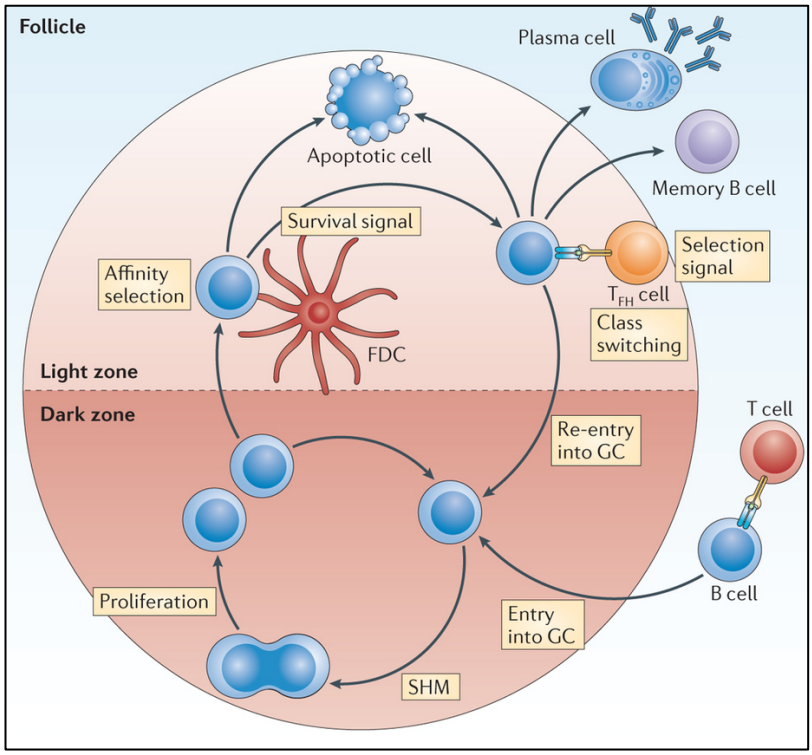


Figure 14: Somatic hypermutation and class-switch recombination in the germinal centre of lymph nodes

Shown are the processes of SHM and CSR in mature activated B-cells occurring in the germinal centres of lymph nodes. Figure taken from (Heesters, *et al.*, 2014).

CSR involves the different heavy chain constant region genes (μ, δ, γ, ε and α, which encode isotypes IgM, IgD, IgG, IgE and IgA, respectively) encoded on chromosome 14. The different isotypes, with the exception of Cδ (encoding IgD), are separated by a switch (S) region, consisting of tandemly repetitive unit sequences that contain many palindromes, and a intronic promoter (I) region (Figure 15, (Kinoshita, *et al.*, 2001)). Cytokines released from T_H-cells determines the isotype the centrocyte should switch to by activating the intronic promoter 5' of the S region of the isotype of choice (Kinoshita, *et al.*, 2001; King, 2009). IL-4 and IL-13 have been associated with class-switching to IgG4 and IgE, IL-10 with IgG1 and IgG3 while IL-10 in combination with TNF-β is associated with IgA (Tangye, *et al.*, 2002).

Once activated the same enzyme (AID) used in somatic hypermutation joins the two S regions between the current isotype and the isotype of choice and splices out the DNA in between the two regions (Figure 15, (Kinoshita, *et al.*, 2001)). This splicing event brings the coding region of the new isotype closer to the VDJ exon of the heavy chain and

leaves everything downstream of the new isotype intact (Kinoshita, *et al.*, 2001). Transcription and translation of the VDJ exon and the constant region results in the B-cell producing an antibody of a new isotype.

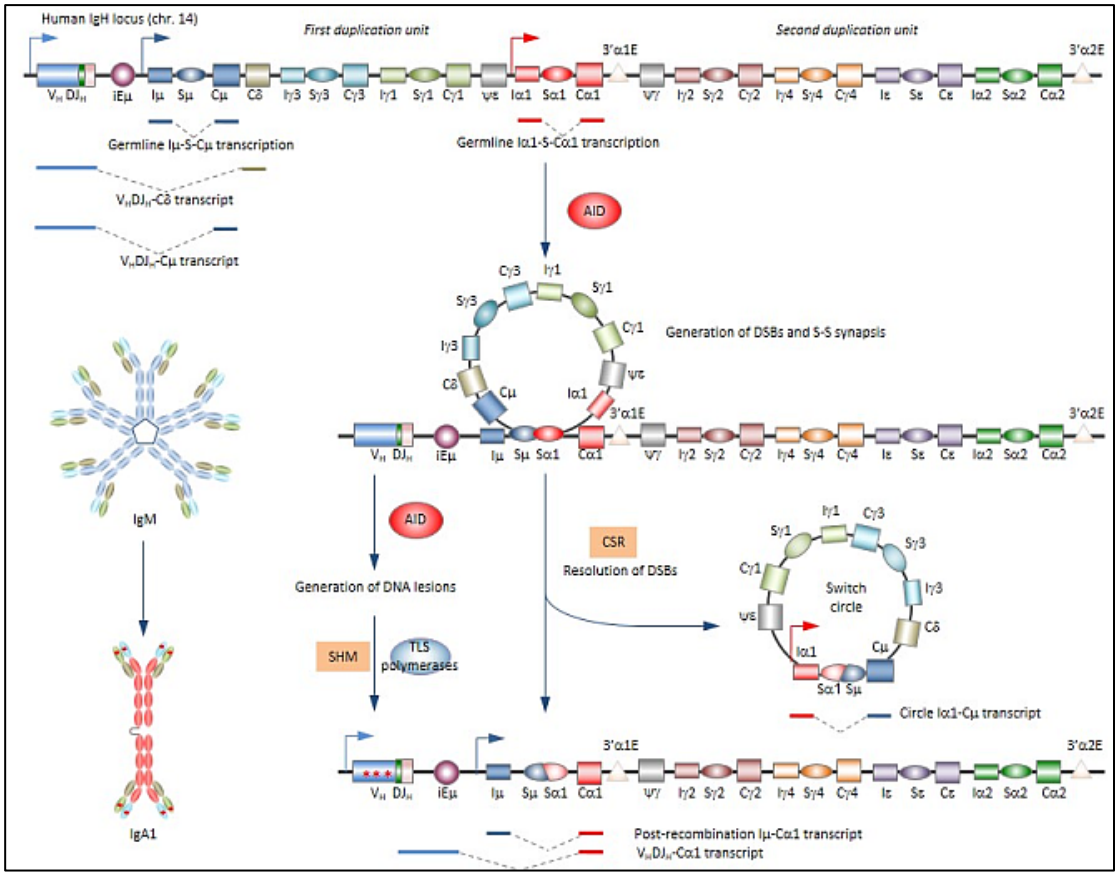


Figure 15: Process of class-switch recombination

Shown are steps involved in CSR that occur in the germinal centre of lymph nodes. Figure taken from (http://uthscsa.edu/micro-immunology/faculty/pclab/casalilab.asp).

1.8.4. Antibody Constant Region Effector Functions

IgM antibodies are the first to be produced and are often a marker of the acute immune response (Schroeder, *et al.*, 2010). They are initially produced as a cell surface antibody but following antigen stimulation are secreted as pentameric molecules held together by a J-chain, which facilitates secretion in mucosal surfaces. IgM antibodies usually have low affinity, due to their lack of maturation, but high avidity due to their pentameric structure.

These antibodies are involved in opsonisation, in which they coat an antigen thereby facilitating ingestion by macrophages and complement.

IgD antibodies are generally co-expressed with IgM on the cell surface of antigen naive mature B-cells, however low levels of circulating IgD molecules have been found in human serum (Schroeder*, et al.*, 2010). Their function in the immune response is largely unknown. Both IgM and IgD antibodies have a very short half-life, 5-8 days and 2.8 days, respectively (Vladutiu, 2000; Brekke*, et al.*, 2003).

IgG antibodies are the most commonly expressed isotype in human serum, the only isotype that crosses the placenta and has the longest half-life (~20 days) (Brekke*, et al.*, 2003). This class of antibodies is further subdivided into four subclasses: IgG1, IgG2, IgG3 and IgG4 and ranked according to their serum levels where IgG1 is the most abundant and IgG4 the least (Schroeder*, et al.*, 2010). All four subclasses, with the exception of IgG4, are able to activate complement. IgG1 and IgG3 generally bind protein antigens, while IgG2 and IgG4 tend to bind polysaccharide antigens. IgG antibodies are also involved in agglutination, antibody-dependent cellular-cytotoxicity (ADCC) and are able to neutralize toxins and viruses, where IgG3 has been shown to be particularly effective at neutralizing HIV (Schroeder*, et al.*, 2010).

IgA antibodies are the most abundant in mucosal surfaces and secretions such as breast milk and saliva where it is expressed as a dimer (Woof*, et al.*, 2005). They are involved in protection against toxins, bacteria and viruses in mucosal surfaces through direct neutralization or preventing binding to mucosal surfaces. IgA is generally not involved in activating complement but may be involved in ADCC. IgA is also found at low levels in the serum, usually as a monomer. There are two subclasses of IgA: IgA1 and IgA2, which differ at their hinge regions, where IgA1 is longer than IgA2 making it more sensitive to bacterial protease digestion (Woof*, et al.*, 2005). Thus, IgA1 predominates in

the serum, while IgA2 predominates in mucosal surfaces, such as the genital tract where bacterial infections are more frequent.

IgE is the least abundant antibody in human sera and has the shortest half-life (~2 days) (Stone*, et al.*, 2010). It is involved in hypersensitivity, allergic reactions and response to parasitic worms. IgE antibodies are highly potent due to their high affinity for the FcεRI on mast cells, basophils, Langerhans and eosinophil cells.

1.9. Antibody Repertoire Analysis and Next Generation Sequencing

In depth analysis of the antibody repertoire has been made possible with advances in next generation sequencing (NGS) technology such as Illumina MiSeq and HiSeq, Roche 454, Ion Torrent and PacBio. The depth of sequencing allowed by NGS enables one to sequence a greater representative of the ~10^{13} antibody repertoire, far greater than was previously achieved through Sanger sequencing. Additional advances in the sequencing chemistry have enabled greater read lengths covering more of the antibody sequence. One of the challenges of NGS, however, is correcting the errors created during sequencing as well as through PCR used to create the libraries that are being sequenced. Numerous bioinformatic tools have been developed, such as Sickle, BayesHammer and PANDAseq, specifically for Illumina MiSeq error correction (Schirmer*, et al.*, 2015). Other experimental techniques such as cell or sample barcoding and digital drop PCR have been used for error correction through consensus sequencing and to ensure the amplification of single molecules (Robinson, 2014). Techniques such as single-cell sorting and linkage-PCR have enabled the sequencing of paired heavy and light chain sequences from single B-cells (DeKosky*, et al.*, 2013). Sequencing fully functional heavy and light chains from single B-cells opens up opportunities for identifying potent antibody responses to diseases or infections that could be useful in vaccine development.

In HIV, antibody repertoire analysis has been used to identify antibodies against the virus as well as studying the evolution of antibodies during infection (Zhu, 2012; Liao*, et al.*, 2013; Zhu*, et al.*, 2013b; Zhu*, et al.*, 2013a; Doria-Rose*, et al.*, 2014; Wu*, et al.*, 2015). Two studies used NGS to elucidate the co-evolutionary pathway of the virus and bNAbs targeting the CD4bs and the V1V2 region of gp120 (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014), discussed in more detail below. Other studies have used NGS to pair heavy and light chain sequences of antibodies that are clonally related to bNAbs 10E8, PGT135-137 and PGT121 (Zhu, 2012; Zhu*, et al.*, 2013b; He*, et al.*, 2014), thereby enabling faster identification of potentially more potent and broad antibodies against HIV that may be useful in vaccine design.

1.10. Current Status of Vaccine Development

The development of an effective vaccine remains a high priority in the fight against the HIV pandemic. Traditionally vaccine design relies on the ability of the host during a natural infection to mount an immune response against the pathogen, clear the pathogen and protect the host from future infections through B-cell memory. The antigens that elicit the response are then mimicked in an immunogen designed to elicit the same natural host response. However, in HIV natural immune responses do not clear the virus nor do they protect against future infections, making vaccine design particularly complex.

Early HIV vaccine strategies focused on the traditional approach of initiating a humoral response (Wang*, et al.*, 2015a). Although antibodies were induced they were not broadly neutralizing and failed to protect against HIV infection. Based on these failures the strategy of vaccine design changed from humoral to cellular responses, with some support given by non-human primate models in which T-cell responses were associated with prolonged survival and decreased viral set-point (Hirsch*, et al.*, 1994; Shiver*, et al.*, 2002).

The STEP trial (or HVTN 502 trial), focused on T-cell immunity, however, this vaccine failed to protect against HIV infection or decrease viral loads and was stopped prematurely due to an apparent increase in susceptibility to infection in the vaccinated arm compared to the placebo arm (Casimiro*, et al.*, 2005).

The only human trial (RV144) that has shown some efficacy (~31.2%) was designed to elicit both T-cell and B-cell responses (Rerks-Ngarm*, et al.*, 2009). This trial was conducted in Thailand with a prime-boost strategy using a vCP1521 canarypox vector and a AIDSVAX B/E gp120 subunit. The correlates of protection in this trial were antibodies against the V1V2 region of gp120, re-igniting the importance of the humoral response (Haynes*, et al.*, 2012a).

A major HIV vaccine strategy is focused on the B-cell lineage immunogen design (Haynes*, et al.*, 2012b). This strategy relies on the identification/isolation of antibodies with a high affinity and ability to neutralize numerous strains of the virus (Haynes*, et al.*, 2012b). Using the knowledge of the antibody heavy and light chains the co-evolution of the antibody lineage and virus during natural infection is studied through deep sequencing. The aim of this approach is to identify the unmutated common ancestor that was activated and gave rise to the antibody lineage as well as the viral strain that activated the lineage. Two longitudinal studies have observed the co-evolution of broadly neutralizing antibodies targeting the CD4bs and V1V2 regions of the HIV envelope (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014). The study involving CD4bs antibody CH103, revealed that modest levels of SHM (13-15%) accumulated over years of infection was required for broadly neutralizing activity and that it had an average CDRH3 length (Liao*, et al.*, 2013). However, this lineage only achieved ~50% breadth. Often CD4bs bNAbs have high levels of SHM, such as VRC01 (30-34%) (Wu*, et al.*, 2015). Conversely CAP256.VRC26, which targets the V1V2 region, required extremely long CDRH3 lengths (35aa) and had average

SHM (8-15%) (Doria-Rose*, et al.*, 2014). The UCA antibody sequence from the CD4bs bNAb lineage was able to bind but not neutralize the transmitted founder virus of the respective donors, whereas the UCA of the CAP256.VRC26 lineage could both bind and neutralize the transmitted founder virus (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014; Wu*, et al.*, 2015). These studies have suggested that immunization with a similar set of viral envelopes that resulted in breadth could be a useful vaccine strategy.

2. STUDY OBJECTIVES

Broadly neutralizing antibodies against HIV-1 are able to neutralize various strains of the virus and are likely to be a correlate of protection in an HIV vaccine. However, only ~20% of individuals infected with HIV-1 develop these types of antibodies and only after years of infection. The mechanisms by which these antibodies develop remain poorly understood. The aim of this project was to study host aspects associated with the development broadly neutralizing antibodies, specifically:

To investigate the germline IGHV gene profiles in individuals who do and do not develop broadly neutralizing antibodies during HIV-1 infection

The germline IGHV gene encodes the heavy chain variable region of rearranged antibodies, which is central to antigen binding. Studies on broadly neutralizing antibodies have shown biases in germline IGHV gene usage. Our hypothesis was that individuals who develop broadly neutralizing antibodies have distinct germline IGHV repertoires. Using Roche 454 and Illumina MiSeq next generation sequencing technologies, the germline IGHV repertoires of 28 South African women, including 13 who developed broadly neutralizing antibodies and 13 who did not despite chronic HIV infection were analysed. This project is described in Section B, Chapter 1.

To assess strain-specific antibody evolution during HIV-1 infection

Broadly neutralizing antibodies develop from strain-specific antibodies that arise within the first few months of infection and acquire neutralization breadth through somatic hypermutation. Our aim was to study the evolution of a strain-specific antibody (CAP88-

CH06) lineage in a donor (CAP88) who failed to develop neutralization breadth to determine why some strain-specific antibodies do not become broad. Using Illumina MiSeq technology the heavy chain antibody repertoire of CAP88 was sequenced at seven time points (5, 8, 11, 15, 34, 38 and 121 weeks post-infection (w.p.i)) and used to trace the CAP88-CH06 lineage. We hypothesized that this lineage failed to become broadly neutralizing due to limited somatic hypermutation and evolution. This project is described in Section B, Chapter 2.

To determine the glycan-binding profiles of HIV antibodies

The envelope of HIV-1 is covered with glycans that protect vulnerable sites from the host antibody response. However, these glycans are often the target of broadly neutralizing antibodies. Our aim was to determine if increases in glycan-binding antibodies are a response to HIV infection or unique to individuals who develop broadly neutralizing antibodies. Glycan microarrays were used to determine the longitudinal glycan-binding profiles of serum IgG antibodies from 47 women, 27 of whom were HIV-positive and 20 HIV-negative. The HIV-negative samples were used as controls to determine natural fluctuations in glycan-binding over-time. Of the 27 HIV-positive women, 12 developed bNAbs and 13 did not develop broadly neutralizing antibodies despite chronic HIV infection. We hypothesized that women with broadly neutralizing antibodies would have higher levels of glycan-binding antibodies than women who did not. This project is described in Section B, Chapter 3.

3. METHODS SUMMARY

3.1. Samples

Serum and peripheral blood mononuclear cells (PBMCs) were obtained from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 and 004 cohorts (Van Loggerenberg*, et al.*, 2008; Abdool Karim*, et al.*, 2010). These cohorts included adult women (>18 years of age) of African ancestry with a high risk of HIV infection who have been followed throughout infection, and in some cases prior to infection, in urban and rural sites in KwaZulu-Natal, South Africa. All of the women used in this study were infected with HIV-1 subtype C.

We focused on women who developed bNAbs (referred to as BCN individuals) and a control group of women who do not (referred to as non-BCN individuals). A panel of 18 viruses were used to test for neutralization breadth within these women. The panel consisted of 6 subtype A, 6 subtype B and 6 subtype C viruses. BCN individuals were defined as those whose serum was capable of neutralizing 30%-100% of the panel from 2 years post-infection, while non-BCN individuals were only able to neutralized 0%-11%. Since high viral loads at 6 months post-infection have been associated with the development of neutralization breadth (Gray*, et al.*, 2011b), BCN and non-BCN individuals were matched according to viral loads. Longitudinal serum samples from pre-infection and yearly for three years from 47 CAPRISA women, involving 20 HIV-negative women and 27 HIV-positive women were used in Chapter 3 (Section B).

PBMCs were used to extract total nucleic acids (RNA and DNA) from 28 CAPRISA participants used in Chapter 1 (Section B) and from CAP88 used in Chapter 2

(Section B). A single time point was used for each participant in Chapter 1 (Section B). While longitudinal PBMC samples were taken from 5, 8, 11, 17, 34, 38 and 121 weeks post-infection from CAP88, used in Chapter 2 (Section B).

Several monoclonal antibodies (mAbs) have been isolated from CAPRISA participants at the Centre for HIV and STI, HIV Virology Section at the National Institute of Communicable Diseases (NICD). Sequences from these mAbs as well as sequences from bNAbs, obtained from CATNAP (CATNAP, 2014), were used in Chapter 1 (Section B). The bNAbs used in Chapter 3 (Section B) were obtained through the AIDS Reagent Program (NIH AIDS Reagent, 2013) and IAVI's Neutralizing Antibody Center (IAVI, 2013).

3.2. Nucleic Acid extraction and cDNA Synthesis

Genomic DNA was extracted from PBMCs in Chapter 1 (Section B) using a Promega Wizard Genomic purification kit, while total nucleic acids (RNA and DNA) were extracted from PBMCs using a Qiagen AllPrep DNA/RNA mini kit for Chapter 2 (Section B). The quality and quantity of the resulting RNA and DNA extractions were checked on a Nano spectrophotometer (Section B, Chapters 1 and 2). RNA was converted to cDNA using Random Hexamers, Invitrogen First Strand Buffer, DTT, RNAseOUT and Superscript III RT Enzyme (described in Section B, Chapter 2).

3.3. Primer Design

Novel primers were designed to amplify all seven germline IGHV subgroups in Chapter 1 (Section B), based on the sequences of all known IMGT (IMGT, 2014) listed IGHV genes. In addition to the gene-specific sequence the primers were adapted for both 454 and MiSeq

sequencing. The 454 primers included forward and reverse adapters, tag sequences and four unique MID sequences as recommended by Roche for pooling four samples per run. The MiSeq primers included index binding tags and read sequences for paired-end reads, described in more detail in Section B, Chapter 1.

One IGHV4 forward primer and two reverse primers (IgG and IgA) were designed to amplify the CAP88-CH06 antibody lineage from CAP88. These primers were also adapted for MiSeq sequencing with the addition of index binding tags and read sequences for paired-end sequencing. This is described in more detail in Section B, Chapter 2.

3.4. Library Construction for Next Generation Sequencing

All 7 IGHV subgroups were PCR amplified 3x from gDNA for each of the 28 CAPRISA individuals used in Chapter 1 (Section B). All replicates from each subgroup were pooled to create a full germline IGHV repertoire per individual. For 454 sequencing, four individuals were pooled per sequencing run using MID tags encoded into the primer sequences to distinguish each individual from the sequencing data. For MiSeq sequencing 13 individuals along with some replicates from the 454 sequencing were pooled and individuals distinguished using the Nextera indexing kit to add indexes to the PCR products prior to pooling and sequencing. This is described in more detail in Section B, Chapter 1.

The CAP88-CH06 antibody lineage was PCR amplified 10x from cDNA from each of the 7 time points (5, 8, 11, 17, 34, 38 and 121 w.p.i) from CAP88. Each replicate was pooled per time point and a Nextera indexing kit used to add indexes to the PCR products from each time point. All time points were pooled for sequencing on the Illumina MiSeq. This is described in more detail in Section B, Chapter 2.

3.5. Next Generation Sequencing

Next generation sequencing (NGS) was used in Chapters 1 and 2 (Section B). In Chapter 1, both Roche 454 and Illumina MiSeq technologies were used to elucidate the germline IGHV repertoires in 28 South African women (described in more detail in Section B, Chapter 1). In Chapter 2, Illumina MiSeq was used to sequence the CAP88-CH06 antibody lineage during infection (described in more detail in Section B, Chapter 2). Both studies used a custom amplicon approach.

3.6. Sequence Data Analysis

Sequence data analysis in Chapter 1 (Section B) was done in collaboration with Dr Ram Shrestha, Imogen Wright and Professor Simon Travers from the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape and Dr Katherine Jackson from Stanford University in California, USA. Sequences obtained from the Roche 454 and Illumina MiSeq were separated based on the quality of the base calls and sequence length. High quality sequences were compared to a database of known germline IGHV alleles, obtained from the global Immunogenetics Database, IMGT (IMGT, 2014). Sequences were classified as IMGT alleles if they had perfect matches to IMGT listed germline IGHV alleles; non-IMGT alleles if they had mismatches to IMGT listed germline IGHV alleles (due to SNPs or indels) but matched non-IMGT immunogenetics databases or immunogenetics publications; and novel if they had no matches to either IMGT or non-IMGT alleles and were observed in multiple individuals or across both 454 and MiSeq in one individual. The repertoires between BCN and non-BCN individuals were compared. The sequences from mAbs isolated in house and well characterised bNAbs, obtained from CATNAP (CATNAP, 2014) were used to identify germline gene usage in functional antibodies (described in more detail in Section B, Chapter 1).

Sequence data analysis in Chapter 2 (Section B) was done in collaboration with Dr Zizhang Sheng, Dr Chaim Schramm and Professor Lawrence Shapiro at Columbia University in New York, USA. Sequences were separated according to the quality of the base calls and sequence length and adjusted for sequencing and PCR error. Using an antibodyomics pipeline, developed at Professor Shapiro's lab, high quality sequences were assigned V and J genes and compared to the sequence of the known strain-specific antibody (CAP88-CH06). Sequences with the same IGHV gene (IGHV4-39), IGHJ gene (IGHJ4) and ≥ 80% identity to the CDRH3 CAP88-CH06 were selected as clonally related. These clonally related sequences were then used to construct phylogenetic trees of the CAP88-CH06 lineage (described in more detail in Section B, Chapter 2).

3.7. Glycan Microarrays

Glycan microarrays were used in Chapter 3 (Section B). The printing and running of the glycan microarrays were done in collaboration with Sudipa Chowdhury, Shea Wright and Dr Jeff Gildersleeve at the National Cancer Institute (NCI) in Frederick, Maryland, USA. Microarrays containing 337 components were spotted in duplicate onto glass slides using a MicroGrid II array printer. Each glass slide contained 16 arrays to allow pooling of samples per slide. Each slide was incubated with serum and DyLight^{TM}459 labelled anti-human IgG and scanned on a Genepix microarray scanner (described in more detail in Section B, Chapter 3).

3.8. Glycan Microarray Data Analysis

Fluorescence intensities for each glycan array used in Chapter 3 (Section B) were calculated using the Genepix Pro 6.0 software. Average intensities were calculated for

duplicate spots and data was normalized to reference sera values. Fluorescence intensities were log transformed (log_2) and used to determine the level of IgG binding for each glycan per sample. bNAbs were used to determine the array's ability to detect HIV-binding. Longitudinal HIV-negative sera were used as controls to determine natural levels in glycan-binding fluctuations over a three-year time frame. Longitudinal sera from HIV-positive individuals (including pre-infection samples) were used to establish baseline values. Comparisons between HIV-positive individuals who develop neutralization breadth (BCN) and HIV-positive individuals who do not develop neutralization breadth (non-BCN) despite equivalent antigenic load were used to determine if increases in glycan-binding antibodies were a response to HIV infection or unique to BCN individuals. A gp120 competition assay, in which sera was incubated with gp120 or PBS prior to glycan array binding, was used to determine if glycan-binding detected on the arrays were HIV-specific (described in more detail in Section B, Chapter 3).

3.9. Statistical Analysis

Two-tailed Fisher's exact tests with 95% confidence intervals were used to test the difference in germline IGHV gene repertoires between the BCN and non-BCN sample groups (Section B, Chapter 1), compare the glycan-binding profiles of gp120 and PBS incubated sera and the glycan-binding profiles of BCN and non-BCN individuals (Section B Chapter 3). Fold-change differences were used to assess fluctuations in glycan-binding over time in the HIV-negative and positive sera (Section B, Chapter 3). Statistical analysis was not used in Chapter 2 (Section B).

4. REFERENCES

Abdool Karim, Q*, et al.* (2010) Effectiveness and safety of Tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science,* 329 (5996): 1168-1174

Amin, MN*, et al.* (2013) Synthetic glycopeptides reveal the glycan specificity of HIV-neutralizing antibodies. *Nature Chemical Biology,* 9 (8): 521-526

An, P and Winkler, CA (2010) Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in Genetics,* 26 (3): 119-131

Arts, EJ and Hazuda, DJ (2012) HIV-1 antiretroviral drug therapy. *Cold Spring Harbor Perspectives in Medicine,* 2 (4): a007161

Astronomo, R*, et al.* (2008) A glycoconjugate antigen based on the recognition motif of a broadly neutralizing human immunodeficiency virus antibody, 2G12, is immunogenic but elicits antibodies unable to bind to the self glycans of gp120. *Journal of Virology,* 82 (13): 6359-6368

Barouch, D*, et al.* (2013) Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature,* 503: 224-228

Barre-Sinoussi, F*, et al.* (2013) Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology,* 11 (12): 877-883

Bartesaghi, A*, et al.* (2013) Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nature Structural and Molecular Biology,* 20: 1352-1357

Binley, J*, et al.* (2000) The relationship between T cell proliferative responses and plasma viremia during treatment of Human Immunodeficiency Virus type 1 infection with combination antiretroviral therapy. *The Journal of Infectious Diseases,* 181: 1249-1263

Binley, J*, et al.* (2004) Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *Journal of Virology,* 78 (23): 13232-13252

Bonomelli, C*, et al.* (2011) The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One,* 6 (8): e23521

Brekke, OH and Sandlie, I (2003) Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nature Reviews Drug Discovery,* 2 (1): 52-62

Buchacher, A*, et al.* (1994) Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Research and Human Retroviruses,* 10 (4): 359-369

Buonaguro, L*, et al.* (2007) Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of Virology,* 81 (19): 10209-10219

Calarese, D*, et al.* (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science,* 300 (5628): 2065-2071

Cambier, JC*, et al.* (2007) B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nature Reviews Immunology,* 7 (8): 633-643

Campbell, C*, et al.* (2013) High-throughput profiling of anti-glycan humoral responses to SIV vaccination and challenge. *PLoS One,* 8 (9): e75302

Casimiro, DR*, et al.* (2005) Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with DNA and recombinant adenoviral vaccine vectors expressing Gag. *Journal of Virology,* 79 (24): 15547-15555

Chun, T-W and Fauci, AC (1999) Latent reservoirs of HIV: obstacles to the eradication of virus. *Proceedings of the National Academy of Sciences,* 96: 10958-10961
Cohen, J (2013) Early treatment may have cured infant of HIV infection. *Science,* 339: 1134
DeKosky, BJ*, et al.* (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology,* 31 (2): 166-169
Doores, KJ*, et al.* (2010) Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences,* 107 (31): 13800-13805
Doores, KJ*, et al.* (2015) Two classes of broadly neutralizing antibodies within a single lineage directed to the high-mannose patch of HIV envelope. *Journal of Virology,* 89 (2): 1105-1118
Doria-Rose, NA*, et al.* (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature,* 509 (7498): 55-62
Doria-Rose, NA*., et al.* (2015) A new member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. *Journal of Virology*: In Press
Dunlop, DC*, et al.* (2010) Polysaccharide mimicry of the epitope of the broadly neutralizing anti-HIV antibody, 2G12, induces enhanced antibody responses to self oligomannose glycans. *Glycobiology,* 20 (7): 812-823
Engelman, A and Cherepanov, P (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology,* 10 (4): 279-290
Enriquez, M and McKinsey, DS (2011) Strategies to improve HIV treatment adherence in developed countries: clinical management at the individual level. *HIV/AIDS,* 3: 45-51
Euler, Z and Schuitemaker, H (2012) Cross-reactive broadly neutralizing antibodies: timing is everything. *Frontiers in Immunology,* 3: 215
Falkowska, E*, et al.* (2014) Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity,* 40 (5): 657-668
Fauci, AS*, et al.* (2013) HIV-AIDS: much accomplished, much to do. *Nature Immunology,* 14 (11): 1104-1107
Frankel, AD and Young, JAT (1998) HIV-1: Fifteen proteins and an RNA. *Annual Reviews in Biochemistry,* 67: 1-25
Go, EP*, et al.* (2013) Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *Journal of Proteome Research,* 12 (3): 1223-1234
Gonzalez, SF*, et al.* (2011) Trafficking of B cell antigen in lymph nodes. *Annual Review of Immunology,* 29: 215-233
Gray, ES*, et al.* (2007) Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *Journal of Virology,* 81 (12): 6187-6196
Gray, ES*, et al.* (2011b) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology,* 85 (10): 4828-4840
Haynes, BF*, et al.* (2012a) Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine,* 366 (14): 1275-1286
Haynes, BF*, et al.* (2012b) B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnology,* 30 (5): 423-433
He, L*, et al.* (2014) Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Scientific Reports,* 4: 6778
Heesters, BA*, et al.* (2014) Follicular dendritic cells: dynamic antigen libraries. *Nature Reviews Immunology,* 14 (7): 495-504
Hirsch, VM*, et al.* (1994) Prolonged clinical latency and survival of macaques given a whole inactivated simian immunodefiniciency virus vaccine. *Journal of Infectious Diseases,* 170 (1): 51-59
Huang, J*, et al.* (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature,* 491 (7424): 406-412

Huang, J, *et al.* (2014) Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature,* 515 (7525): 138-142
Julien, J-P, *et al.* (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science,* 342 (6165): 1477-1483
King, C (2009) New insights into the differentiation and function of T follicular helper cells. *Nature Reviews Immunology,* 9 (11): 757-766
Kinoshita, K and Honjo, T (2001) Linking class-switch recombination with somatic hypermutation. *Nature Reviews Immunology,* 2: 493-503
Klein, U and Dalla-Favera, R (2008) Germinal centres: role in B-cell physiology and malignancy. *Nature Reviews Immunology,* 8 (1): 22-33
Ko, S-Y, *et al.* (2014) Enhanced neonatal Fc receptor function improves protection against primate SHIV infection. *Nature,* 514 (7524): 642-645
Kong, L, *et al.* (2013) Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature Structural and Molecular Biology,* 20 (7): 796-803
Lavinder, JJ, *et al.* (2014) Systematic characterization and comparative analysis of the rabbit immunoglobulin repertoire. *PLoS One,* 9 (6): e101322
Lefranc, M-P, *et al.* (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research,* 37 (Database Issue): D1006-D1012
Liao, H-X, *et al.* (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature,* 496 (7446): 469-476
Liu, R, *et al.* (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell,* 86: 367-377
Lyumkis, D, *et al.* (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science,* 342 (6165): 1484-1490
Market, E and Papavasiliou, FN (2003) V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biology,* 1 (1): E16
McLellan, J, *et al.* (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature,* 480 (7377): 336-343
Mlisana, K, *et al.* (2014) Rapid disease progression in HIV-1 subtype C-infected South African women. *Clinical Infectious Diseases,* 59 (9): 1322-1331
Moir, S and Fauci, AS (2009) B cells in HIV infection and disease. *Nature Reviews Immunology,* 9 (4): 235-245
Moldt, B, *et al.* (2012) Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. *Proceedings of the National Academy of Sciences,* 109 (46): 18921-18925
Moore, PL, *et al.* (2015) Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends in Microbiology,* 23 (4): 204-211
Mouquet, H, *et al.* (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proceedings of the National Academy of Sciences,* 109 (47): E3268-E3277
Pancera, M, *et al.* (2013) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature,* 514 (7523): 455-461
Pieper, K, *et al.* (2013) B-cell biology and development. *Journal of Allergy and Clinical Immunology,* 131 (4): 959-971
Pone, EJ, *et al.* (2010) Toll-like receptors and B-cell receptors synergize to induce immunoglobulin class-switch DNA recombination: relevance to microbial antibody responses. *Critical Reviews in Immunology,* 30 (1): 1-29
Rerks-Ngarm, S, *et al.* (2009) Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine,* 361 (23): 2209-2220
Robinson, WH (2014) Sequencing the functional antibody repertoire-diagnostic and therapeutic discovery. *Nature Reviews Rheumatology,* 11: 171-182
Samson, M, *et al.* (1996) Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature,* 382 (6593): 722-725
Saunders, KO, *et al.* (2015) Sustained delivery of a broadly neutralizing antibody in nonhuman primates confers long-term protection against simian/human immunodeficiency virus infection. *Journal of Virology,* 89 (11): 5895-5903

Scanlan, C*, et al.* (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of α 1→2 mannose residues on the outer face of gp120. *Journal of Virology,* 76 (14): 7306-7321
Scharf, L*, et al.* (2014) Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports,* 7 (3): 785-795
Schirmer, M*, et al.* (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research,* 43 (6): e37
Schroeder, HW, Jr. and Cavacini, L (2010) Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology,* 125 (2 Suppl 2): S41-S52
Sharp, PM and Hahn, BH (2011) Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine,* 1 (1): a006841
Shivatare, SS*, et al.* (2013) Efficient convergent synthesis of bi-, tri-, and tetra-antennary complex type N-glycans and their HIV-1 antigenicity. *Journal of the American Chemical Society,* 135 (41): 15382-15391
Shiver, JW*, et al.* (2002) Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature,* 415: 331-335
Stone, KD*, et al.* (2010) IgE, mast cells, basophils, and eosinophils. *Journal of Allergy and Clinical Immunology,* 125 (2 Suppl 2): S73-S80
Tangye, SG*, et al.* (2002) Isotype switching by human B cells is division-associated and regulated by cytokines. *The Journal of Immunology,* 169 (8): 4298-4306
Taylor, BS*, et al.* (2008) The challenge of HIV-1 subtype diversity. *New England Journal of Medicine,* 358 (15): 1590-1602
Tobon, GJ*, et al.* (2013) B lymphocytes: development, tolerance, and their role in autoimmunity-focus on systemic lupus erythematosus. *Autoimmune Diseases,* 2013: 827254
Tomaras, GD*, et al.* (2008) Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology,* 82 (24): 12449-12463
Van Gent, D*, et al.* (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews Genetics,* 2: 196-206
Van Loggerenberg, F*, et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One,* 3 (4): e1954
Vladutiu, AO (2000) Immunoglobulin D: properties, measurement and clinical relevance *Clinical and Diagnostic Laboratory Immunology,* 7 (2): 131-140
Walker, LM*, et al.* (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science,* 326 (5950): 285-289
Walker, LM*, et al.* (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477 (7365): 466-470
Wang, HB*, et al.* (2015a) HIV vaccine research: the challenge and the way forward. *Journal of Immunology Research,* 2015: 503978
Wang, L-X (2013) Synthetic carbohydrate antigens for HIV vaccine design. *Current Opinion in Chemical Biology,* 17 (6): 997-1005
Wang, S*, et al.* (2015b) Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies. *Cell,* 160 (4): 785-797
Watson, CT and Breden, F (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes and Immunity,* 13 (5): 363-373
Watson, CT*, et al.* (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *American Journal of Human Genetics,* 92 (4): 530-546
West, AP*, et al.* (2012) Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proceedings of the National Academy of Sciences,* 109 (30): E2083–E2090
West, AP*, et al.* (2014) Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell,* 156 (4): 633-648

Wibmer, CK*, et al.* (2015) HIV broadly neutralizing antibody targets. *Current Opinion in HIV and AIDS,* 10: 1-8
Woof, JM and Mestecky, J (2005) Mucosal immunoglobulins. *Immunological Reviews,* 206: 64-82
Wu, TT*, et al.* (1993) Length distribution of CDRH3 in antibodies. *Proteins: Structure, Function and Genetics,* 16: 1-7
Wu, X*, et al.* (2015) Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell,* 161: 1-16
Wyatt, R and Sodroski, J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science,* 280: 1884-1888
Zhu, J (2012) Somatic populations of PGT135–137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatic. *Frontiers in Microbiology,* 3: 315
Zhu, J*, et al.* (2013a) De novo identification of VRC01 class HIV-1–neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proceedings of the National Academy of Sciences,* 110 (43): E4088-E4097
Zhu, J*, et al.* (2013b) Mining the antibodyome for HIV-1–neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proceedings of the National Academy of Sciences,* 110 (16): 6470-6475
Zhu, X*, et al.* (2000) Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry,* 39 (37): 11194-11204
Zolla-Pazner, S (2014) A critical question for HIV vaccine development: which antibodies to induce? *Science,* 345 (6193): 167-168
Zwick, MB*, et al.* (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *Journal of Virology,* 75 (22): 10892-10905

Books

Janeway, C*, et al.* (2001) Immunobiology: the immune system in health and disease. New York: Garland Science. ISBN: 081533642
Parham, P (2005) The immune system. Garland Science. ISBN: 0815340931
Parham, P (2009) The immune system. Garland Science. ISBN: 9780815341468

Online Databases

CATNAP, HIV sequence database [Online] Available: http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp [Accessed: June 2014]
CFG - Consortium for Fuctional Glycomics [Online] Available: http://www.functionalglycomics.org/ [Accessed: December 2014]
IAVI's neutralizing antibody center [Online] Available: http://www.scripps.edu/research/nac/ [Accessed: February 2013]
IMGT - The international immunogenetics information system [Online] Available: http://www.imgt.org/ [Accessed: June 2014]
Medscape: HIV-1 Clades [Online] Available: http://img.medscape.com/article/708/915/708915-fig1.jpg
NIH AIDS Reagent Program [Online] Available: www.aidsreagent.org [Accessed: February 2013]
UNAIDS global report on the global AIDS epidemic 2013 [Online] Available: http://www.unaids.org/
World Health Organization (WHO) HIV/AIDS factsheet [Online] Available: http://www.who.int/mediacentre/factsheets/fs360/en/

SECTION B: RESULTS

1. ABILITY TO DEVELOP BROADLY NEUTRALIZING HIV-1 ANTIBODIES IS NOT RESTRICTED BY THE GERMLINE IMMUNOGLOBULIN GENE REPERTOIRE[1]

Cathrine Scheepers[*,†], Ram K. Shrestha[‡], Bronwen E. Lambson[*,†], Katherine J. L. Jackson[§,¶], Imogen A. Wright[‡], Dshanta Naicker[*], Mark Goosen[*], Leigh Berrie[*], Arshad Ismail[*], Nigel Garrett[#,], Quarraisha Abdool Karim[#,††], Salim S. Abdool Karim[#,††], Penny L. Moore[*,†,#], Simon A. Travers[‡] and Lynn Morris[*,†,#]**

[*]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, 2131 South Africa; [†]School of Pathology, Division of Virology and Communicable Disease Surveillance, University of the Witwatersrand, Johannesburg, 2050, South Africa; [‡]South African National Bioinformatics Institute (SANBI), South African Medical Research Council Bioinformatics Unit, University of the Western Cape, Bellville 7535, South Africa; [§]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia; [¶]Department of Pathology, School of Medicine, Stanford University, Stanford, CA 94305, USA; [#]Centre for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu-Natal, 4013 South Africa; [**]Department of Infectious Diseases, Nelson R. Mandela School of Medicine, University of KwaZulu–Natal, 4041, Durban, South Africa; and [††]Department of Epidemiology, Columbia University, New York City, 10032, USA

Journal of Immunology, 194(9): 4371-4378 (Attached in the Appendix)

---

[1] This project was supported by the Poliomyelitis Research Foundation (PRF), University of the Witwatersrand Health Sciences Faculty Research Council and the National Research Foundation (NRF). Cathrine Scheepers was supported by the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) through the Fogarty International Center, National Institutes of Health (grant # 5 D43 TW000231). CAPRISA is funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes for Health (NIH), and U.S. Department of Health and Human Services (grant: AI51794). SANBI received funding from the South African Department of Science and Technology, South African National Research Foundation DAAD Study Bursary, Atlantic Philanthropies and the South African Medical Research Council. Penny L. Moore is a Wellcome Trust Intermediate Fellow in Public Health and Tropical Medicine (Grant 089933/Z/09/Z).

1.1. ABSTRACT

The human immunoglobulin repertoire is vast, producing billions of unique antibodies from a limited number of germline immunoglobulin genes. The immunoglobulin heavy chain variable region (IGHV) is central to antigen binding and is comprised of 48 functional genes. Here we analysed whether HIV-1 infected individuals who develop broadly neutralizing antibodies show a distinctive germline IGHV profile. Using both 454 and Illumina technologies we sequenced the IGHV repertoire of 28 HIV-infected South African women from the CAPRISA 002 and 004 cohorts, 13 of whom developed broadly neutralizing antibodies. Of the 259 IGHV alleles identified in this study, approximately half were not found in IMGT. This included 85 entirely novel alleles and 38 alleles that matched rearranged sequences in non-IMGT databases. Analysis of the rearranged VH genes of mAbs isolated from 7 of the CAPRISA women and previously isolated broadly neutralizing antibodies from other donors provided evidence that at least 8 novel or non-IMGT alleles contributed to functional antibodies. Importantly, we found that despite a wide range in the number of IGHV alleles in each individual, including alleles used by known broadly neutralizing antibodies, there were no significant differences in germline IGHV repertoires between individuals who do and do not develop broadly neutralizing antibodies. This study reports novel IGHV repertoires and highlights the importance of a fully comprehensive immunoglobulin database for germline gene usage prediction. Furthermore, these data suggest a lack of genetic bias in broadly neutralizing antibody development in HIV-1 infection, with positive implications for HIV vaccine design.

1.2. INTRODUCTION

The induction of broadly neutralizing antibodies (bNAbs) is likely to be crucial for an efficacious HIV vaccine. While the majority of chronically HIV-infected individuals develop some level of cross-neutralizing activity (Hraber, *et al.*, 2014), bNAbs are generally found in less than 20% of HIV-infected individuals (Stamatatos, *et al.*, 2009). The mechanisms underlying bNAb emergence are largely unknown, but a better understanding of how these antibodies arise in natural infection would provide a blueprint for a vaccine designed to elicit them. Over the last few years a large number of potent and broad bNAbs have been isolated from selected HIV-infected donors. These bNAbs target conserved epitopes on the HIV envelope including the membrane proximal external region (MPER), V2 glycans, V3 glycans, CD4 binding site (CD4bs) and the gp120/gp41 interface (Kwong, *et al.*, 2012; Scharf, *et al.*, 2014). However, most bNAbs have unusual genetic features including high levels of somatic hypermutation (SHM), long CDRH3s (the complementarity determining region 3 on the heavy chain) (Breden, *et al.*, 2011; Kwong, *et al.*, 2012) and for some classes, biases in germline IGHV gene usage (Gorny, *et al.*, 2009; Breden, *et al.*, 2011; Gorny, *et al.*, 2011; Gorny, *et al.*, 2012; Kwong, *et al.*, 2012; Jardine, *et al.*, 2013).

The VRC01 class of antibodies to the CD4bs (including NIH45-46, 12A12, 3BNC117, VRC-PG04 and VRC-CH31 isolated from multiple donors) have been shown to preferentially use either IGHV1-2*02 or IGHV1-46*02 germline alleles (Scheid, *et al.*, 2011; Kwong, *et al.*, 2012; West, *et al.*, 2012). This preference is thought to be due to the electrostatic and hydrophobic contacts afforded by conserved residues in framework 1 (FR1), CDRH2 and FR3 of this allele (some of these features are conserved in IGHV1-

46*02 and IGHV1-3*01 which are also used by CD4bs antibodies) (Scheid, *et al.*, 2011; Jardine, *et al.*, 2013). Anti-HIV antibodies frequently use IGHV1 genes compared to healthy individuals (Gorny, *et al.*, 2009; Breden, *et al.*, 2011). In particular, IGHV1-69 is used by mAbs that target V2, the CD4 induced site (CD4i) and gp41 in HIV infection as well as other viral infections such as influenza (Corti, *et al.*, 2010; Gorny, *et al.*, 2012). The preference for this particular gene can be attributed to the interaction of hydrophobic residues in the CDRH2 with helical elements or hydrophobic β-sheets like those found on the hemagglutinin (HA) of influenza, and gp41 and gp120 of the HIV envelope (Gorny, *et al.*, 2012; Lingwood, *et al.*, 2012). The use of IGHV5-51*01 and IGHV5-51*03, which are underrepresented in mature antibodies, has also been reported to be favoured by anti-V3 HIV antibodies (Gorny, *et al.*, 2009; Gorny, *et al.*, 2011).

The human germline IGHV repertoire consists of 7 IGHV subgroups, which are described in the International Immunogenetics Database (IMGT, 2014). These IGHV1-7 subgroups include functional genes, open-reading frames (ORF) and pseudogenes with only functional genes being involved in antibody production. IGHV3 is the largest of the IGHV subgroups, with 21 functional genes, followed by IGHV1 and IGHV4 both with 10 and IGHV2, IGHV5, IGHV6 and IGHV7 with three or less genes (Lefranc, 2001). Most of these genes have multiple alleles, including functional and non-functional alleles, which differ by either a single nucleotide polymorphism (SNP) or by multiple SNPs, which can be either synonymous or nonsynonymous, or frameshift mutations caused by indels that contribute diversity to the immunoglobulin gene repertoire. Furthermore, whole IGHV genes have been reported as duplicated or deleted from the germline repertoire of some individuals resulting in varied gene copy numbers (Watson, *et al.*, 2012; Watson, *et al.*, 2013). IGHV genes make up the majority of the heavy chain variable region (VH) of mature antibodies and are central to antigen binding. Differences in germline IGHV

repertoires between different populations were recently highlighted in an extensive study of the human immunoglobulin gene locus (Watson*, et al.*, 2013), where African individuals were found to be particularly diverse.

Given the propensity of some HIV mAbs to use restricted IGHV genes, we examined whether HIV-infected individuals who develop bNAbs have unique IGHV repertoires compared to individuals who do not, in order to determine whether the ability to develop these types of antibodies is genetically restricted.

1.3. METHODS AND MATERIALS

1.3.1. Samples

This study involved 28 adult (> 18 years) women of African ancestry (mostly Zulu speaking) with HIV-1 subtype C infection from the CAPRISA 002 (Van Loggerenberg*, et al.*, 2008) and 004 cohorts (Abdool Karim*, et al.*, 2010) who were being followed at urban and rural clinics in KwaZulu-Natal, South Africa. Of the 28 women, 13 developed bNAbs (referred to as BCN individuals) and 13 did not develop bNAbs (referred to as non-BCN individuals) despite chronic HIV infection and two were intermediate neutralizers (Appendix, Supplementary Table 1.1). BCN individuals were defined as those whose sera from 2 years post-infection were able to neutralize 33-94% (median 56%) of a panel of 18 viruses, made up of 6 subtype A, 6 subtype B and 6 subtype C viruses of which 2 viruses were isolated from the women in the cohort. Non-BCN individuals neutralized 0-11% of the 18 viruses and had the same viral loads as the BCN individuals at 6 months post-infection (to remove viral load biases associated with the development of bNAbs (Gray*, et al.*, 2011b)). This study was given ethics clearance from the Human Research Ethics Committee for Medical Research in Johannesburg, South Africa (clearance number: M111104 for this study and M080470 for the CAPRISA parent study).

1.3.2. DNA Extraction

Genomic DNA was extracted from peripheral blood mononuclear cells (PBMCs) from each individual. Prior to extraction, PBMCs were thawed (at 37°C) and washed with 10ml

RPMI media with 10% FBS. Genomic DNA was extracted from the pellet from all 28 individuals using a Promega Wizard Genomic purification kit.

1.3.3. Primer Design and Amplicon Library Preparation

Alignments were created for each IGHV subgroup using sequences obtained from ENSEMBL (ENSEMBL, 2013) and IMGT (IMGT, 2014). Forward primers (with the exception of IGHV1-F) were mapped to the intron in the leader sequence of each gene alignment. The reverse primers were mapped to the intron after the CDR3 region of each gene alignment. IGHV1, IGHV3 and IGHV4 primers were designed based on previously published primer sets (Wang*, et al.*, 2011). New primers for IGHV2, IGHV5, IGHV6 and IGHV7 were designed based on related sequences for each subgroup. A total of eight primer sets were used to amplify all seven subgroups. The binding properties of the primer sets were determined using UCSC's BLAT (BLAT, 2012) and NCBI's BLAST (BLAST, 2012). Additional next generation sequencing (NGS) specific sequences were added to the gene specific primers to allow sequencing of the Roche 454 and Illumina MiSeq. The 454 primers included the 454-adapter sequence, key sequence and 10bp MID sequences (4 unique sequences in total) (listed in Appendix, Supplementary Table 1.2). The Illumina MiSeq primers included a MiSeq index binding tag and the read 1 or read 2 tags (used for paired-end reads) (listed in Appendix, Supplementary Table 1.3). Nextera XT Indexing tags, which allow pooling of samples on the MiSeq, were added during library construction rather than included into the primer (further described in Section 1.3.4).

Each IGHV subgroup was amplified three times for each individual to ensure adequate coverage of the subgroup and minimise PCR bias. The PCR conditions for all eight amplicons and both 454 and Illumina primers were the same with the exception of the annealing temperatures: IGHV1, IGHV3a and IGHV5 at 56°C, IGHV2, IGHV6 and

IGHV7 at 59°C and IGHV3b and IGHV4 at 55°C. The PCR conditions were as follows: initial denaturation at 94°C for 3 minutes, 35 cycles of denaturation at 94°C for 15 seconds, annealing for 45 seconds, extension for 1 minute at 72°C, final extension at 72°C for 8 minutes and held at 4°C. Each PCR contained 17.5µl dH_2O, 2.5µl Roche FastStart HiFi 10x buffer with 18mM $MgCl_2$, 0.5µl dNTP mix (10mM each), 1µl each primer (10uM), 0.25µl of Roche FastStart HiFi Enzyme (5U/µl) and 1µl of 10ng/µl DNA. The PCR amplicon lengths ranged from ~390-440bp. All replicates of the eight amplicons from each individual were pooled to create a full IGHV repertoire per individual.

1.3.4. Next Generation Sequencing

Fifteen individuals were sequenced on the Roche 454 GS Junior, with four individuals pooled per sequencing run. All clean-up procedures and sequencing on the Roche 454 GS Junior (Titanium) was done as per the manufacturers recommendations.

In order to confirm rare alleles that were only found in single individuals, eleven of the individuals sequenced on the GS Junior were re-sequenced on the Illumina MiSeq along with thirteen additional individuals. A custom amplicon sequencing approach was used for the Illumina MiSeq sequencing. Nextera XT Indexing tags were added to the pooled MiSeq amplicon libraries for each individual using 5µl of each Nextera XT index (two per sample), 1µl of cleaned PCR product (amplicon library prep), 0.5µl EpiCentre FailSafe enzyme, 25µl EpiCentre FailSafe PCR PreMix and 13.5µl dH_2O. Thermocycler conditions: 72°C for 3 minutes, 95°C for 30 seconds, 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds followed by a final extension at 72°C for 5 minutes. All products were checked on an Agilent bioanalyser and cleaned-up using 0.75x Ampure Beads, using the manufacturers protocol. Each sample was quantified on a Qubit and diluted to 8nM. A single 8nM pooled library was then created by pooling 4µl of each

diluted sample, 5µl of which was denatured using 5µl 0.2N NaOH, according to MiSeq protocol. A final concentration of 12pM denatured DNA library with 15% PhiX control was run onto the Illumina MiSeq, using the MiSeq reagent kit (version 2) with 2 x 250 paired-end reads.

1.3.5. Sequence Data Analysis

Various bioinformatics languages were used to analyse the NGS data, including python, R and java as well as publically available bioinformatics tools. FASTQ files were extracted from the raw GS Junior output (.sff files) or automatically generated on the MiSeq. The resulting sequences were quality trimmed using QTrim (Shrestha*, et al.*, 2014). Sequences with a minimum read length of 260bp and a mean quality score of 23 or higher (≥99% confidence) were considered high quality and used for downstream analyses. Paired-end MiSeq reads were merged using PEAR (Zhang*, et al.*, 2014) after quality trimming. All identical sequences were collapsed to single unique reads. Unique reads compiled from 6 or more identical sequences were used for further analysis while all unique reads compiled from less than 6 sequences were discarded from further analysis. The unique sequences were compared to a database of IGHV alleles downloaded from IMGT (IMGT, 2014) using a custom BLAST approach with match reward score of 2, mismatch penalty of 5, a gap penalty for insertions of 16 and a gap extension for deletions with a penalty of 4. All related sequences for each IGHV gene were aligned using RAMICS (Wright*, et al.*, 2014).

Sequences with 100% matches to functional IMGT alleles were assigned as that particular IMGT allele. Sequences matching open reading frames or pseudogenes were removed from further analysis. Sequences with nonsynonymous and/or synonymous single nucleotide polymorphisms (SNPs), insertions or deletions compared to their top matched functional IMGT sequence were given the name of the top scoring IMGT match with the

suffix “m” to denote a single mismatch (nonsynonymous and synonymous SNPs), e.g. IGHV1-2*5m and “mm” for multiple SNPs, insertions or deletions, e.g. IGHV1-2*5mm. Where more than one sequence with a single mismatch or multiple mismatches to the same top scoring IMGT functional sequence was observed, an additional numerical identifier was given to the name, for example: IGHV1-2*5m2 and IGHV1-2*5mm2. These sequences were compared to those listed in the Immunoglobulin Polymorphism Database (IgPdb, 2014), GenBank (GenBank, 2013), NCBI dbSNP (dbSNP, 2013) and NCBI’s IgBLAST (IgBLAST, 2014) and other published IGHV data (Boyd*, et al.*, 2009; Jackson*, et al.*, 2014; Parameswaran*, et al.*, 2014; Wang*, et al.*, 2014a; Wang*, et al.*, 2014c). If 100% matches were found in any of these databases or publications these sequences are described as non-IMGT alleles. If no match was found and sequences were observed in multiple individuals or in both sequencing platforms in a single individual, we report these as novel alleles.

Broadly neutralizing antibody (bNAb) sequences were obtained from CATNAP (CATNAP, 2014) or GenBank (GenBank, 2013). The germline IGHV gene usage for these bNAbs were also obtained from CATNAP (CATNAP, 2014) or the relevant bNAb isolation publication. The bNAb sequences were compared to our IGHV database, to see if any novel alleles might match these sequences better than their currently predicted germline IGHV gene usage. To do this we used blastn from BLAST 2.2.29+ with a match reward of 1, mismatch penalty of -1, gap deletions of 5, gap extension of 2, word size of 7 and an E value threshold of $1e^{-10}$. Monoclonal antibodies (mAbs) have previously been isolated from 7 of the CAPRISA participants examined in this study using single-cell sorting and PCR amplification of the heavy and light chains (data unpublished). The majority of these mAbs were not HIV-specific. The sequences obtained from these mAbs were used to assign germline IGHV gene usage in each participant.

1.3.6. Statistical Analysis

Two-tailed Fisher's exact tests with 95% confidence intervals were used to test the difference in germline IGHV gene repertoires between the BCN and non-BCN sample groups.

1.4. RESULTS

1.4.1. Comparison of NGS technologies

The MiSeq gave 47x more data (7,093,765 quality reads vs. 150,542 quality reads on the 454) in the 11 individuals sequenced on both technologies, even though some of the subgroups were not re-sequenced on the MiSeq. Based on the depth of coverage on the MiSeq we were able to pool more individuals per run (~3.5x as many, ~14 per MiSeq run compared to 4 on the 454). Approximately 70% (192/276) of the unique sequences amplified in this study were observed in both 454 and MiSeq. Of these 276 sequences, 16 were excluded from further analysis as they were only observed in one individual on the 454 only and not found in any other publications or databases. Based on the lower level of depth in the 454 data, the germline IGHV repertoires from the 4 individuals that were only sequenced on the 454 may be underrepresented (highlighted in Supplementary Table 1.1).

1.4.2. Novel Germline IGHV Alleles in South African Women

The germline IGHV gene repertoire of 28 HIV-infected women in the CAPRISA 002 and 004 cohorts based in KwaZulu-Natal, South Africa were sequenced using both the 454 and Illumina NGS platforms. Sequences with perfect matches to IMGT (IMGT, 2014) listed alleles were given the IMGT allele name and are referred to as IMGT alleles. Sequences that did not match IMGT alleles due to synonymous or nonsynonymous mutations or indels but were matches to non-IMGT immunoglobulin gene databases (dbSNP, 2013; GenBank, 2013; IgBLAST, 2014; IgPdb, 2014) or immunoglobulin gene publications (Boyd*, et al.*, 2009; Jackson*, et al.*, 2014; Parameswaran*, et al.*, 2014; Wang*, et al.*, 2014a;

Wang*, et al.*, 2014c) were referred to as non-IMGT alleles. Sequences with no matches to either IMGT or non-IMGT alleles and were observed in multiple individuals or across both technologies in one individual are hereafter referred to as novel alleles.

A total of 47 functional IGHV genes representing all seven subgroups were found in our cohort (Figure 1.1A). Only one published gene, IGHV3-NL1 first identified in Papua New Guinea (Wang*, et al.*, 2011), was not detected. There was a wide range in the number of alleles associated with each gene with 20 alleles in the IGHV2-70 gene compared to only a single allele for the IGHV3-20 gene. Of the 259 alleles identified in this population, just over a half (~52%, n=136) had an exact match to those listed in IMGT, while ~15% (n=38) were non-IMGT alleles and ~33% (n=85) were novel alleles. Eighty-one of the 85 novel alleles (95%) were observed in more than one individual (Figure 1.1B). The 4 novel alleles observed in single individuals were confirmed with both 454 and Illumina sequences. The most commonly observed novel alleles were IGHV6-1*1m1 found in 25 individuals (~89%) and IGHV2-26*1m2 found in 20 individuals (~71%) both of which contained a single SNP compared to the known alleles. Overall, 48% (n=60) of the novel and non-IMGT alleles identified in this study were found in at least 4 individuals indicating that many are fairly common. Furthermore this study significantly expanded the number of alleles for certain genes. For example, IGHV6-1, which previously only had 2 alleles reported in IMGT, was found to have 10 novel alleles and 1 non-IMGT allele (Figure 1.1). Similarly, IGHV2-26 was found to have 13 alleles (including 10 novel alleles and 2 non-IMGT) compared to 1 recorded in IMGT. Other significant expansions were in the IGHV1-69 and IGHV2-70 genes.

The majority (~97%, n=119) of novel and non-IMGT alleles had single or multiple SNPs (synonymous and nonsynonymous), with single nonsynonymous SNPs being the most common (Figure 1.2). One novel allele (IGHV2-26*1mm4, observed in 15

individuals) had 8 nonsynonymous SNPs (Figures. 1.1B and 1.2A). Four novel alleles had indels relative to their most closely matched IMGT allele and three of the four had additional SNPs. IGHV3-11*5mm, observed in two individuals, had a full codon insertion, while frameshifts as a result of single nucleotide deletions were found in IGHV4-59*8mm and IGHV4-61*2mm, each observed in two individuals, and IGHV6-1*1m8 that was found in six individuals (alignments of which can be seen in Appendix, Supplementary Figures 1.1-1.4). It is likely that the three alleles with the frameshift mutations are not functional but rather novel pseudogenes.

When the repertoires for each of the 28 individuals were analysed separately, it was apparent that there was a wide range in the number of alleles per individual (Figure 1.3). CAP295 had the largest number of alleles with 105 while CAP322 had the fewest at 23 alleles. Each individual's repertoire included IMGT, non-IMGT and novel alleles. There was variation in the number of alleles per IGHV family as some individuals lacked alleles for IGHV6 or IGHV7 while others had smaller numbers of IGHV2, IGHV4 or IGHV5 alleles. This data suggests variations in copy number within this population with CAP295 having duplicated genes and CAP322 having gene deletions, which would have to be confirmed through digital drop PCR. A phylogenetic tree of all 259 alleles, including IMGT, non-IMGT and novel alleles, sequenced in this study revealed separate clustering of each of the IGHV subgroups, with all of the novel and non-IMGT alleles clustering with their respective IMGT genes (Figure 1.4).

1.4.3. Novel Germline IGHV Allele Usage in bNAbs

In order to determine whether these novel and non-IMGT germline alleles are being used by the human immune system to generate functional antibodies, we analysed the germline gene usage of monoclonal antibodies (mAbs) isolated by single-cell PCR from seven

CAPRISA participants who were part of this study and 57 well known HIV-1 bNAbs. We identified 15 mAbs that made use of seven novel or non-IMGT alleles identified from NGS sequencing of the 28 individuals (Figure 1.5). Of these, two clonally related mAbs were HIV-1 specific (CAP255-37C and CAP255-18F). An example of one of the new non-IMGT alleles used by two mAbs isolated from two different individuals (CAP177 and CAP248) is shown in Figure 1.5B. Based on the IGHV repertoires from CAP177 and CAP248 the germline IGHV gene predicted to be used by these two mAbs is IGHV2-5*9m which has a Serine (S) at position 62 rather than a Glycine (G) present in the closest related IMGT allele IGHV2-5*09. All other SNPs found in these two mAbs, compared to the germline allele are likely the results of somatic hypermutation (SHM). Of the remaining seven novel and non-IMGT alleles used by these mAbs, 4 (57%) had a better match at the amino acid level and 3 (43%) at the nucleotide level compared to the closest matching IMGT allele (Figure 1.5A). Within the bNAbs we found 14 instances where the novel or non-IMGT alleles, identified in this study, provided the same or a better match than their currently predicted IMGT allele (Figure 1.5C). Two of the predicted alleles (IGHV4-39*7m2 and IGHV4-59*1m2) were also observed among the mAbs isolated from the CAPRISA participants. An example of alternate IGHV allele usage for the bNAbs is shown in Figure 1.5D, where PGT130 and PGT131 are more likely to be using IGHV4-39*7m1 or IGHV4-39*7mm, which like the mAbs have a Valine (V) at the second amino acid, rather than a Leucine (L) as seen in IGHV4-39*07. Alignments of all other mAbs listed in figure 1.5A can be found in Appendix, Supplementary Figures 1.5-1.10. While alignments for all other bNAbs listed in Figure 1.5C can be found in Appendix, Supplementary Figures 1.11-1.15.

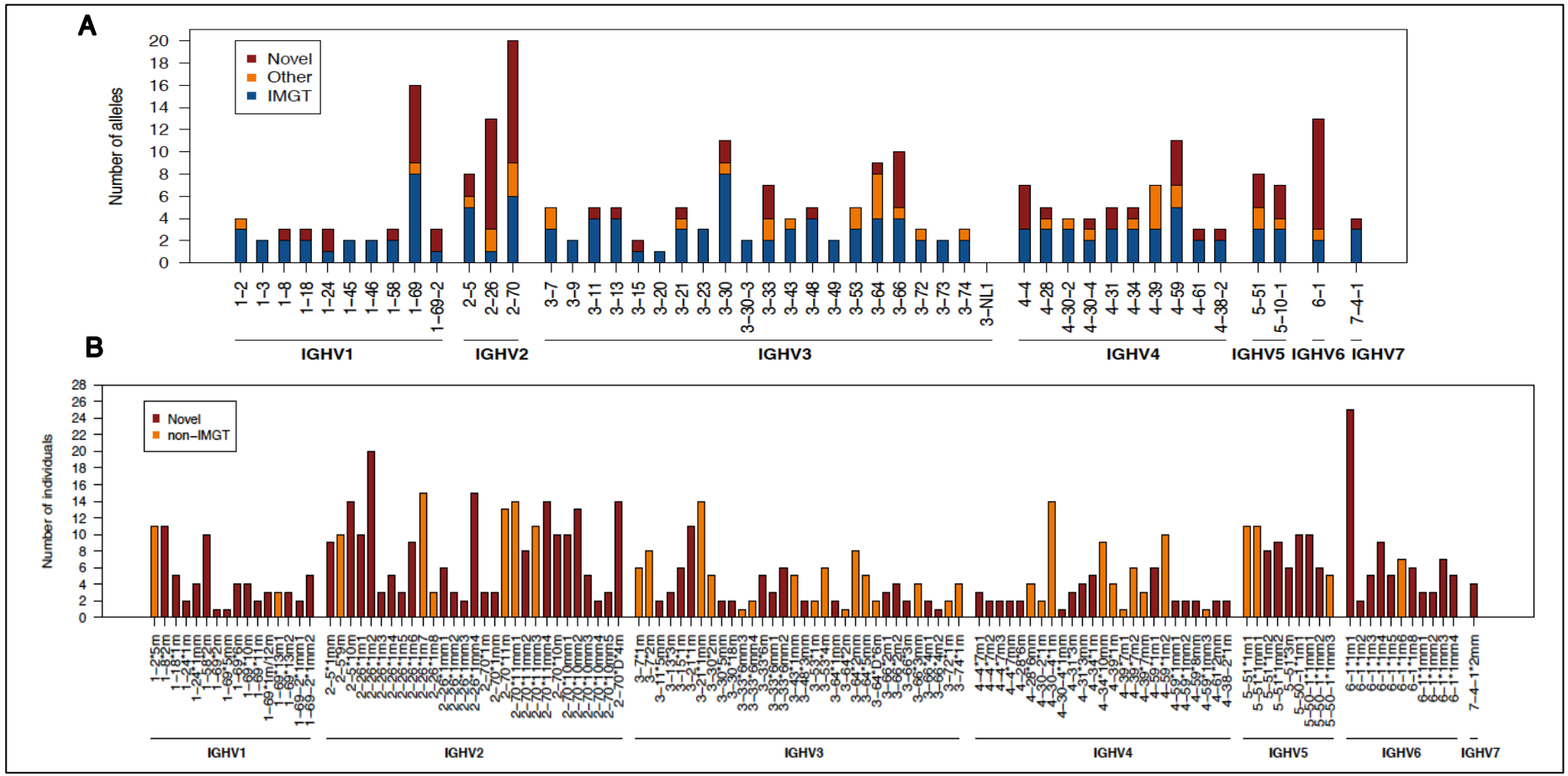


Figure 1.1: Novel germline IGHV alleles in 28 South African women

A) Number of alleles observed for each IGHV gene. Shown are the numbers of novel (red), non-IMGT (orange) and IMGT (blue) alleles. **B)** Prevalence of novel and non-IMGT alleles germline IGHV alleles.

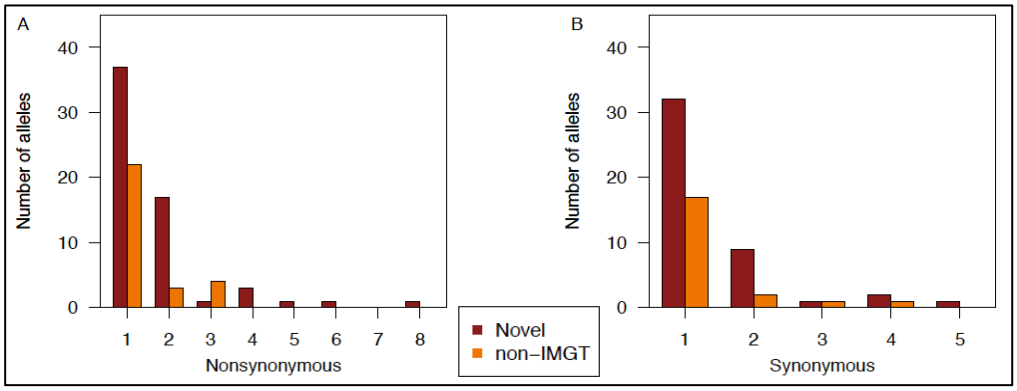


Figure 1.2: Number of single nucleotide changes (SNPs) in the novel and non-IMGT germline IGHV alleles

A) Number of nonsynonymous SNPs observed in the novel and non-IMGT germline IGHV alleles. Shown are the number of novel (red) and non-IMGT (orange) alleles. **B)** Number of synonymous SNPs observed in the novel and non-IMGT germline IGHV alleles.

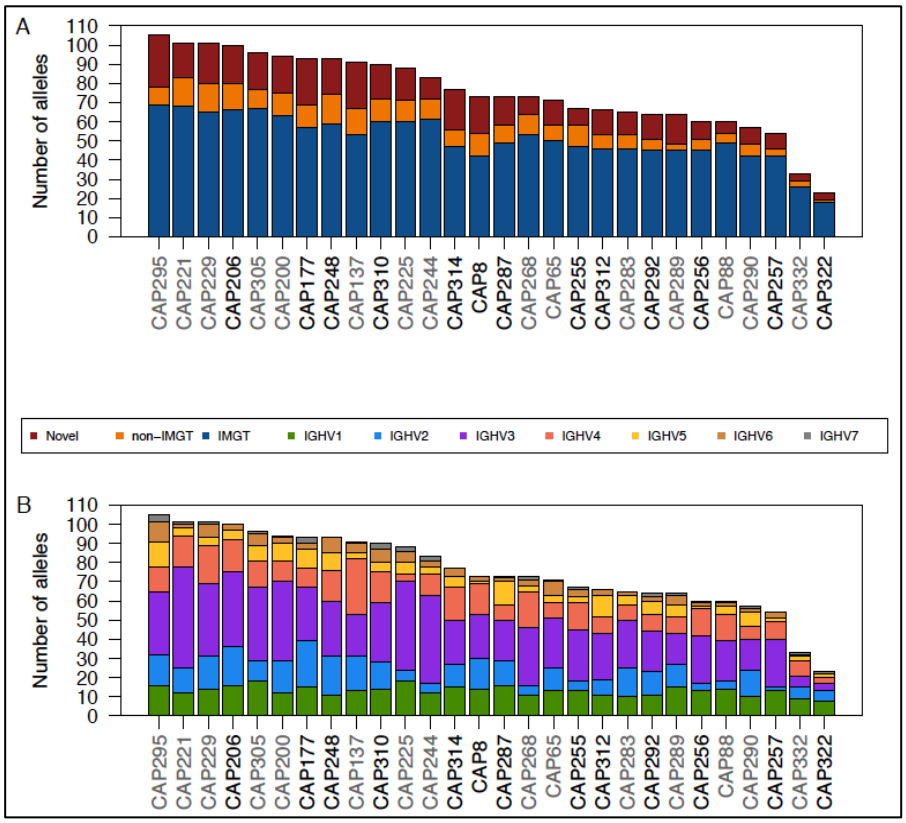


Figure 1.3: Number of germline IGHV alleles for each individual

A) Total number of novel, non-IMGT and IMGT alleles observed. Shown are the numbers of novel (red), non-IMGT (orange) and IMGT (blue) alleles. BCN individuals are highlighted in black and the non-BCN individuals in grey. **B)** Total number of alleles for each of the seven IGHV subgroups. Shown are the total numbers of alleles coloured according to IGHV subgroup.

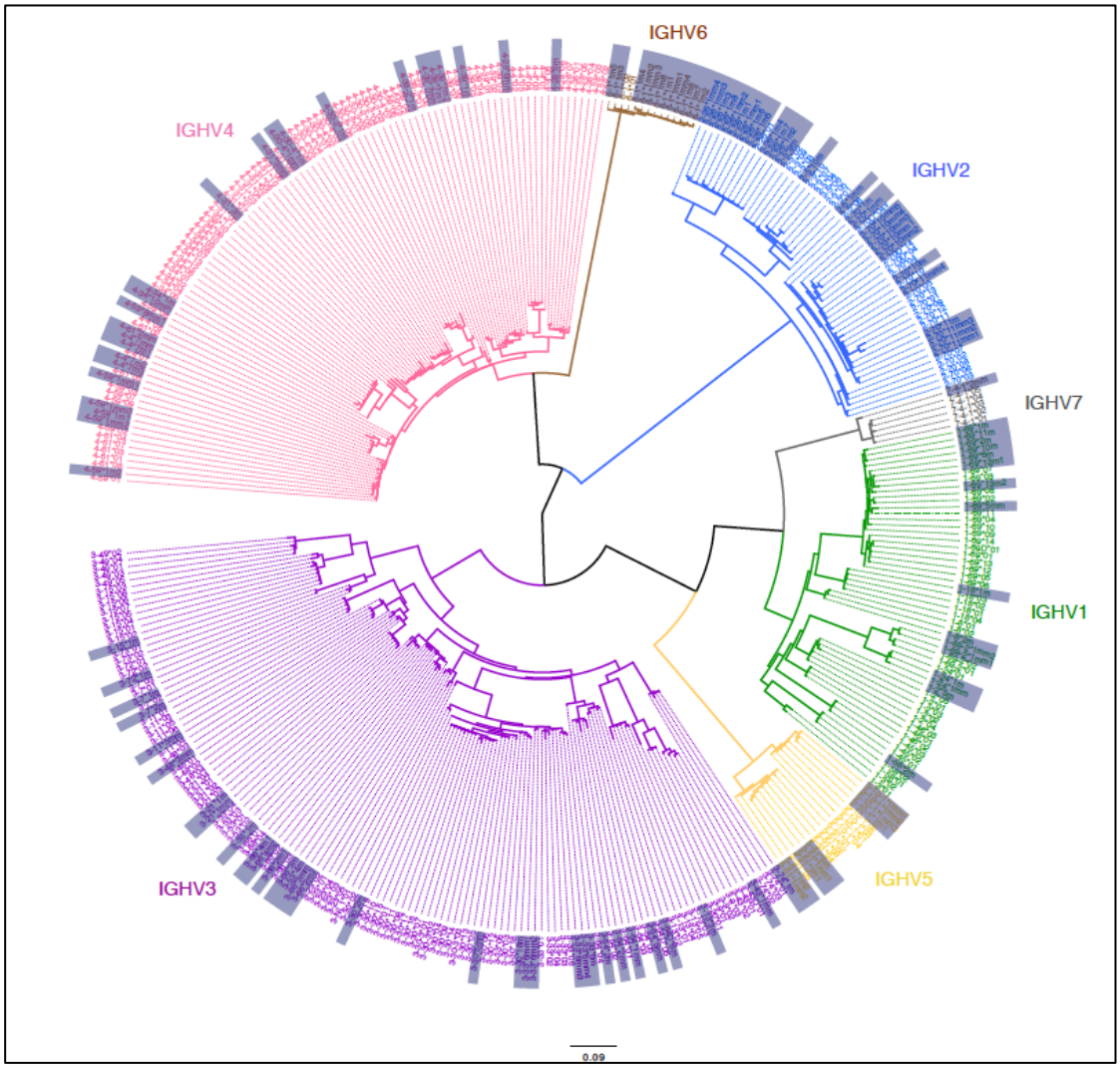
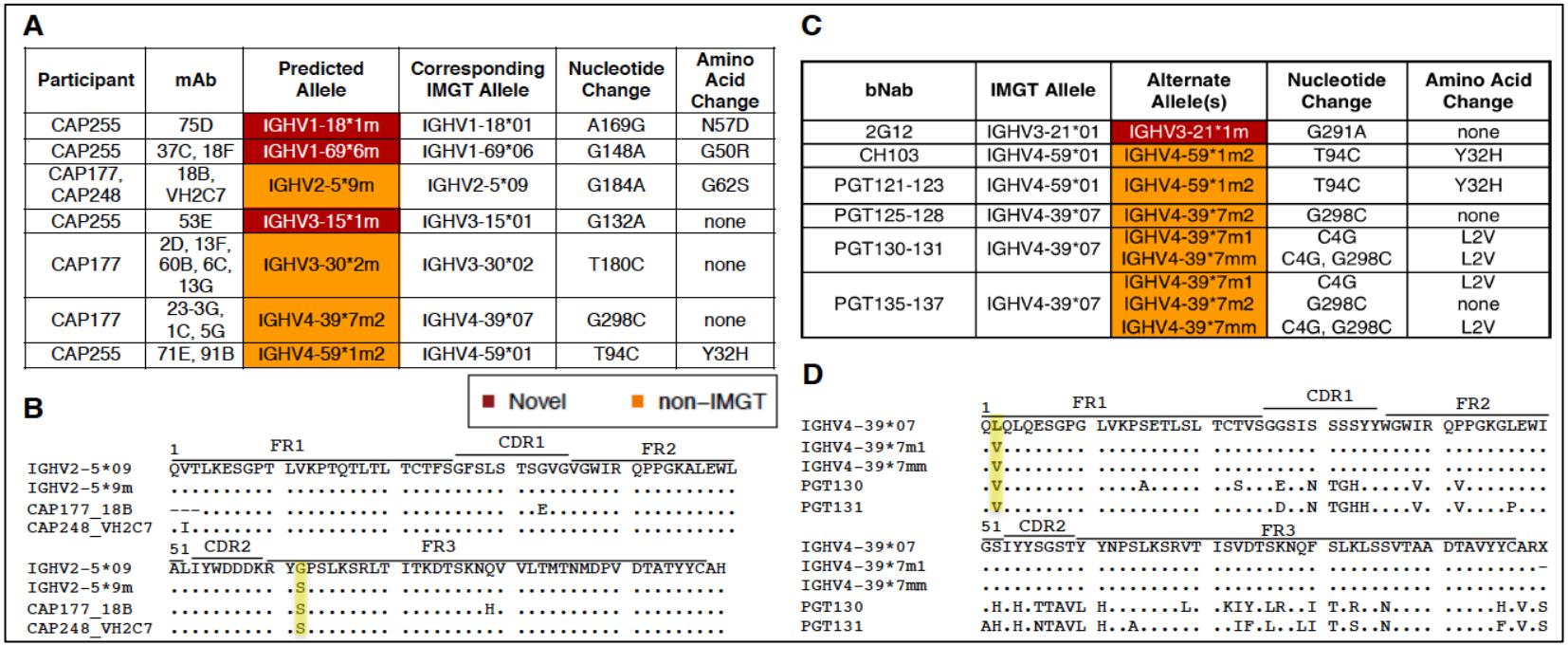


Figure 1.4: Phylogenetic tree of all germline IGHV alleles observed in 28 South African individuals

Shown are all IMGT, non-IMGT and novel alleles observed in this study. Alleles are coloured according to IGHV subgroup and the non-IMGT and novel alleles are highlighted in the grey boxes.



A

Participant	mAb	Predicted Allele	Corresponding IMGT Allele	Nucleotide Change	Amino Acid Change
CAP255	75D	IGHV1-18*1m	IGHV1-18*01	A169G	N57D
CAP255	37C, 18F	IGHV1-69*6m	IGHV1-69*06	G148A	G50R
CAP177, CAP248	18B, VH2C7	IGHV2-5*9m	IGHV2-5*09	G184A	G62S
CAP255	53E	IGHV3-15*1m	IGHV3-15*01	G132A	none
CAP177	2D, 13F, 60B, 6C, 13G	IGHV3-30*2m	IGHV3-30*02	T180C	none
CAP177	23-3G, 1C, 5G	IGHV4-39*7m2	IGHV4-39*07	G298C	none
CAP255	71E, 91B	IGHV4-59*1m2	IGHV4-59*01	T94C	Y32H

B

■ Novel ■ non–IMGT

```
                1          FR1                          CDR1          FR2
IGHV2-5*09      QVTLKESGPT LVKPTQTLTL TCTFSGFSLS TSGVGVGWIR QPPGKALEWL
IGHV2-5*9m      .......... .......... .......... .......... ..........
CAP177_18B      ---....... .......... .......... ..E....... ..........
CAP248_VH2C7    .I........ .......... .......... .......... ..........
                51  CDR2                  FR3
IGHV2-5*09      ALIYWDDDKR YGPSLKSRLT ITKDTSKNQV VLTMTNMDPV DTATYYCAH
IGHV2-5*9m      .......... .S........ .......... .......... .........
CAP177_18B      .......... .S........ ........H. .......... .........
CAP248_VH2C7    .......... .S........ .......... .......... .........
```

C

bNab	IMGT Allele	Alternate Allele(s)	Nucleotide Change	Amino Acid Change
2G12	IGHV3-21*01	IGHV3-21*1m	G291A	none
CH103	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT121-123	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT125-128	IGHV4-39*07	IGHV4-39*7m2	G298C	none
PGT130-131	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7mm	C4G C4G, G298C	L2V L2V
PGT135-137	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7m2 IGHV4-39*7mm	C4G G298C C4G, G298C	L2V none L2V

D

```
                1          FR1                          CDR1          FR2
IGHV4-39*07     QLQLQESGPG LVKPSETLSL TCTVSGGSIS SSSYYWGWIR QPPGKGLEWI
IGHV4-39*7m1    .V........ .......... .......... .......... ..........
IGHV4-39*7mm    .V........ .......... .......... .......... ..........
PGT130          .V........ ....A..... ..S...E..N TGH.....V. .V........
PGT131          .V........ .......... ......D..N TGHH....V. .V....P...
                51  CDR2                  FR3
IGHV4-39*07     GSIYYSGSTY YNPSLKSRVT ISVDTSKNQF SLKLSSVTAA DTAVYYCARX
IGHV4-39*7m1    .......... .......... .......... .......... .........-
IGHV4-39*7mm    .......... .......... .......... .......... ..........
PGT130          .H.H.TTAVL H.......L. .KIY.LR..I T.R..N.... .....H.V.S
PGT131          AH.H.NTAVL H..A...... ..IF.L..LI T.S..N.... .....F.V.S
```

Figure 1.5: Novel IGHV alleles used by isolated mAbs and some bNAbs

(**A**) Germline IGHV allele usage for isolated mAbs from study participants, showing the germline IGHV alleles used, their corresponding IMGT allele, nucleotide and amino acid substitutions between the two alleles. (**B**) Alignment of germline IGHV allele usage for isolated mAbs from two study participants. Highlighted in yellow are the differences in amino acids at position 62, G (Glycine) and S (Serine). Amino acid changes not highlighted in the figure are a likely result of somatic hypermutation during antibody maturation. Missing data in the sequence are represented by "-". (**C**) Germline IGHV allele usage for selected bNAbs. Showing the published IMGT predicted allele and alternate allele(s) predicted from IGHV alleles in 28 South African individuals. (**D**) Alignment of predicted germline IGHV allele usage for PGT130 and PGT131. Highlighted in yellow are the differences at the second amino acid V (Valine) and L (Leucine). Amino acid changes not highlighted in the figure are likely a result of somatic hypermutation during antibody

1.4.4. Comparison of Germline IGHV Repertoires in BCN and non-BCN Individuals

Given that some HIV broadly neutralizing mAbs show biased heavy chain variable gene usage, we analysed the germline IGHV repertoires of BCN and non-BCN individuals to determine whether the ability to develop bNAbs was associated with a particular genotypic profile. Among the 28 CAPRISA women, 13 developed bNAbs in their plasma after 2-4 years of HIV infection (referred to as BCN individuals), 13 women did not develop bNAbs (referred to as non-BCN individuals), despite matching viral loads, and two were intermediate neutralizers (Appendix, Supplementary Table 1.1, (Gray*, et al.*, 2011b) and unpublished data). Comparison of the BCN and non-BCN individuals revealed that both groups had similar profiles of IGHV genes (Figure 1.6), approximately the same number of alleles for each of those genes (Figure 1.7) and the same number of potential alleles (Figure 1.3). The only gene with a difference between the two groups was IGHV3-30 (p-value=0.035, although not significant following Bonferroni multiple testing correction); where non-BCN individuals had five more alleles (n=11) compared to the BCN individuals (n=6) (Figure 1.7).

We also compared the frequency of the specific germline IGHV alleles used by bNAbs between the non-BCN and BCN groups. Of the 57 monoclonal bNAbs targeting one of 4 different epitopes on the HIV envelope glycoprotein, most made use of IGHV1, IGHV3 and IGHV4 germline genes (Figure 1.8) and as reported above some may use novel/non-IMGT alleles (Figure 1.5C). Some alleles used by bNAbs were not observed in this study group. This included IGHV1-46*02, used by some CD4bs antibodies, however there were two other alleles for this gene, including IGHV1-46*01, that were found in all individuals. Similarly, IGHV3-15*05 (used by bNAb 10E8) and IGHV3-33*05 (used by bNAbs PG9 and PG16) were not observed in any of the individuals in this study, but other alleles were more commonly observed including some novel/non-IMGT alleles. The

IGHV1-2*02 allele which is also preferentially used by CD4bs antibodies (shown in blue in Figure 1.8) occurred at high frequencies in both the non-BCN and BCN groups. The most prevalent IGHV alleles in both groups were IGHV4-59*01 (used by the CD4bs bNAb CH103 and the V3/glycan bNAbs PGT121-123), IGHV1-2*02 (used by CD4bs mAbs), IGHV1-3*01 (used by the CD4bs bNAb IgG1b12) and IGHV2-5*01/05 (used by bNAb 2F5 that targets the MPER). Importantly, there was no difference in the frequency of any of these alleles between the 13 BCN and 13 non-BCN individuals. Thus, the inability of all HIV-infected individuals to develop bNAbs is not due to a paucity of the relevant alleles in their IGHV germline gene repertoires.

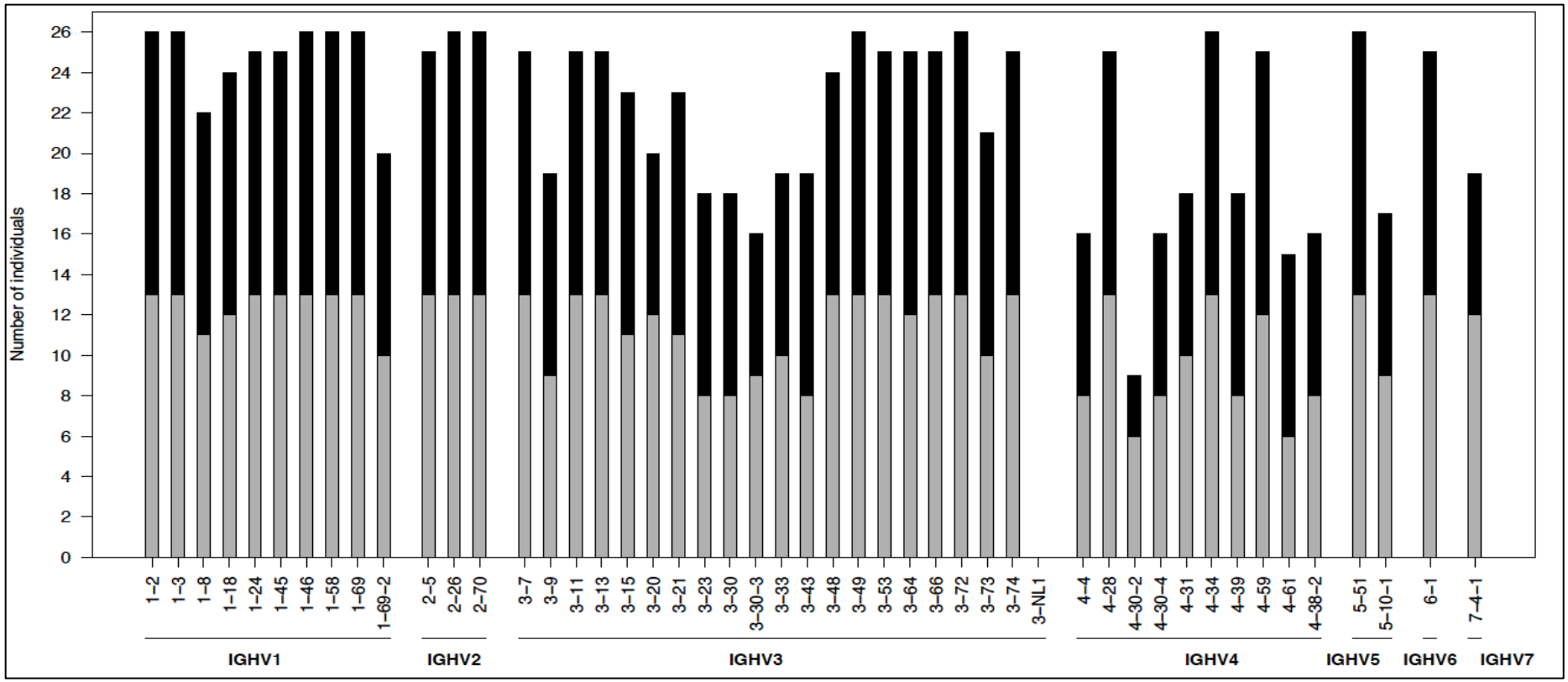


Figure 1.6: Comparison of germline IGHV genes in the repertoires of BCN and non-BCN individuals

Shown are the number of BCN (black) and non-BCN (grey) individuals with a particular functional IGHV gene.

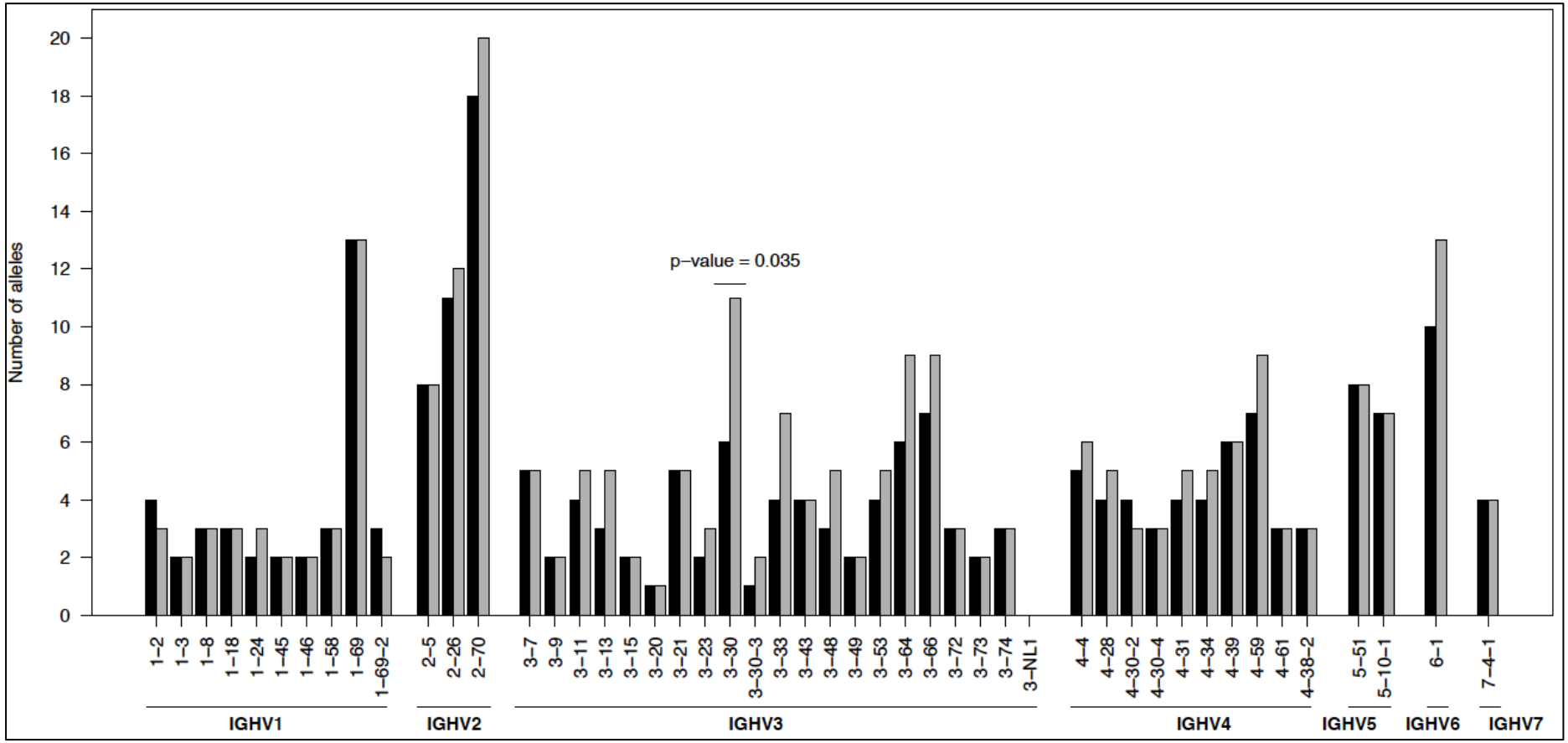


Figure 1.7: Comparison of the germline IGHV alleles in the repertoires of BCN and non-BCN individuals

Shown are the number of BCN (black) and non-BCN (grey) individuals with functional IGHV alleles for each IGHV gene.

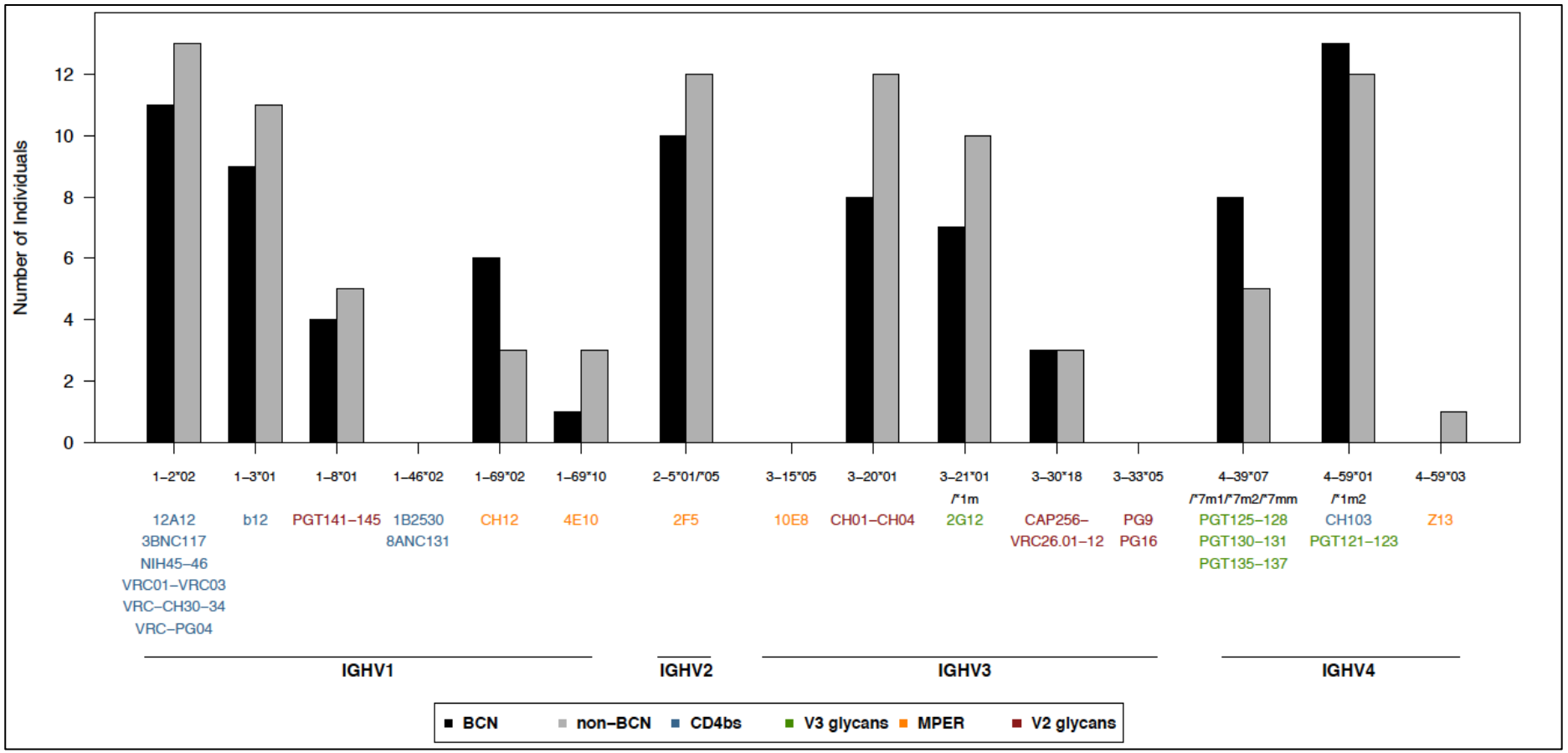


Figure 1.8: Comparison of bNAb germline IGHV alleles between BCN and non-BCN individuals

Shown are the number of BCN (black) and non-BCN (grey) individuals and bNAbs coloured according to epitope.

1.5. DISCUSSION

We analysed the germline immunoglobulin heavy chain variable gene repertoire encoded in the genomic DNA of individuals from KwaZulu-Natal, South Africa and noted that ~48% of the alleles seen in this population are not reported in IMGT. Some of these alleles (non-IMGT alleles) were described in published studies of re-arranged antibodies, although most alleles were novel and are described here for the first time to our knowledge. Further analysis of these IGHV repertoires revealed there to be no differences between those individuals who developed bNAbs to HIV compared to those who did not, despite equivalent antigenic load. Since the induction of these types of antibodies is considered essential for an effective HIV vaccine, these data suggest that the ability to develop bNAbs is not restricted by the IGHV repertoire in an individual.

Previous studies have reported differences in the frequencies of IGHV genes between different populations, with Africans showing particularly unique profiles (Watson*, et al.*, 2013). The presence of IGHV3-64D, IGHV5-10-1 and IGHV7-4-1 has been reported to be lower in African (Luyha, Maasai and Yoruba) populations compared to Asian and European groups (Watson*, et al.*, 2013). We found a similarly low frequency of IGHV3-64D (14%) in this Zulu-speaking South African population. However, IGHV5-10-1 was observed at a higher prevalence (64%) compared to the studied African, Asian and European groups (0.03-16%, 20-21% and 34-48%, respectively) (Watson*, et al.*, 2013), while IGHV7-4-1 frequencies were similar (75%) to those seen in Asian (~78%) and European (54-75%) groups (Watson*, et al.*, 2013). IGHV1-69-2, IGHV3-43D and IGHV4-38-2 genes have been reported to be common in African populations (Watson*, et al.*, 2013) which we corroborated in our study for IGHV1-69-2 and IGHV4-38-2 (75% and 61%,

respectively). However, we observed IGHV3-43D*01 in only 29% (8/28) of individuals studied which is lower than the prevalence reported in other African groups (45-65%, (Watson*, et al.*, 2013)).

We have shown that both novel and non-IMGT alleles are being used by mAbs isolated from individuals in this study as well as some well-characterised anti-HIV bNAbs. This included CH103 and 12 antibodies in the PGT121-137 series all of which were isolated from African donors (Simek*, et al.*, 2009; Walker*, et al.*, 2011; Liao*, et al.*, 2013). Given that three of the novel and non-IMGT alleles (IGHV3-21*1m, IGHV4-39*1m2 and IGHV4-39*7m2) potentially being used by these bNAbs were fairly common in the South African individuals (39%, 36% and 21%, respectively), it is perhaps not surprising, that these have been found to contribute to functional antibodies in other African individuals. IGHV4-39*7m1 and IGHV4-39*7mm were less common in our study group, at 4% and 11% respectively. The other novel and non-IMGT alleles used by mAbs isolated from the study participants were also fairly commonly observed ranging from 14-36%. Only two of the 15 mAbs isolated from the study participants were able to bind HIV, highlighting that the use of novel and non-IMGT alleles by functional antibodies is not HIV-specific and thus could play a role in immune responses to other diseases or infections.

Despite the wide range in the number of IGHV alleles present in each individual there were no differences in the overall germline IGHV repertoires between BCN and non-BCN individuals. This extended to a sub-analysis of the genes and alleles used by known bNAbs whose frequency differed within the cohort but not between BCN and non-BCN groups. The only gene that showed a significant difference between the two groups was IGHV3-30, where non-BCN individuals had more alleles than BCN individuals. This gene is used by the CAP256-VRC26 family of broad and potent V1V2 antibodies, which were isolated from CAPRISA donor CAP256, a participant in this study (Doria-Rose*, et al.*,

2014). We also noted that alleles IGHV1-46*02, IGHV3-15*05 and IGHV3-33*05 used by CD4bs, MPER and V2 glycan bNAbs, respectively, were not found in this South African cohort. Other alleles of these genes were however observed in this cohort, as well as allelic variants used by other bNAbs targeting the same epitopes. Thus, although the specific alleles used by these bNAbs were not present, it did not preclude these individuals from generating antibodies to these epitopes. In addition, CH103 and PGT121-123 which both use IGHV4-59*01 (or the non-IMGT allele IGHV4-59*1m2), target different epitopes, (CD4bs and V3 glycans, respectively), demonstrating that a single germline allele can be used by antibodies against numerous HIV Env targets.

The germline IGHV gene makes the greatest contribution to the unmutated common ancestor (UCA) antibody encoding the entire CDRH1, CDRH2 and FR1, FR2 and FR3 regions. Elegant studies on the UCAs of the VRC01 class of antibodies, which use the VH1-2*02 allele, have shown how two glycans in the V5 region of the HIV envelope obstruct binding of the UCA (Jardine*, et al.*, 2013; McGuire*, et al.*, 2014). Such studies guide the design of suitable envelope immunogens able to trigger the VRC01-class of antibodies (Jardine*, et al.*, 2013). However, other studies aimed at identifying UCA-binding envelopes for PG9 and other bNAb lineages have been less successful, primarily because the UCA approximations used in such studies were predicted from highly mutated antibodies and using incomplete germline databases. The public availability of a comprehensive database of germline immunoglobulin genes, including data from this study, will significantly enhance the accuracy with which UCAs can be inferred.

By studying an under-represented population group in southern Africa, this investigation has greatly expanded the repertoire of germline IGHV genes. We further hypothesise that the IGHD and IGHJ germline genes, which make up the remainder of the

VH region of functional antibodies, are likely to be highly variable within this population as well, and thus warrant investigation.

1.6. CONCLUSIONS

In this study we identified 85 novel germline IGHV alleles within the South African population that have not been previously described. We also identified 38 non-IMGT alleles, which have previously only been described in rearranged antibodies and not in the germline IGHV repertoire. The identification of non-IMGT alleles in the germline IGHV repertoire is an important discovery as it shows that the SNPs observed in these alleles compared to the IMGT alleles are already encoded in the genome rather than a consequence of somatic hypermutation. We have thus expanded the knowledge of the germline IGHV repertoire that is currently described in the global Immunogenetics database, IMGT. In addition we have shown that both novel and non-IMGT alleles are being used by mAbs isolated from study participants and possibly also by well described bNAbs. This knowledge will contribute to a better understanding of antibody response to HIV as well as other infections, immunizations and B cell pathologies. Importantly, we have also shown that the development of bNAbs against HIV is not restricted by the germline IGHV repertoire, which is significant for vaccine development, as it suggests that everyone has the potential to make antibodies capable of neutralizing all strains and subtypes of HIV.

1.7. REFERENCES

Abdool Karim, Q*, et al.* (2010) Effectiveness and safety of Tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science,* 329 (5996): 1168-1174

Boyd, SD*, et al.* (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel V-D-J pyrosequencing. *Science Translational Medicine,* 1 (12): 12ra23

Breden, F*, et al.* (2011) Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease. *PLoS One,* 6 (3): e16857

Corti, D*, et al.* (2010) Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *Journal of Clinical Investigation,* 120 (5): 1663-1673

Doria-Rose, NA*, et al.* (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature,* 509 (7498): 55-62

Gorny, MK*, et al.* (2009) Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Molecular Immunology,* 46 (5): 917-926

Gorny, MK*, et al.* (2011) Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PLoS One,* 6 (12): e27780

Gorny, MK*, et al.* (2012) Functional and immunochemical cross-reactivity of V2-specific monoclonal antibodies from HIV-1-infected individuals. *Virology,* 427 (2): 198-207

Gray, ES*, et al.* (2011a) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology,* 85 (10): 4828-4840

Hraber, P*, et al.* (2014) Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS,* 28 (2): 163-169

Jackson, K*, et al.* (2014) Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host and Microbe,* 16 (1): 105-114

Jardine, J*, et al.* (2013) Rational HIV immunogen design to target specific germline B cell receptors. *Science,* 340 (6133): 711-716

Kwong, PD and Mascola, JR (2012) Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity,* 37 (3): 412-425

Liao, H-X*, et al.* (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature,* 496 (7446): 469-476

Lingwood, D*, et al.* (2012) Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature,* 489 (7417): 566-570

McGuire, A*, et al.* (2014) Diverse recombinant HIV-1 envs fail to activate B cells expressing the germline B cell receptors of the broadly neutralizing anti-HIV-1 antibodies PG9 and 447-52D. *Journal of Virology,* 88 (5): 2645-2657

Parameswaran, P*, et al.* (2014) Convergent antibody signatures in human dengue. *Cell Host and Microbe,* 13 (6): 691-700

Scharf, L*, et al.* (2014) Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports,* 7 (3): 785-795

Scheid, JF*, et al.* (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science,* 333 (6049): 1633-1637

Shrestha, R*, et al.* (2014) QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BioMedCentral Bioinformatics,* 15 (1): 33

Simek, MD*, et al.* (2009) Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *Journal of Virology,* 83 (14): 7337-7348

Stamatatos, L*, et al.* (2009) Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature Medicine,* 15 (8): 866-870

Van Loggerenberg, F*, et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One,* 3 (4): e1954

Walker, LM*, et al.* (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477 (7365): 466-470

Wang, C*, et al.* (2014a) Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *Journal of Immunology* 192 (2): 603-611

Wang, Y*, et al.* (2011) Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics,* 63 (5): 259-265

Wang, Y*, et al.* (2014c) IgE-associated IGHV genes from venom and peanut allergic individuals lack mutational evidence of antigen selection. *PLoS One,* 9 (2): e89730

Watson, CT and Breden, F (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes and Immunity,* 13 (5): 363-373

Watson, CT*, et al.* (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *American Journal of Human Genetics,* 92 (4): 530-546

West, AP*, et al.* (2012) Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proceedings of the National Academy of Sciences,* 109 (30): E2083–E2090

Wright, I and Travers, S (2014) RAMICS: trainable, high-speed and biologically relevant alignment of high-throughput sequencing reads to coding DNA. *Nucleic Acids Research,* 42 (13): e106

Zhang, J*, et al.* (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics,* 30 (5): 614-620

Books

Lefranc, M-P (2001) The immunoglobulin factsbook. Academic Press. ISBN: 012441351

Online Databases

NCBI's BLAST Alignment Tool [Online] Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi [Accessed: January 2012]

UCSC Genome Bioinformatics, BLAT [Online] Available: https://genome.ucsc.edu/cgi-bin/hgBlat?command=start [Accessed: January 2012]

CATNAP, HIV sequence database [Online] Available: http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp [Accessed: June 2014]

NCBI's dbSNP short genetic variations tool. [Online] Available: http://www.ncbi.nlm.nih.gov/SNP/ [Accessed: November 2013]

ENSEMBL [Online] Available: www.ensembl.org [Accessed: November 2013]

NCBI's Genetic Sequence Database, GenBank [Online] Available: http://www.ncbi.nlm.nih.gov/genbank/ [Accessed: November 2013]

NCBI's IgBLAST tool [Online] Available: http://www.ncbi.nlm.nih.gov/igblast/ [Accessed: June 2014]

IgPdb: Immunoglobulin Polymorphism Database [Online] Available: http://cgi.cse.unsw.edu.au/~ihmmune/IgPdb/display_ref.php?ref=2 [Accessed: June 2014]

IMGT - The international immunogenetics information system [Online] Available: http://www.imgt.org/ [Accessed: June 2014]

2. LIMITED EVOLUTION OF A POTENT STRAIN-SPECIFIC ANTIBODY LINEAGE IN AN HIV-INFECTED INDIVIDUAL WHO FAILS TO DEVELOP NEUTRALIZATION BREADTH[b]

Cathrine Scheepers[1, 2], Zizhang Sheng[3], Chaim Schramm[3], Arshad Ismail[1], Bronwen E. Lambson[1, 2], Nigel Garrett[4, 5], Salim S. Abdool Karim[4, 6], Penny L. Moore[1, 2, 4], Lawrence Shapiro[3] and Lynn Morris[1, 2, 4]

[1]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, 2131 South Africa; [2]School of Pathology, Division of Virology and Communicable Disease Surveillance, University of the Witwatersrand, Johannesburg, 2050, South Africa; [3]Columbia University, Department of Biochemistry and Molecular Biophysics, New York City, New York, 10032, USA; [4]Centre for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu-Natal, 4013 South Africa; [5]Department of Infectious Diseases, Nelson R. Mandela School of Medicine, University of KwaZulu–Natal, 4041, Durban, South Africa; and [6]Department of Epidemiology, Columbia University, New York City, 10032, USA

Manuscript in preparation.

---

[b] This project was supported by the Poliomyelitis Research Foundation (PRF), University of the Witwatersrand Health Sciences Faculty Research Council and the National Research Foundation (NRF). Cathrine Scheepers was supported by the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) through the Fogarty International Center, National Institutes of Health (grant # 5 D43 TW000231). CAPRISA is funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes for Health (NIH), and U.S. Department of Health and Human Services (grant: AI51794). Penny L. Moore is a Wellcome Trust Intermediate Fellow in Public Health and Tropical Medicine (Grant 089933/Z/09/Z).

2.1. ABSTRACT

Strain-specific HIV-1 antibodies play an important role in shaping broadly neutralizing antibody responses during infection. We previously identified an HIV-1 subtype C infected individual (CAP88) with a potent strain-specific antibody response targeting the C3-V4 region of gp120 that did not develop neutralization breadth. Plasma neutralization against the C3-V4 was first detected from 15 weeks post-infection, peaked at 26 weeks and then waned following viral escape. A strain-specific monoclonal antibody (CAP88-CH06) was isolated as an IgA1 isotype from donor CAP88 at 34 weeks post-infection. Here we studied the evolution of the CAP88-CH06 heavy chain by next generation sequencing from 5 weeks through to 121 weeks of infection. IgA sequence transcripts that were identical matches to the mature CAP88-CH06 antibody, with 8.8% divergence, were detected at all 7 time-points. This included the earliest time-point collected 10 weeks prior to the detection of plasma neutralizing activity and the latest time-point, collected long after the virus had escaped this response. The presumed unmutated common ancestor of this lineage was also detected at 5 weeks but disappeared after 11 weeks with very few intermediate sequences observed. We also identified clonally related IgG transcripts at 11 to 38 weeks but not at other time-points. This is the first study to examine the development of an early strain-specific HIV antibody lineage over years of infection. The limited evolution of the IgA isotype and the disappearance of the IgG isotype may have contributed to the inability of this lineage to develop neutralization breadth.

2.2. INTRODUCTION

HIV-1 specific binding and neutralizing antibodies of varying breadth and potency targeting the envelope glycoprotein arise shortly after infection. Initial antibody responses can be detected as early as ~12 days post-infection that bind gp41 but are not neutralizing (Tomaras*, et al.*, 2008). These responses are followed by strain-specific antibodies detected from 4-14 weeks after infection, typically gp120-directed and persisting throughout infection (Gray*, et al.*, 2007; Euler*, et al.*, 2012; West*, et al.*, 2014). Strain-specific antibodies exert significant pressure on the virus forcing neutralization escape, often resulting in the development of other strain-specific antibodies targeting different regions (Moore*, et al.*, 2009b; Wibmer*, et al.*, 2013). In approximately 15-30% of HIV-1 infected individuals this co-evolution of virus and antibody leads to the development of broadly neutralizing antibodies (bNAbs) (Stamatatos*, et al.*, 2009; West*, et al.*, 2014; Moore*, et al.*, 2015). BNAbs are able to neutralize numerous strains of the virus but take approximately 2-3 years post-infection to appear (Gray*, et al.*, 2011b).

The targets of strain-specific neutralizing antibodies are usually localised to the variable regions of the viral envelope, allowing the virus to escape neutralization fairly easily (Moore*, et al.*, 2009a). Most commonly this includes the V1V2 and C3-V4 regions of gp120, but V4 and V5 regions have also been implicated (Gray*, et al.*, 2007; Moore*, et al.*, 2008; Moore*, et al.*, 2009b). In contrast the targets of bNAbs cluster within five more conserved regions of the HIV envelope; the V2 site, N332 supersite, CD4-binding site, gp120-gp41 interface and membrane external proximal region (MPER) (Wibmer*, et al.*, 2015). BNAbs tend to have unusual features such as long CDRH3 lengths and high levels of somatic hypermutation that are often required to neutralize at these sites. Thus, CD4bs

antibodies typified by the VRC01 bNAb lineage have levels of somatic hypermutation in excess of 30% (Wu*, et al.*, 2015) while V1V2 bNAbs such as the CAP256-VRC26 lineage have extremely long CDRH3s (35 amino acids) with lower levels of SHM (8-15%) (Doria-Rose*, et al.*, 2014). In contrast, strain-specific antibodies tend to have average levels of somatic hypermutation (1-12%) and normal CDRH3 lengths (~16 amino acids) (West*, et al.*, 2014; Zolla-Pazner, 2014).

The current paradigm in HIV vaccine development focuses on B-cell lineage immunogen design. This involves studying B-cell ontogeny through deep sequencing of highly affinity matured antibody heavy and light chains to identify the early B-cell unmutated common ancestor (UCA) that gave rise to the lineage (Haynes*, et al.*, 2012b). To date two studies have analysed the evolution of bNAbs from acute through to chronic infection. While the UCA of the CH103 CD4bs lineage bound the early autologous virus, neutralization occurred only after a sufficient level of SHM had been achieved (Liao*, et al.*, 2013). In contrast the UCA of the CAP256.VRC26 V1V2 lineage both bound and neutralized the early autologous virus as a result of selection of a B-cell with a pre-existing long CDRH3 (Doria-Rose*, et al.*, 2014). These data highlight the different pathways to acquire neutralizing activity although in both cases, SHM of the earlier strain-specific antibodies was required to achieve broad neutralization.

Strain-specific antibodies have also been shown to play an active role in shaping bNAb responses. In two subtype C infected individuals, neutralization escape from early strain-specific antibodies resulted in a shift of a glycan at position 334 to 332 in the C3 region of gp120, which led to the development of bNAbs targeting the glycan at 332 (Moore*, et al.*, 2012). Similarly neutralization escape, through an Asparagine (N) to Aspartic Acid (D) change at position 167, driven by strain-specific antibodies targeting the V1V2 region of gp120 resulted in the development of bNAbs targeting the same region in

another donor (Wibmer*, et al.*, 2013). In the CH103 lineage, breadth was achieved with the aid of a separate helper antibody lineage (CH235) that was initially strain-specific (Gao*, et al.*, 2014).

We have previously described a strain-specific neutralizing antibody response in a CAPRISA 002 participant (CAP88), which was mapped to the C3-V4 region of gp120 (Moore*, et al.*, 2008; Moore*, et al.*, 2009a). Plasma neutralization of the C3-V4 region was detected from 15 weeks post-infection, peaked at 26 weeks post-infection and then waned, concomitant with viral escape, but persisted throughout infection (Moore*, et al.*, 2009a; Gray*, et al.*, 2011a). The strain-specific monoclonal antibody (mAb) CAP88-CH06 was later isolated at 34 weeks post-infection (Gray*, et al.*, 2011a). The heavy chain of this IgA1 antibody was mapped to the germline IGHV4-39*01, D3-3*01 and J4*02 gene families with 8.8% SHM and a CDRH3 length of 18 aa. The light lambda chain was mapped to IGLV3-21*03 and IGLJ2*01 germline genes (Gray*, et al.*, 2011a). Although potent strain-specific responses developed in this participant, she did not develop breadth even after five years of infection. Here we report the evolution of the heavy chain of this CAP88-CH06 lineage from 5-121 weeks post-infection in order to provide insight into why some strain-specific antibodies do not develop neutralization breadth.

2.3. METHODS AND MATERIALS

2.3.1. Samples

Stored peripheral blood mononuclear cells (PBMCs) collected at seven time-points (5, 8, 11, 17, 34, 38 and 121 weeks post-infection) from a participant (CAP88) in the Centre for the AIDS Programme of Research In South Africa (CAPRISA) 002 Acute Infection cohort were used for this study (Van Loggerenberg*, et al.*, 2008). Ethics clearance was obtained from the Human Research Ethics Committee for Medical Research in Johannesburg, South Africa (clearance number: M111104 for this study and M080470 for the CAPRISA parent study).

2.3.2. Total Nucleic Acid Extraction and cDNA Synthesis

PBMCs were thawed at 37°C and decanted into a preheated 10ml RPMI buffer with 10% FBS. The samples were mixed by pipetting gently and centrifuged at 1200rpm for 10 minutes. The supernatant was discarded and the cells re-suspended in 1ml RPMI buffer with 10% FBS and further centrifuged at 8000rpm for 3 minutes. The supernatant was once again removed leaving ~10-50µl with the pellet. A Qiagen AllPrep DNA/RNA mini kit was used to extract genomic DNA and total RNA from all seven samples as per the manufacturers specifications. Concentrations of the resulting nucleic acids were determined on a NanoDrop spectrophotometer. The RNA was converted to cDNA in two replicates as follows: 5µl Random Hexamers (150ng/µl) (obtained from Integrated DNA Technologies (IDT)) was added to 50µl total RNA, 5µl dNTPs (10mM each, Roche) and 5µl dH_2O which was heated to 65°C for 5 minutes and placed on ice for 1 minute. The

samples were centrifuged and the following added: 20µl 5x First strand buffer (Invitrogen, Life Technologies) 5µl 0.1M DTT (Invitrogen, Life Technologies), 5µl RNAseOUT (Invitrogen, Life Technologies) and 5µl Superscript III RT Enzyme (Invitrogen, Life Technologies) and placed in a thermocycler with the following conditions: 25°C for 5 minutes, 50°C for 60 minutes, 70°C for 15 minutes and held indefinitely at 4°C. The two replicates of each sample were then pooled resulting in a final volume of 200µl cDNA. The gDNA extracted from these time points were stored for later use, while cDNA was used to sequence the CAP88-CH06 antibody lineage during infection.

2.3.3. Primer Design

An IGHV4 forward primer (CAGSTGCAGCTGCAGGAGTCGG) located at the beginning of framework 1 (FR1) and two reverse constant region primers matching IgA (CCAGCCCCAAGGTCTTCCCG) and IgG (GTCTTCCCCCTGGCRCCCTC) were used to amplify clonally related antibodies to CAP88-CH06 from seven time points. The forward primer was designed to amplify all IGHV4 genes including IGHV4-39 (the germline gene used by CAP88-CH06). The reverse IgA and IgG primers were taken from DeKoksy *et al*., (DeKosky*, et al.*, 2013). MiSeq specific index binding site sequences (TCGTCGGCAGCGTC for forward and GTCTCGTGGGCTCGG for the reverse primers) and paired-end read sequences (AGATGTGTATAAGAGACAG) were incorporated into both the forward and reverse primers to allow sequencing on the Illumina MiSeq.

2.3.4. Library Construction and Illumina Sequencing

Each sample was amplified ten times to ensure adequate coverage of the IGHV4 antibody repertoire and minimise PCR bias. A plasmid containing the CAP88-CH06 sequence was

also amplified, once using an IgA reverse primer and once using an IgG reverse primer. Both reactions used the IGHV4 forward primer and the same conditions as those used for the cDNA samples. The PCR conditions were as follows: initial denaturation at 95°C for 5 minutes, 25 cycles of denaturation at 95°C for 30 seconds, annealing for 1 minute at 58°C, extension for 1 minute 30 seconds at 72°C, final extension at 72°C for 7 minutes and held at 4°C. Each PCR contained 29.2µl dH_2O, 5µl Roche FastStart HiFi 10x buffer with 18mM $MgCl_2$, 0.4µl dNTP mix (10mM each), 1µl each primer (10µM), 0.4µl of Roche FastStart HiFi Enzyme (5U/µl) and 2µl of 10ng/µl gDNA/cDNA. The PCR amplicon sizes were 550-570bp. The amplicons were cleaned using 0.75x Ampure Beads as per the manufacturers recommendations. A limited cycle PCR was used to incorporate MiSeq Nextera XT indices into the amplicons, to allow pooling of multiple samples, using the Nextera XT indexing kit and the EpiCentre FailSafe PCR PreMix Selection Kit. The limited cycle PCR conditions were as follows: 5µl of each index (two per sample) was added to 1µl of DNA, 0.5µl FailSafe enzyme and 13.5µl dH_2O, which was then placed into a thermocycler for 72°C for 3 minutes, 95°C for 30 seconds, 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds followed by a final extension at 72°C for 5 minutes. All products were checked on an Agilent bioanalyser and cleaned-up using 0.75x Ampure Beads, using the manufacturers protocol. Each sample (including the two CAP88-CH06 plasmid control samples) was then quantified on a Qubit and diluted to 8nM. A single 8nM pooled library was then created by pooling 4µl of each diluted sample, 5µl of which was denatured using 5µl 0.2N NaOH, according to MiSeq protocol. A final concentration of 12pM denatured library with 15% PhiX control was run on the Illumina MiSeq using the reagent version 3 kit with 2 x 300bp paired-reads.

In order to confirm the presence of the mature CAP88-CH06 antibody in the 5 and 8 weeks post-infection time points, cDNA from these original samples were re-amplified

and re-sequenced on the MiSeq at separate times, following the same protocol as described. A total of three MiSeq runs were used for this study, one including all seven time-points and two plasmid controls, the second using the 5 w.p.i sample only and the third using the 8 w.p.i sample only.

2.3.5. Antibody Sequence Data Analysis and Antibodyomics Pipeline

Paired-end FASTQ files for each time point were obtained from the MiSeq. An average of 130 nucleotides were trimmed from the 3' ends of each reverse mate read to remove poor quality bases. The paired-end reads were merged into full-length reads for each time point using USEARCH (Edgar, 2010). The number of errors present in each read was estimated using USEARCH and all reads with an estimated value above 1 were removed (Edgar, 2010). The merged FASTQ files were converted to FASTA files using fastq_to_fasta part of the FASTX tool kit (http://hannonlab.cshl.edu/fastx_toolkit/).

To measure the rate of sequencing error, we sequenced the CAP88-CH06 plasmid DNA with both IgA and IgG primers, as controls for the IgA and IgG sequences (see Section 2.4.1). For each control sample, the merged reads were aligned to CAP88-CH06 using ClustalO (Sievers*, et al.*, 2011) and the numbers of single nucleotide mismatch errors were counted for each read. This distribution was used to determine the threshold for a two-step clustering process for removing sequences potentially containing errors due to PCR or sequencing. In the first step all sequences in the dataset were clustered using 100% identity. All clusters containing only one read were defined as low coverage and likely contain sequencing errors and thus removed. The remaining non-redundant sequences were then clustered at 99% sequence identity with USEARCH (Edgar, 2010) and sequences represented by larger numbers of raw reads were given priority to be used as centroids for each cluster. The final coverage of a cluster was then calculated by summing the number of

raw reads of each member of a cluster, and clusters with coverage of less than 3 were removed. After the two steps of clustering, we successfully removed 99.9% of sequences containing sequencing errors in the control samples. The following parameters were used for the two-step clustering process: 1) usearch8.0.1517_i86osx32 -derep_fulllength input_file_name.fa -threads 24 -fastaout output_file_name.fa -sizeout -uc output_file_name.txt, 2) usearch8.0.1517_i86osx32-sortbysize input_file_name.fa -minsize 2 -fastaout output_file_name.fa, 3) usearch8.0.1517_i86osx32-cluster_smallmem input_filename.fa -sortedby size -id 0.99 -sizein -sizeout -uc output_file_name.fa -centroids output_file_name.txt, 4) usearch8.0.1517_i86osx32-sortbysize input_file_name.fa -minsize 3 -fastaout output_file_name.fa

An antibodyomics bioinformatics pipeline (in preparation for publication, Zhang *et al.*) was used to study the evolution of CAP88-CH06, by identifying lineage related antibodies from the MiSeq sequencing data of all time points. Briefly, merged reads with a read length less than 300 nucleotides or more than 600 nucleotides were removed and germline V and J genes were assigned to the remaining reads using NCBI Blast (Altschul*, et al.*, 1997). Based on the boundaries of assigned V and J genes, the VDJ region was identified for each read and 5' upstream and 3' downstream sequences were removed. The 3' downstream sequence was further aligned to IgA and IgG constant regions separately using ClustalO (Sievers*, et al.*, 2011) and assigned to either the IgA or IgG subtype based on sequence identity. The boundaries of CDRH3 were then identified using the conserved CXR motif at the 3' end of V gene and WGXG motif in the J gene.

To find lineage related transcripts from each time point, all raw reads having the same germline V (IGHV4-39) and J (IGHJ4) gene assignment as CAP88-CH06 antibody were collected. The CDRH3 of each read was then aligned to that of CAP88-CH06 using ClustalO (Sievers*, et al.*, 2011) and the sequence identity calculated. Sequences from the

raw read dataset with CDRH3 identity ≥ 80% to that of CAP88-CH06 were selected as lineage related raw reads. To remove reads containing PCR errors and/or sequencing errors, we performed the two steps of clustering as described previously. After the second step of clustering, a representative sequence was chosen (by USEARCH, (Edgar, 2010)) for each cluster and formed the unique sequence dataset. We further removed unique sequences containing errors manually and the curated sequences were used for downstream analyses.

To trace the developmental pathway of CAP88-CH06 antibody, a Maximum Likelihood phylogenetic tree was constructed with the lineage related unique reads from all time points using a GTR phylogenetic maximum likelihood model in seaview (Galtier*, et al.*, 1996) and displayed in Dendroscope (Huson*, et al.*, 2012).

2.4. RESULTS

2.4.1. CAP88-CH06 Neutralization and Sequencing

CAP88 is a subtype C infected participant from the CAPRISA 002 cohort who developed a potent autologous strain-specific HIV-1 antibody response (Van Loggerenberg*, et al.*, 2008; Gray*, et al.*, 2011a). The antibody specificity was mapped to the C3-V4 region of the envelope and was first detected from 15 weeks post-infection (w.p.i) with a peak in neutralization at 26 w.p.i. This response was replaced by a second response targeting the V1V2 from 26 w.p.i., however, plasma neutralization against the C3-V4 region was still detectable even after 2 years post-infection (~106 w.p.i) (Figure 2.1, (Moore*, et al.*, 2009a)). Using gp120 containing the C3-V4 region of CAP88, seven antibodies were isolated from this participant at 34 w.p.i. by antigen-specific B-cell sorting (Gray*, et al.*, 2011a). One of these (CAP88-CH06) neutralized the transmitted/founder virus and acute infection viral clones isolated from CAP88 with an IC_{50} of 13ng/ml but failed to neutralize later clones containing escape mutations (Gray*, et al.*, 2011a). CAP88-CH06 was isolated as an IgA1 strain-specific antibody with 8.8% divergence from the heavy chain germline IGHV4-39*01 gene.

In this study we examined the evolution of CAP88-CH06 heavy chain lineage throughout HIV-1 infection by next generation sequencing (NGS). Total nucleic acids were extracted from CAP88 peripheral blood mononuclear cells (PBMCs) taken at seven time points (5, 8, 11, 17, 34, 38 and 121 w.p.i). These time points were chosen to represent the various stages of evolution of this antibody response from first appearance in the plasma, peak breadth, and long after viral escape (Figure 2.1). The cDNA from each time

point was PCR amplified with an IGHV4 forward and two reverse primers for IgA and IgG and sequenced on the Illumina MiSeq. A plasmid sequence of CAP88-CH06 with a hybrid constant region derived from the original IgA antibody and the IgG1 vector was used as a control to determine PCR and sequencing error.

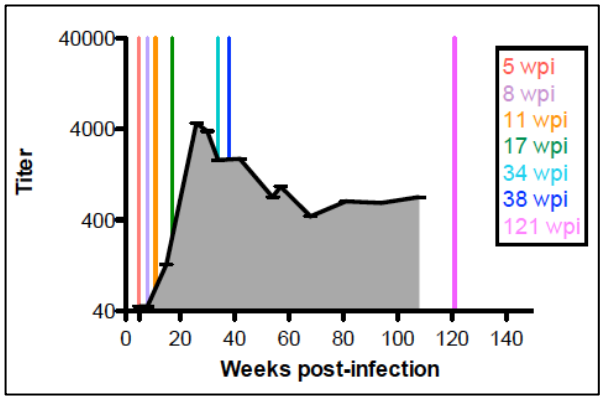


Figure 2.1: Plasma neutralization of the C3-V4 region of CAP88 Env during HIV-1 infection

Shown are the plasma neutralization titres against the C3-V4 region of CAP88 Env in grey and the seven time points (represented by the coloured bars) at which RNA was isolated to study the evolution of this antibody lineage.

A single MiSeq run was used to sequence all seven time points and the two IgA and IgG plasmid controls. Between 1.6 and 1.9 million raw reads were obtained from each time point with IgA and IgG sequences being equally distributed (Table 2.1). The most sequences were obtained at 5 weeks post-infection and the least at 17 weeks post-infection. The plasmid controls for IgA and IgG gave 109,556 and 168,527 paired reads, respectively.

Table 2.1: Number of reads obtained from the pooled MiSeq run containing all seven time points and plasmid controls

Sample	Number of raw reads	Number of paired reads	Number reads assigned to IgA	Number reads assigned to IgG
5 w.p.i	1,958,859	1,289,179	520,690	505,821
8 w.p.i	1,765,059	1,191,063	521,546	427,381
11 w.p.i	1,717,050	1,155,963	492,013	427,454
17 w.p.i	1,660,629	1,114,587	466,903	422,808
34 w.p.i	1,709,224	1,215,663	518,297	469,004
38 w.p.i	1,684,162	1,105,684	472,869	401,787
121 w.p.i	1,836,466	1,265,332	276,696	310,461
IgA control	251,348	168,527	168,527	NA
IgG control	193,182	109,556	NA	109,556

Shown are the number of paired reads obtained from the first MiSeq run containing all time points and plasmid controls. The cDNA samples are coloured according to time point.

2.4.2. CAP88-CH06 Plasmid Controls

NGS is known for sequencing errors, which are often difficult to identify. We therefore used a CAP88-CH06 plasmid sequence with a hybrid IgA and IgG constant region as a control to identify the level of PCR and sequencing error. The plasmid was PCR amplified with the IgA and IgG reverse primers separately and sequenced under the same conditions as the seven cDNA samples taken from CAP88. Following sequencing, the reads were mapped back to the original plasmid sequence and identity-divergence plots used to show the percentage identity of the sequences to the CAP88-CH06 plasmid and the percentage of divergence or somatic hypermutation of the sequences from the germline gene (IGHV4-39) used by CAP88-CH06 (Figure 2.2).

The majority of the sequences (~80%) were ≥99% identical to CAP88-CH06 with ~8.8% somatic hypermutation. The remaining 20% of sequences had PCR or sequencing errors resulting in decreased identity to CAP88-CH06 (80-99%) and increased divergence from the germline (5-40%). Using a two-step clustering process, we removed 99.9% of the reads that contained sequencing errors (Figure 2.2). Some error containing reads remained in the IgG control even after the quality control (forming a "t" shape in Figure 2.2D).

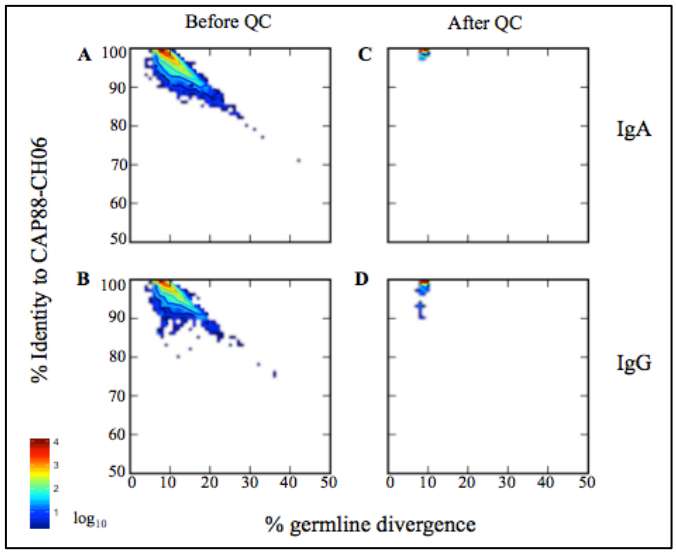


Figure 2.2: Identity-divergence plots for CAP88-CH06 IgA and IgG plasmid controls

Shown are the percentage identity to CAP88-CH06 on the y-axis and the percentage divergence from the germline IGHV4-39 gene on the x-axis for each of the sequences obtained from the plasmid amplification using the IgA and IgG reverse primers before and after correcting for sequencing and PCR error. The $\log_{10}$ scale bar represents the number of sequences in the plot.

2.4.3. CAP88-CH06 Clonally Related Antibodies

An antibodyomics pipeline (described in Section 2.3.5) was used to assign V and J gene usage to the paired reads obtained from each of the seven time points, as well as the percentage identity to CAP88-CH06 and percentage divergence from IGHV4-39. Identity-divergence plots were used to visualise the distribution of sequences obtained from each time point based on percentage identity to CAP88-CH06 and divergence from IGHV4-39. These sequences represent the raw data, which have not undergone the two-step clustering process (Figure 2.3A).

Within the IgA sequences, at 5 and 8 weeks post-infection we observed a small group of raw sequences with >90% identity to CAP88-CH06 with ~0% divergence from the germline. In addition we found a group of raw sequences that were 100% identical to CAP88-CH06 mAb with 8.8% divergence, each represented by a single unique sequence after QC (highlighted within the red block in Figure 2.3A). These two groups were also present in the IgA sequences at 11 weeks post-infection. By 17 weeks post-infection and throughout infection the germline-like group of sequences (~91% identity to CAP88-CH06

and ~0% divergence) disappeared and only the CAP88-CH06-like sequences (100% identity to CAP88-CH06 and 8.8% divergence) remained. At 34 weeks post-infection three distinct groups of sequences were observed; one which was CAP88-CH06 like, one with lower levels of divergence from the germline (6.7%) and one with increased levels of divergence from the germline (9.1%) compared to CAP88-CH06.

In order to confirm the presence of the fully matured CAP88-CH06 antibody at the early time-points, two additional MiSeq runs were performed. New PCR amplicons were amplified from the 5 and 8 week samples separately and sequenced on different MiSeq runs. Approximately 2.1 and 1.6 million reads were obtained from 5 and 8 weeks, respectively. Following the same two-step clustering process used in the first round of sequencing, we observed IgA transcripts that were identical to CAP88-CH06 in both time-points confirming the presence of the mature antibody at 5 and 8 weeks of infection.

Within the IgG group we observed sequences with 90-92% identity to CAP88-CH06 with less than 5% divergence from 11-17 weeks post-infection. At 34 weeks post-infection very few sequences were observed with >90% identity to CAP88-CH06 and by 38 weeks post-infection there were no sequences with >90% identity to CAP88-CH06.

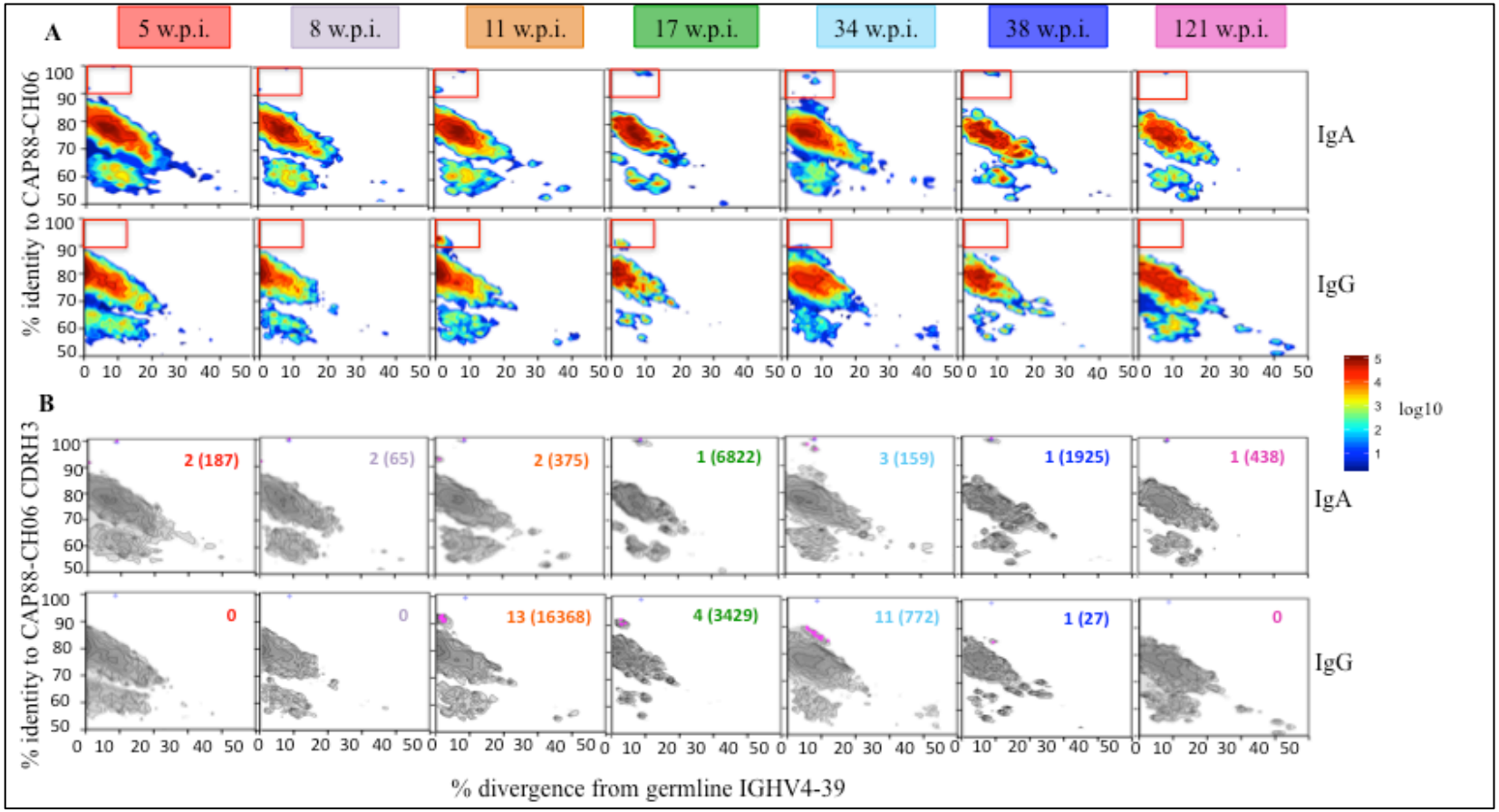


Figure 2.3: Identity-divergence plots of CAP88 cDNA sequences over 2 years of infection

A) Shown is the percentage identity to CAP88-CH06 (y-axis) and percentage divergence from the germline IGHV4-39 gene (x-axis) for each of the sequences obtained from all seven time points. The red block highlights clonally related sequences. **B)** Shown are the sequences with ≥ 80% identity to the CDRH3 of CAP88-CH06 highlighted as pink dots and the number of unique sequences given at the top right-hand corner for each time point, with the total number of related sequences given in parentheses. The $\log_{10}$ scale bar represents the number of sequences in the plot.

In order to identify additional clonally related sequences to CAP88-CH06 a lower cut-off of ≥80% identity to the CAP88-CH06 CDRH3 was used as a filter. The number of unique clonally related sequences (obtained following the two-step clustering) for each time point are given in the top right corner of each plot and highlighted in pink in Figure 2.3B. The coverage of the raw clonally related sequences for each time point are given in parenthesis. We observed clonally related sequences throughout infection within the IgA group, with 5, 8 and 11 weeks post-infection having 2 unique clonally related sequences, one of which represented the germline-like sequences (0.3% divergence and 91.9% identity to CAP88-CH06), and the other matching CAP88-CH06 (8.8% divergence and 100% identity). At 17, 38 and 121 weeks post-infection only one unique sequence was observed, which was identical to CAP88-CH06. At 34 weeks post-infection 3 unique sequences were observed; one which was identical to CAP88-CH06, a second with 6.7% divergence and 97.9% identity and a third with 9.1% divergence and 96.1% identity (Table 2.2).

Within the IgG group, clonally related sequences were only observed from 11 to 38 weeks post-infection, all of which had ≤93% identity to CAP88-CH06. At 11 weeks post-infection we observed 13 unique sequences with 1.7-2.8% divergence and 91.1-93.2% identity to CAP88-CH06. At 17 weeks post-infection only 4 unique sequences were observed with 2.8-4.1% divergence and 90.3-92.1% identity to CAP88-CH06. The number of clonally related sequences increased from 17-34 weeks post-infection. However the identities of these sequences to CAP88-CH06 were lower and they showed more divergence (84.6-90% and 5.8-11.7%, respectively) at 34 weeks post-infection compared to 17 weeks post-infection. At 38 weeks post-infection only one unique clonally related sequence was observed with 85% identity to CAP88-CH06 and 9.3% divergence.

Table 2.2: CAP88-CH06 clonally related IgA and IgG sequences

Time - Point	IgA			IgG		
	Number of Unique Sequences	% Identity to CAP88-CH06	% Divergence from IGHV4-39, (coverage)	Number of Unique Sequences	Identity to CAP88-CH06 (%)	% Divergence from IGHV4-39, (coverage)
5 weeks	2	91.9	0.3 (86)	0		
		100	**8.8 (13)**			
8 weeks	2	91.9	0.3 (14)	0		
		100	8.8 (8)			
11 weeks	2	92.7	1.4 (192)	13	90.8	2.1 (260)
					91.1	2.4 (544)
					91.6	1.7 (31)
					91.6	2.8 (208)
					91.9	1.7 (138)
					91.9	2.1 (49)
		100	**8.8 (39)**		91.9	2.1 (6)
					92.1	1.7 (4)
					92.1	2.4 (4)
					92.4	1.7 (20)
					92.4	2.8 (9)
					92.7	1.4 (8949)
					93.2	2.1 (624)
17 weeks	1	**100**	**8.8 (5462)**	4	90.3	3.1 (620)
					90.6	2.8 (128)
					90.8	4.1 (685)
					92.1	2.8 (593)
34 weeks	3	96.1	9.1 (51)	11	84.6	11.7 (4)
					85.3	10.0 (39)
					85.8	9.3 (69)
		97.9	6.7 (20)		86.4	9.6 (7)
					86.6	8.6 (85)
					87.1	7.9 (27)
					87.4	7.2 (29)
		100	**8.8 (19)**		87.9	7.9 (73)
					88.7	6.5 (3)
					89.2	5.8 (19)
					90.0	5.8 (27)
38 weeks	1	**100**	**8.8 (1349)**	1	85.0	9.3 (12)
121 weeks	1	100	8.8 (298)	0		

Shown are the numbers of unique clonally related sequences for each time point for IgA and IgG isotypes. The numbers in parenthesis represent the total number of raw reads obtained for each unique sequence. Blocks highlighted in grey and text in bold represent the sequences that were identical to CAP88-CH06.

2.4.4. CAP88-CH06 Phylogeny

All unique clonally related sequences (listed in Table 2.2) obtained by selecting those using IGHV4-39 and IGHJ4, with ≥80% sequence identity to CAP88-CH06 CDRH3 and following the two-step clustering process, were used to create phylogenetic trees to study the evolution of this lineage.

The IgA clonally related sequences showed limited evolution with the mature antibody (CAP88-CH06) detected at 5 weeks of infection as a fully mature antibody with 8.8% divergence that was maintained throughout infection (Figure 2.4). There was evidence of low divergence antibodies from 5-11 weeks post-infection (0.3-1.4% divergence) but no intermediate sequences (with divergence between 2-8.8%), with the exception of one sequence detected at 34 weeks post-infection. We were not able to distinguish between the two subtypes of IgA, namely IgA1 and IgA2 within these clonally related sequences, due to the primers used. It is however likely that these sequences are IgA1, since CAP88-CH06 was isolated as an IgA1 antibody.

Within the IgG sequences a greater level of evolution was observed (Figure 2.5). IgG is sub-divided into four subtypes (IgG1-4). Based on the primers used we determined that none of the IgG clonally related sequences were IgG4 but could be IgG1, IgG2 or IgG3 (we were not able to distinguish between these subtypes due to the primers used). CAP88-CH06 was included into this phylogenetic tree, even though the antibody was not observed as an IgG isoform, to give an indication of where the antibody would cluster. There appeared to be two distinct clusters of antibodies arising from the IgG group, the first (observed at the top of the phylogenetic tree) included two sequences from 11 weeks post-infection, three from 17 weeks post-infection and one each from 34 and 38 weeks post-infection. The second much larger cluster is composed mostly of 11 and 34 weeks post-infection samples and also contains the native CAP88-CH06 sequence.

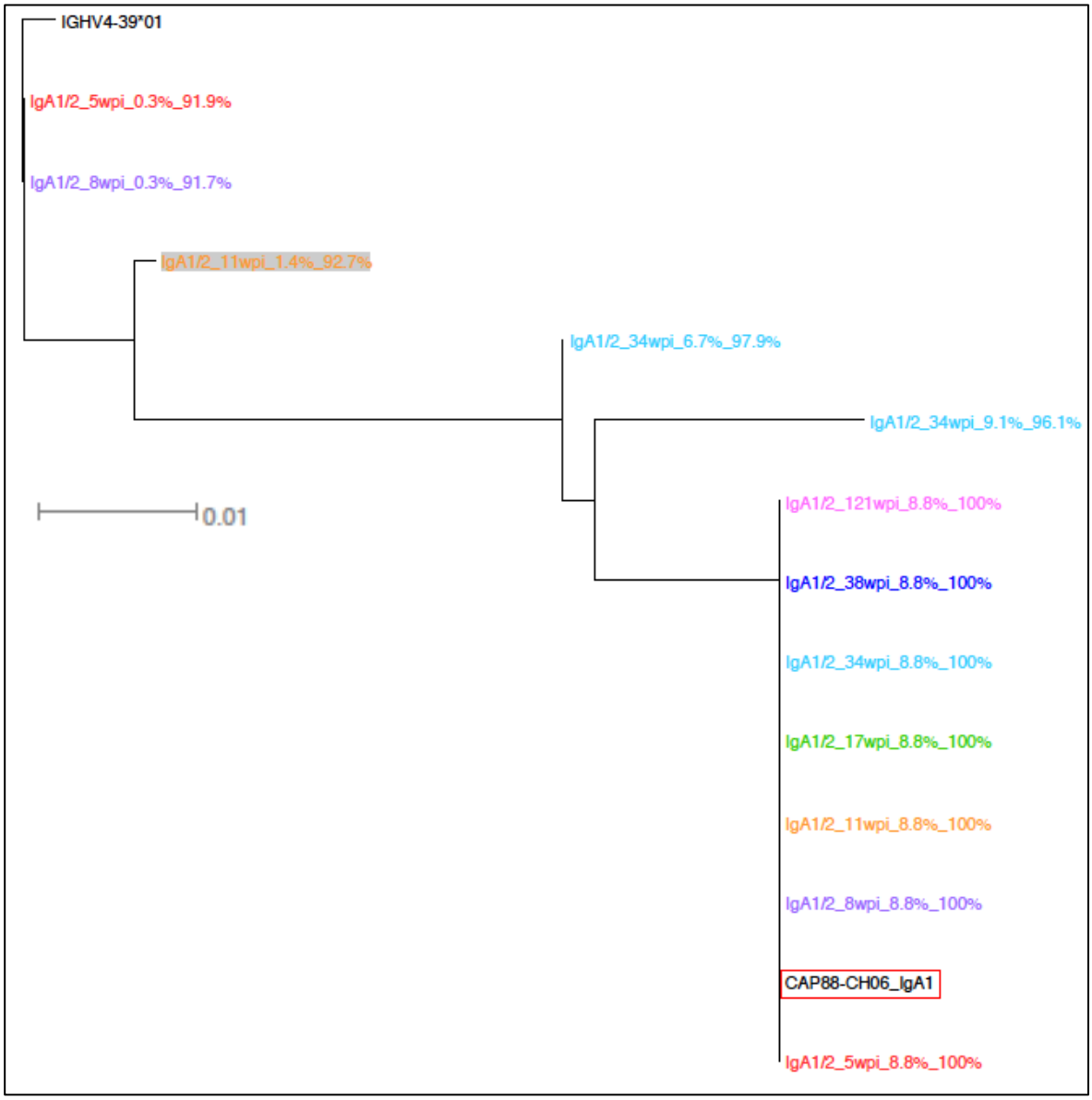


Figure 2.4: Phylogenetic tree using clonally related IgA sequences

Shown is a outgroup rooted (with IGHV4-39*01, top) phylogenetic tree of the 12 clonally related sequences to CAP88-CH06 (red square box). Sequences are coloured according to time point: 5 w.p.i (red), 8 w.p.i (purple), 11 w.p.i (orange), 17 w.p.i (green), 34 w.p.i (light blue), 38 w.p.i (dark blue) and 121 w.p.i (pink). Sequences are labelled according to isotype, time point, percentage divergence and percentage identity to CAP88-CH06. The 1.4%-divergent sequence observed in both IgA and IgG transcripts is highlighted in grey.

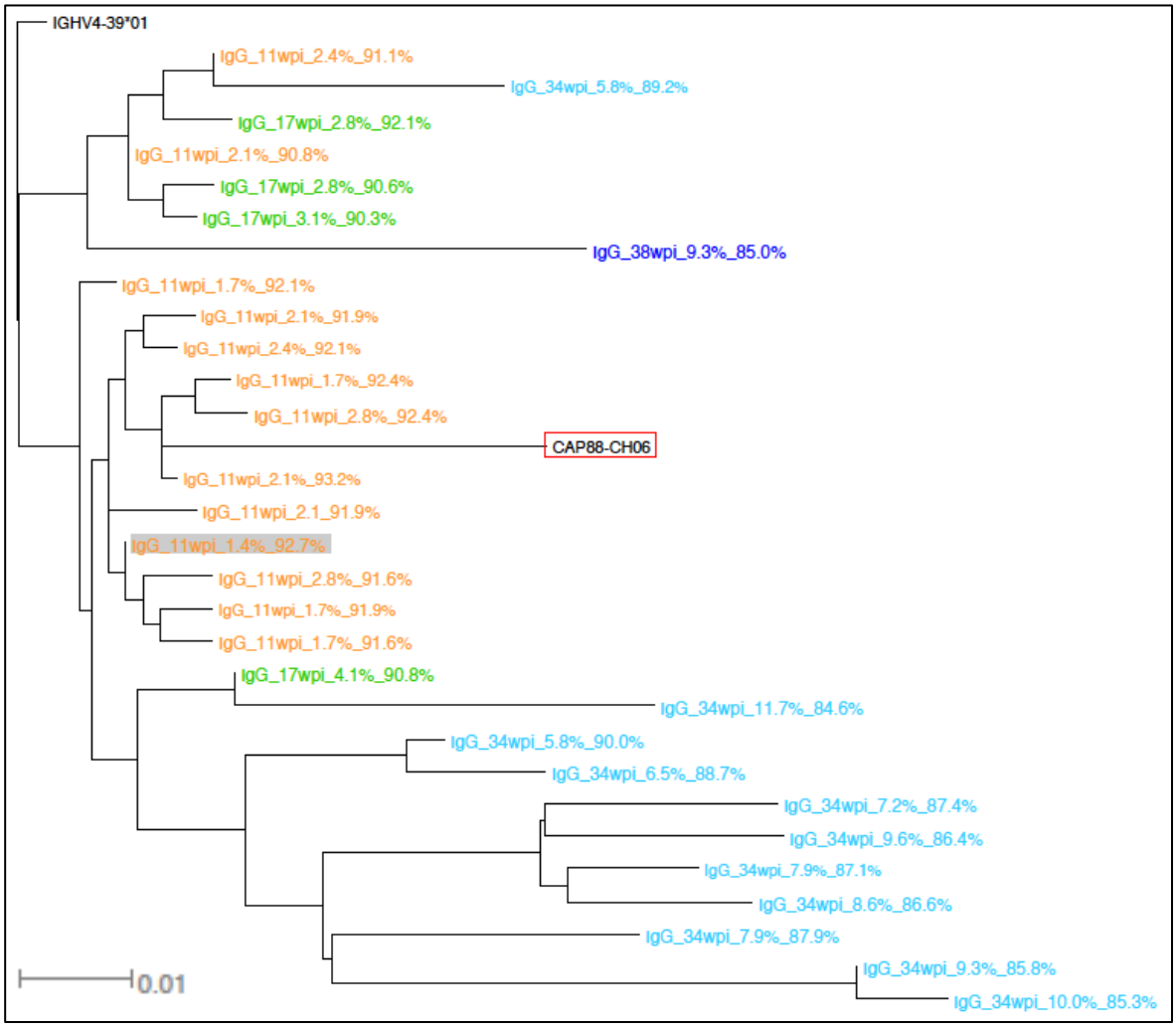


Figure 2.5: Phylogenetic tree using clonally related IgG sequences

Shown is a phylogenetic tree of the 29 clonally related sequences to CAP88-CH06 (red square box). Sequences are coloured according to time point: 11 w.p.i (orange), 17 w.p.i (green), 34 w.p.i (light blue) and 38 w.p.i (dark blue). Sequences are labelled according to isotype, time point, percentage divergence from IGHV4-39 and percentage identity to CAP88-CH06. The 1.4% divergent sequence that also found in IgA is highlighted in grey.

When joining isotypes to show the evolution of all clonally related sequences, we observed the lowest divergence (0.3-1.4%) sequences as an IgA isotype from 5 and 8 w.p.i (Figure 2.6). From 11 w.p.i the two isotypes form two distinct clusters. The first cluster includes the mature antibody (CAP88-CH06), which remained unchanged from 5 to 121 w.p.i. The second cluster represents the IgG isoform, which appeared from 11 w.p.i with

1.4% divergence and evolved throughout infection up to 9.3% at 38 weeks post-infection after which it was no longer detectable.

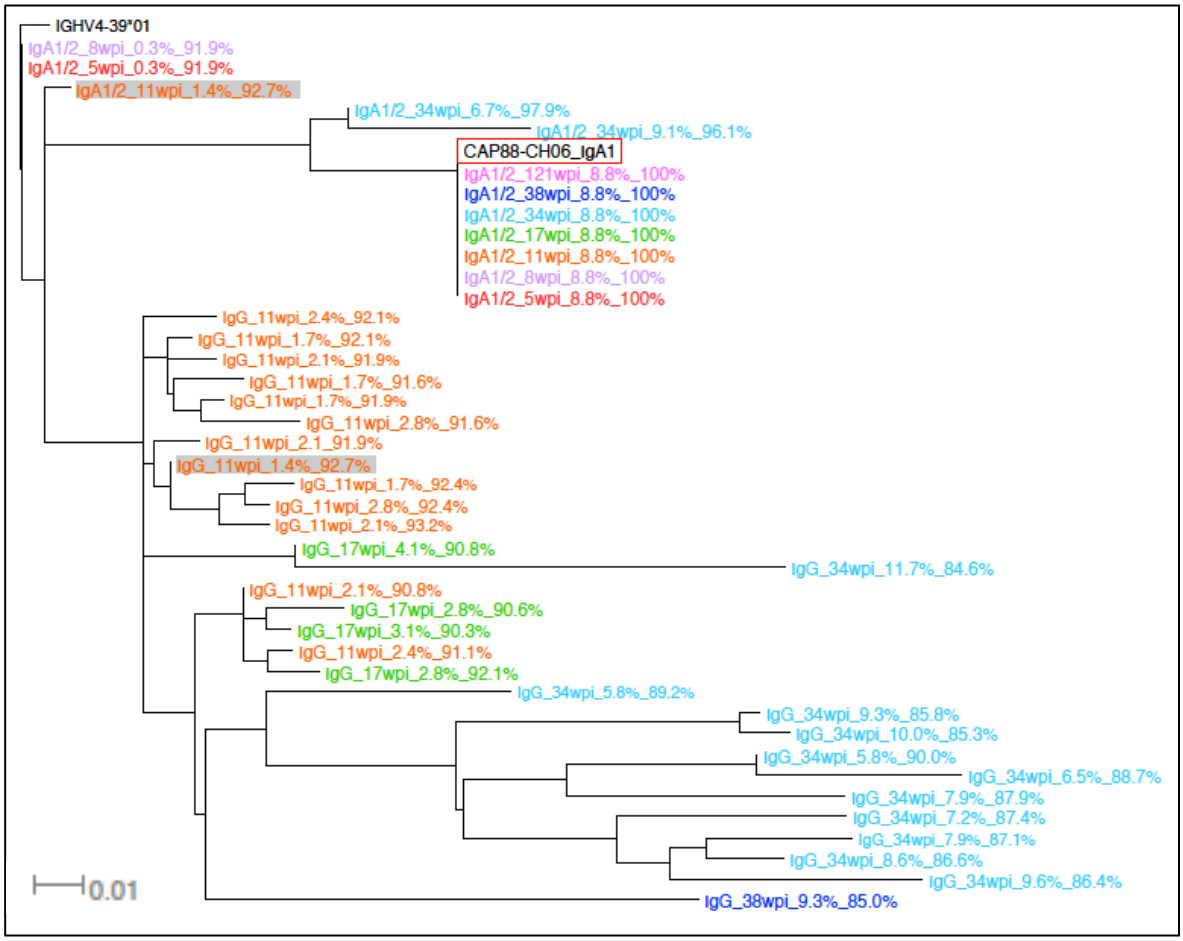


Figure 2.6: Phylogenetic tree of all clonally related sequences to CAP88-CH06

Shown are each of the sequences (coloured according to time point: 5 w.p.i (red), 8 w.p.i (purple), 11 w.p.i (orange), 17 w.p.i (green), 34 w.p.i (light blue), 38 w.p.i (dark blue) and 121 w.p.i (pink)) that are clonally related to CAP88-CH06. Sequences are labelled according to isotype, time point, percentage divergence from IGHV4-39 and percentage identity to CAP88-CH06. The 1.4%-divergent sequence that was found in both IgA and IgG is highlighted in grey.

2.4.5. CAP88-CH06 Lineage Evolution

All clonally related sequences were aligned according to the percentage divergence, with the less divergent sequences shown first and the SHM compared to the germline IGHV4-39*01 and IGHJ4*02 genes (Figure 2.7). Two sequences were seen at multiple time-points and were thus only represented once with the time-points given in the sequence name and coloured accordingly. SHM was largely observed within the CDRH regions (highlighted in yellow) and FR3 with the majority occurring in the CDRH2 with all of the amino acids covering this region showing SHM. This was followed by CDRH1, CDRH3 and FR3 having 92% (12/13), 61% (11/18) and 59% (22/38) of the amino acids in each region subject to SHM during evolution, respectively. We observed no SHM within the constant regions that were sequenced.

The majority of the somatic hypermutations observed within these sequences were single amino acid substitutions, however, some indels were observed. A three amino acid insertion (NSI at positions 30-32) in the CDRH1 was seen in one sequence obtained at 34 weeks post-infection. A two amino acid insertion (RP, positions 61 and 62, highlighted in the dark yellow box in Figure 2.7) was observed within 3 IgA sequences, including the CAP88-CH06 antibody sequence. A deletion of a Serine (S) was observed at position 34 in one IgG sequence obtained from 38 weeks post-infection. The highest level of divergence seen within these sequences was 11.7%, seen at 34 w.p.i within the IgG transcripts.

The least mutated sequence was observed in the IgA transcripts obtained from 5 and 8 weeks post-infection with 0.3% divergence. This sequence had a single amino acid change (valine, V) at position 2 with no other SHM compared to the germline gene. We also observed an identical sequence with low divergent sequence (1.4%) in both the IgA and IgG transcripts obtained at 11 w.p.i (highlighted in the red box in Figure 2.7). This transcript links the two isotypes in this lineage.

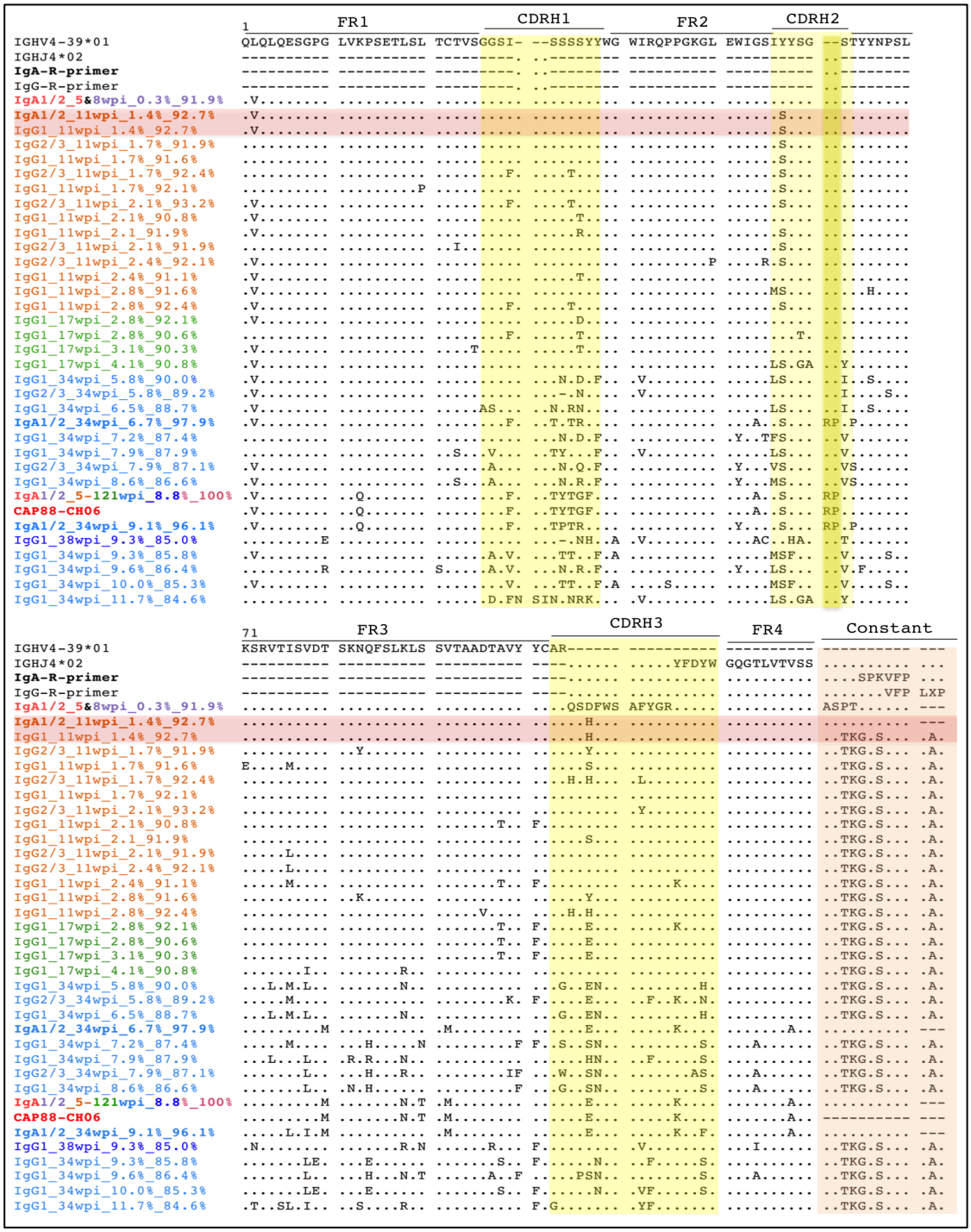


Figure 2.7: Alignment of CAP88-CH06 clonally related sequences

Shown is an alignment of all IgA (bold) and IgG clonally related sequences compared to the germline IGHV4-39*01, IGHJ4*02, IgA and IgG primers used. The names of the sequences are coloured according to time point (5 w.p.i in light red, 8 w.p.i in purple, 11 w.p.i in orange, 17 w.p.i. in green, 34 w.p.i in light blue, 38 w.p.i in dark blue, 121 w.p.i in pink and CAP88-CH06 in dark red). Two sequences were seen in multiple time-points and are shown in multi-colours representing the different time-points. The CDRH regions are highlighted in yellow and the constant region highlighted in orange.

2.5. DISCUSSION

We have tracked the evolution of a potent strain-specific anti-HIV antibody (CAP88-CH06) during the course of infection. CAP88-CH06 is an IgA1 antibody with 8.8% divergence from the germline IGHV4-39 gene, targeting the C3-V4 region of gp120. Our sequencing data revealed IgA transcripts that were identical to CAP88-CH06 at all seven time-points (5, 8, 11, 17, 34, 38 and 121 weeks post-infection) examined. Very little evolution was observed within the IgA transcripts during infection. The putative UCA of the lineage was detected within the IgA transcripts at 5 and 8 weeks post-infection. At 11 weeks post-infection an identical low divergent (1.4%) sequence was observed in both the IgA and IgG transcripts. The limited evolution of the IgA transcripts and the disappearance of the IgG transcripts during infection could be a reason why this particular lineage never became broadly neutralizing.

Previous studies from this participant showed plasma neutralization of the C3-V4 region from 15 weeks post-infection, which peaked at 26 weeks post-infection and then waned but remained detectable throughout infection (last time point tested was 108 weeks post-infection, Figure 2.1) (Moore*, et al.*, 2009a; Gray*, et al.*, 2011a). Viral escape was detected by 26 weeks post-infection, as a result of a glycan introduced at position 339 and a Glutamic Acid (E) to Lysine (K) change at either position 343 or 350. Escape was followed by a second neutralization response targeting the V1V2 region (Moore*, et al.*, 2009a; Gray*, et al.*, 2011a). The strain-specific antibody (CAP88-CH06) responsible for the C3-V4 neutralization was isolated at 34 weeks post-infection as an IgA1. The detection of the mature CAP88-CH06 antibody with 8.8% divergence in the IgA sequences from 17 - 121 weeks post-infection in this study is consistent with the observation of plasma

neutralization of the C3-V4 region despite viral escape (Moore*, et al.*, 2009a; Gray*, et al.*, 2011a). However we detected transcripts identical to CAP88-CH06 at 5 and 8 weeks post-infection even though the earliest plasma neutralization response was at 15 weeks post-infection. We have previously noted inconsistent neutralization at 11 weeks (unpublished) suggesting that the TZM-bl neutralization assay used may not be sensitive enough to detect low levels of neutralization or that levels of antibody circulating at 5 and 8 weeks post-infection were insufficient to cause neutralization of the virus.

The ability of an antibody to acquire 8.8% SHM within 5 weeks of HIV infection as we observed with CAP88-CH06 has been noted in other studies. Analysis of autologous responses against gp41 during acute HIV-1 infection showed SHM levels averaging 5% after 17 to 30 days post-infection, with some responses having as much as 15% SHM (Liao*, et al.*, 2011). In non-human primate studies, antibodies with an average of 4-7% divergence were obtained from 6 weeks post $SHIV_{AD8}$-infection, with one animal having an average of 13.2% SHM after 8 weeks post-infection (Francica*, et al.*, 2015). Antibodies against influenza vaccination from newly recruited B-cells had on average 6.4% divergence after 7 days, with some antibodies having as high as 10% SHM (Wrammert*, et al.*, 2008). Likewise antibodies elicited following the Pneumovax®23 had on average 6.6% SHM after 7 days (Smith*, et al.*, 2013). We confirmed the presence of mature CAP88-CH06 transcripts at both 5 and 8 weeks post-infection through the re-amplification of these time-points and re-sequencing with two separate MiSeq runs. These data suggest that upon HIV infection the CAP88-CH06 UCA acquired 8.8% SHM within 5 weeks and thereafter arrested its evolution. A study on early antibody responses to HIV infection has showed that many gp41-targetting antibodies were polyreactive and bound bacterial antigens (Liao*, et al.*, 2011). This data suggested that these responses were a result of a memory response to non-HIV-1 antigens that was activated by HIV. Thus it is also a possibility that this

lineage may have been present prior to HIV infection and became activated upon infection. Pre-infection samples were not available from this participant to explore this possibility, however, polyreactivity of this lineage could be tested.

The sequence with very low divergence (0.3%) is likely to represent a close relative of the CAP88-CH06 UCA. The primers used to amplify these antibodies contained both a Valine (V) and Leucine (L) at position 2, thus the UCA is likely to have a Leucine at this position as is found in the germline IGHV4-39 sequence. The presence of this sequence at 5 weeks could be due either to plasma cells and/or memory cells in response to early HIV infection (before 5 weeks post-infection). We were not able to distinguish between plasma and memory responses since total RNA was isolated from PBMCs.

The lack of intermediate IgA sequences between the putative UCA and mature antibody might suggest that the B-cell remained in the germinal centre acquiring mutations through SHM before leaving the lymph node expressing the high affinity B-cell receptor. Alternatively, intermediates could have been missed as a result of limited sequencing depth or due to our rigorous quality control criteria. One IgA sequence obtained at 34 weeks post-infection showed 9.1% divergence suggestive of some evidence of evolution of the IgA transcripts. However, since transcripts of the 8.8% CAP88-CH06 antibody were detected after this time-point (at 38 and 121 weeks), the 9.1%-divergent sequence could also be the result of sequencing or PCR error that was not corrected through the two-step clustering process or an off-track antibody that later disappeared. Additional investigations including expressing this more mutated antibody would help address this question.

Indels have been identified in a number of bNAbs (Wu*, et al.*, 2010; Walker*, et al.*, 2011). A study examining indels in a large number of HIV antibodies showed that these were associated with chronic HIV infection and not restricted to bNAbs (Kepler*, et al.*, 2014). In this study we reveal that indels are also found in strain-specific lineages.

Furthermore neutralization of CAP88-CH06 was enhanced by the 2 amino acid insertion in the CDRH2 (data unpublished). Mutations in the framework regions have also been found to be important for neutralization (Klein*, et al.*, 2013). Within the CAP88-CH06 lineage we observed many mutations within the third framework (FR3). However, the affects of these mutations on neutralization are yet to be established.

The presence of both IgA and IgG isotypes suggest a complex process of B-cell activation, maturation and class-switching within this lineage. During the process of class-switch recombination some of the constant region genes are spliced out of the genome in an irreversible manner. The order of the constant region genes in the genome therefore determines whether certain class-switches can occur, enabling a switch forward but not backward. The order of the constant region genes are IgM/IgD, IgG3, IgG1, IgA1, IgG2, IgG4, IgE and IgA2 (Kinoshita*, et al.*, 2001). The lack of clonally related IgG transcripts at 5 and 8 weeks p.i. and the identification of an identical sequence at 11 weeks p.i. in both the IgA and IgG isotypes suggest that the IgG antibodies arose some time between 8 and 11 weeks post-infection and that the IgA antibodies were present before the IgG. Since it would not be possible to class-switch from an IgA1 to an IgG1 or IgG3 it is likely that the IgG transcripts in this lineage are IgG2. Thus following activation (by the virus) between 8-11 weeks, the B-cell moved to a germinal centre where it underwent affinity maturation and class-switched from an IgA1 to an IgG2 isotype. Teasing apart the class-switches and isotype specificity would require re-sequencing of the lineage using different IgA and IgG primers which would enable one to distinguish between the different IgA and IgG isotypes.

It is unclear why CAP88 never developed neutralization breadth ever after 5 years of infection. The CAP88-CH06 mAb isolated from this individual did not show high level of somatic hypermutation or long CDRH3 associated with bNAbs. This mAb also showed limited evolution and disappearance of the IgG isotype. Studies are ongoing to determine

the interplay between the virus and antibody in CAP88. Two previous longitudinal studies on individuals who develop bNAbs have shown that viral diversity preceded the emergence of breadth (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014). More specifically mAbs from the CAP256-VRC26 lineage able to tolerate a diversity of immunotypes within the epitope show the greatest neutralization breadth (Bhiman*, et al.*, 2015). Virological studies in CAP88 may shed light on why this antibody failed to develop breadth.

2.6. CONCLUSIONS

We have studied the evolution of an IgA1 autologous strain-specific antibody (CAP88-CH06) during HIV-1 infection. This antibody targets the C3-V4 region of gp120 and plasma neutralization was first detected at 15 w.p.i. The putative unmutated common ancestor of this lineage was observed at 5 and 8 w.p.i as an IgA isotype. Transcripts identical to CAP88-CH06 were observed within all seven time-points sequenced. The presence of CAP88-CH06 sequences well after viral escape (over 2 years post-infection) was consistent with plasma neutralization of this region within CAP88, from which this antibody was isolated. The IgA sequences showed very little evidence of evolution during infection. An identical sequence was observed as both an IgA and IgG isotype at 11 weeks post-infection and likely represents a BCR that was activated between 8-11 weeks, which gave rise to a transient clonally related clade of IgG antibodies. The IgG isotype could be targeting quasispecies circulating during infection, which disappears when the V1V2 neutralization response is well established. The lack of continued evolution of the IgA isotype and the disappearance of the IgG isotype could be the reason why this autologous antibody lineage failed to produce breadth later in infection.

2.7. REFERENCES

Altschul, SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25 (17): 3389-3402

Bhiman, JN*, et al.* (2015) Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nature Medicine,* In Press

DeKosky, BJ*, et al.* (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology,* 31 (2): 166-169

Doria-Rose, NA*, et al.* (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature,* 509 (7498): 55-62

Edgar, RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26 (19): 2460-2461

Euler, Z and Schuitemaker, H (2012) Cross-reactive broadly neutralizing antibodies: timing is everything. *Frontiers in Immunology,* 3: 215

Francica, JR*, et al.* (2015) Analysis of immunoglobulin transcripts and hypermutation following SHIVAD8 infection and protein-plus-adjuvant immunization. *Nature Communications,* 6: 6565

Galtier, N*, et al.* (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computational Applied Biosciences,* 12 (6): 543-548

Gao, F*, et al.* (2014) Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell,* 158 (3): 481-491

Gray, ES*, et al.* (2007) Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *Journal of Virology,* 81 (12): 6187-6196

Gray, ES*, et al.* (2011a) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology,* 85 (10): 4828-4840

Gray, ES*, et al.* (2011b) Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. *Journal of Virology,* 85 (15): 7719-7729

Haynes, BF*, et al.* (2012b) B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnology,* 30 (5): 423-433

Huson, DH and Scornavacca, C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology,* 61 (6): 1061-1067

Kepler, Thomas B*, et al.* (2014) Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host & Microbe,* 16 (3): 304-313

Kinoshita, K and Honjo, T (2001) Linking class-switch recombination with somatic hypermutation. *Nature Reviews Immunology,* 2: 493-503

Klein, F*, et al.* (2013) Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell,* 153 (1): 126-138

Liao, H-X*, et al.* (2011) Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *Journal of Experimental Medicine,* 208 (11): 2237-2249

Liao, H-X*, et al.* (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature,* 496 (7446): 469-476

Moore, PL*, et al.* (2008) The C3-V4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *Journal of Virology,* 82 (4): 1860-1869

Moore, PL*, et al.* (2009a) Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathogens,* 5 (9): e1000598

Moore, PL, *et al.* (2009b) Specificity of the autologous neutralizing antibody response. *Current Opinion in HIV and AIDS,* 4 (5): 358-363
Moore, PL, *et al.* (2012) Evolution of an HIV glycan–dependent broadly neutralizing antibody epitope through immune escape. *Nature Medicine,* 18 (11): 1688-1692
Moore, PL, *et al.* (2015) Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends in Microbiology,* 23 (4): 204-211
Sievers, F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7: 539
Smith, K, *et al.* (2013) Fully human monoclonal antibodies from antibody secreting cells after vaccination with Pneumovax®23 are serotype specific and facilitate opsonophagocytosis. *Immunobiology,* 218 (5): 745-754
Stamatatos, L, *et al.* (2009) Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature Medicine,* 15 (8): 866-870
Tomaras, GD, *et al.* (2008) Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology,* 82 (24): 12449-12463
Van Loggerenberg, F, *et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One,* 3 (4): e1954
Walker, LM, *et al.* (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477 (7365): 466-470
West, AP, *et al.* (2014) Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell,* 156 (4): 633-648
Wibmer, CK, *et al.* (2013) Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. *PLoS Pathogens,* 9 (10): e1003738
Wibmer, CK, *et al.* (2015) HIV broadly neutralizing antibody targets. *Current Opinion in HIV and AIDS,* 10: 1-8
Wrammert, J, *et al.* (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature,* 453 (7195): 667-671
Wu, X, *et al.* (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science,* 329 (5993): 856-861
Wu, X, *et al.* (2015) Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell,* 161: 1-16
Zolla-Pazner, S (2014) A critical question for HIV vaccine development: which antibodies to induce? *Science,* 345 (6193): 167-168

3. INDUCTION OF SERUM GLYCAN-BINDING ANTIBODIES DURING HIV-1 INFECTION[c]

Cathrine Scheepers[1, 2], Sudipa Chowdhury[3], Shea Wright[3], Christopher T. Campbell[3], Nigel Garrett[4, 5], Quarraisha Abdool Karim[4, 6], Salim S. Abdool Karim[4, 6], Penny Moore[1, 2, 4], Jeffrey C. Gildersleeve[3] and Lynn Morris[1, 2, 4]

[1]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, 2131 South Africa; [2]School of Pathology, Division of Virology and Communicable Disease Surveillance, University of the Witwatersrand, Johannesburg, 2050, South Africa; [3]Chemical Glycobiology Section of the Chemical Biology Laboratory, National Cancer Institute, Frederick, MD, 21702-1201 USA; [4]Centre for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu-Natal, 4013 South Africa; [5]Department of Infectious Diseases, Nelson R. Mandela School of Medicine, University of KwaZulu–Natal, 4041, Durban, South Africa; and [6]Department of Epidemiology, Columbia University, New York City, 10032, USA

Manuscript in preparation

---

[c]This project was supported by the Poliomyelitis Research Foundation (PRF), University of the Witwatersrand Health Sciences Faculty Research Council and the National Research Foundation (NRF). Cathrine Scheepers was supported by the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) through the Fogarty International Center, National Institutes of Health (grant # 5 D43 TW000231). This work was supported in part by the intramural research program of the NIH, NCI. CAPRISA is funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes for Health (NIH), and U.S. Department of Health and Human Services (grant: AI51794). Penny L. Moore is a Wellcome Trust Intermediate Fellow in Public Health and Tropical Medicine (Grant 089933/Z/09/Z). Christopher T. Campbell was supported by a Pharmacology Research Assoicated Training Fellowship, NIGMS.

3.1. ABSTRACT

The HIV-1 envelope is surrounded by glycans that are crucial for structural integrity as well as protecting conserved regions from host antibody responses. However, these glycans are also often the targets of broadly neutralizing antibodies that emerge in some HIV-infected individuals many years after infection. We studied the IgG anti-glycan antibody profiles of 20 HIV-negative and 27 HIV-positive women from the CAPRISA 002 and 004 cohorts (this included 12 individuals who developed broadly neutralizing antibodies and 13 individuals who did not), with the aim of determining whether anti-glycan antibodies are associated with HIV-1 infection or specific to individuals who develop broadly neutralizing antibodies. Amongst the HIV-negative individuals, IgG anti-glycan antibodies fluctuated over a three year period but were nevertheless highly correlated within a single individual. In contrast, HIV-positive individuals had elevated levels of antibodies that bound high mannose *N*-linked glycans including Man8 that increased over time, peaking at three years of infection. Competition experiments using gp120 protein confirmed that at least a proportion of these mannose-specific antibodies were HIV-1 specific. In addition we observed high levels of IgG binding to Tn-peptides and glycolipids throughout HIV infection. However, the levels of these glycan-binding antibodies were not different between individuals who did and did not develop broadly neutralizing antibodies. This indicated that elevated levels of glycan-binding antibodies are a response to HIV-1 infection rather than specific to individuals who develop broadly neutralizing antibodies suggesting that the epitopes of broadly neutralizing antibodies are not uniquely immunogenic.

3.2. INTRODUCTION

The HIV envelope is highly variable, complex and undergoes various conformational changes before and after attachment to the CD4 receptor on T-cells, the target cells during infection (Pancera*, et al.*, 2014). The envelope is covered in viral spikes made up of trimeric gp120 and gp41 glycoproteins (Pancera*, et al.*, 2014). gp120 binds the CD4 receptor on the host's cell membrane enabling viral entry, while gp41 forms a link between gp120 and the viral membrane (Pancera*, et al.*, 2014). Both gp120 and gp41 are heavily glycosylated with between 25-30 *N*-linked glycosylation sites that account for half of the molecular mass of the trimer, known as the glycan shield (Wang, 2013; Pancera*, et al.*, 2014). Although glycans are attached to the viral envelope through the host cell glycosylation pathways, they lack the typical processing methods seen in the host glycosylation pathways, resulting in inefficiently trimmed glycans (Bonomelli*, et al.*, 2011). The glycan shield on trimeric gp120 is mostly comprised of densely packed high mannose, particularly Manα1-2Man terminating glycans such as $Man_{6-9}GlcNAc_2$, and complex *N*-linked glycans although some *O*-linked glycans do occur (Zhu*, et al.*, 2000; Doores*, et al.*, 2010b; Go*, et al.*, 2013). In addition to protecting immunogenic sites on the virus, these glycans are also essential for glycoprotein folding and are thus often conserved (Bonomelli*, et al.*, 2011; Pancera*, et al.*, 2014).

HIV-1 infection has been associated with a wide range of B-cell dysfunctions, including hypergammaglobulinemia (increased levels of circulating IgG), polyclonal activation, malignancies, exhaustion and an increase in subsets of B-cells not found in healthy individuals (Moir*, et al.*, 2014). Many of the polyclonal B-cell responses are not HIV-specific, however, antibodies against HIV-1 do occur throughout infection (Morris*, et*

al., 1998; Tomaras*, et al.*, 2008). The earliest antibody responses to HIV-1 can be detected from ~12 days post-infection and although they are able to bind gp41 on the HIV-1 viral spike they are not able to neutralize the virus (Tomaras*, et al.*, 2008; Euler*, et al.*, 2012). These responses are typically followed by non-neutralizing antibodies that bind monomeric gp120 approximately 2 weeks later (Euler*, et al.*, 2012). The first neutralizing antibodies can be detected from approximately 3 months post-infection and are typically strain-specific (only neutralize the infecting virus) and target variable regions on the viral envelope (Euler*, et al.*, 2012; Moore*, et al.*, 2015). The most relevant antibodies against HIV-1 in terms of vaccine development are broadly neutralizing antibodies (bNAbs), which are capable of neutralizing numerous strains of HIV-1 due to their affinity for conserved epitopes on the envelope (Moore*, et al.*, 2015). These can be detected in natural HIV-1 infection from 2 years post-infection and only in approximately 15-30% of infected individuals (Euler*, et al.*, 2012; Moore*, et al.*, 2015). bNAbs typically target five sites on the viral spike which are the CD4 binding site on gp120, the membrane proximal external region (MPER) of gp41, the C-sheet of the V1V2 domain on gp120, V3/C3 site on gp120 and the interface between gp120 and gp41 (Wibmer*, et al.*, 2015).

Despite the glycan shield serving as a protective barrier against antibody responses, some bNAbs, such as 2G12, PG9, PG16, PGTs121-123, PGTs125-131, PGT135 and PGTs141-145 are able to neutralize various strains of HIV by binding glycans in the C-sheet of V1V2 domain and the V3/C3 site on gp120 (Wang, 2013; Moore*, et al.*, 2015). 2G12 was the first identified bNAb capable of binding glycans and thus most widely studied (Trkola*, et al.*, 1996). It however has fairly limited breadth and potency, while PG9, PG16 and the PGTs form some of the most broad and potent bNAbs, with PGT121-123 and PGT125-128 being the most potent of the glycan-binding bNAbs (Walker*, et al.*, 2011; Wibmer*, et al.*, 2015). 2G12, PGT121-123, PGT125-131 and PGT135 target the

N332 supersite on the V3 and V4 loop of gp120, with 2G12, PGT125-131 and PGT135 binding high mannose *N*-linked glycans such as Man_{7-9} and PGT121-123 binding complex type glycans such as A2, N2 and NA2 (Mouquet*, et al.*, 2012; Wang, 2013; Wibmer*, et al.*, 2015). PG9, PG16 and PGT141-145 target high mannose *N*-linked glycans at the N160 site of V2 region of gp120 with PG9 and PG16 having additional contacts with complex sialylated *N*-linked glycans at N156 or N173 (McLellan*, et al.*, 2011; Walker*, et al.*, 2011; Pancera*, et al.*, 2013).

Over the years glycan arrays have been developed to study the glycan-binding properties of serum antibodies (Von Gunten*, et al.*, 2009; CFG, 2014). These have proven to be particularly useful in determining the types of glycan structures that are bound by HIV-1 bNAbs (Walker*, et al.*, 2011; Mouquet*, et al.*, 2012; Shivatare*, et al.*, 2013). In addition anti-glycan antibody responses in sera of non-human primates after Ad5hr-SIV vaccination and SIV infection (Campbell*, et al.*, 2013) and potential HIV vaccine responses in rabbits have been assayed on arrays (Luallen*, et al.*, 2008; Astronomo*, et al.*, 2010; Dunlop*, et al.*, 2010). Longitudinal human sera studies on glycan arrays are usually limited to 1-3 months and since bNAbs take years to develop during infection it is important to understand anti-glycan antibody profiles in humans over greater time frames.

Here we study serum anti-glycan antibody profiles of HIV-positive individuals (including 12 individuals who make broadly neutralizing antibodies during infection and 13 individuals who do not) before and after infection for three years. We have previously mapped the plasma antibody responses in 8 of the 12 individuals who make broadly neutralizing antibodies, of which 6 target glycans at the V3, V2 or MPER. Our aim was to determine whether elevated levels of anti-glycan antibodies are a response to HIV infection or are unique to individuals who make bNAbs. This information might guide the design of

an immunogen needed for an effective HIV vaccine that stimulates glycan reactive antibodies.

3.3. METHODS AND MATERIALS

3.3.1. Serum Samples

Serum samples used in this study were obtained from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) CAP002 Acute Infection Study (Van Loggerenberg*, et al.*, 2008) and the CAP004 Tenofovir Gel Trial (Abdool Karim*, et al.*, 2010) cohorts. These studies involved adult (> 18 years) women of African ancestry, with a high risk of HIV infection that were being followed in urban and rural clinics in KwaZulu-Natal, South Africa. Sera from 47 women (27 HIV-positive and 20 HIV-negative) were obtained from four time points (pre-infection and yearly for 3 years).

The HIV-positive individuals (infected with HIV-1 subtype C) were classified according to their plasma neutralization breadth at 2-3 years post infection. Those with broadly cross-neutralizing activity were capable of neutralizing 33-94% (median 53%) of a panel of 18 viruses (referred to as BCN individuals) and those with no broadly cross-neutralizing activity, neutralized 0-11% of the viruses (referred to as non-BCN individuals). Those who neutralized 11-30% were termed intermediate neutralizers. The 18 viruses used in the panel were made-up of six subtype A, six subtype B and six subtype C viruses, of which two subtype C viruses were isolated from the women in this cohort. Of the 27 HIV-positive individuals used in this study, 12 were BCN, 13 were non-BCN and two were intermediate neutralizers. The neutralization assays were done routinely in the HIV virology laboratory at the National Institute for Communicable Diseases (NICD) and data provided for this study. Since high viral loads at 6 months post-infection are associated with the development of bNAbs (Gray*, et al.*, 2011b) all BCN and non-BCN

individuals were matched based on their viral load. Information regarding breadth, duration of infection and viral loads for these individuals can be found in Appendix, Supplementary Table 3.1. Samples from 20 HIV-negative individuals were used as a control to assess normal fluctuations in glycan-binding over a three-year period. All human serum samples were heat-inactivated, for 30 minutes in a 56°C water bath, prior to binding on the glycan arrays, as is standard procedure when working with serum or plasma and shown not to affect antibody function (Montefiori, 2009).

Both the parent study involving the CAPRISA cohorts and this particular study were granted ethics clearance under the following certificate numbers: M080470 for the parent study and M111104 for this study.

3.3.2. Array Fabrication and Binding

Glycan array fabrication and binding assays were done in the Chemical Biology Laboratory at the National Cancer Institute in Frederick, MD, USA as part of a Fogarty traineeship with Dr Jeff Gildersleeve. Sudipa Chowdhury and Shea Wright performed repeat assays and additional experiments under the supervision of Dr Jeff Gildersleeve, Dr Christopher Campbell and Dr Saddam Muthana.

These glycan arrays were previously used to assess anti-glycan antibody responses in non-human primates after SIV vaccination and infection (Campbell*, et al.*, 2013). In this study arrays were fabricated as described (Campbell*, et al.*, 2010) with the exception of the addition of a washable fluorescent dye, DyLight™649 (0.7 µg/mL, ThermoScientific), to the print buffer as an indicator of successful liquid deposition and spot morphology and Triton X-100 concentration was 0.0001% instead of 0.006%. The array format and binding assay was analysed for reproducibility and validated with numerous antibodies and lectins

as previously described (Manimala*, et al.*, 2005; Manimala*, et al.*, 2006; Manimala*, et al.*, 2007; Oyelaran*, et al.*, 2009; Campbell*, et al.*, 2010; Wang*, et al.*, 2014b).

The arrays were composed of 337 components (of which 245 were unique), including *N*-linked carbohydrates, glycolipids, and glycopeptides (for a full list of array components, see Appendix, Supplementary Table 3.2). Of the 337 array components, ten were controls (BSA, HSA, linkers used to attach the glycans to the glass slides and the fluorochromes used to detect antibody binding) that were used to assess the technical reproducibility of the arrays and were excluded from downstream analysis. Prior to each experiment, slides were pre-scanned on a Genepix 4000A microarray scanner (Molecular Devices Corporation) and analysed with Genepix Pro 6.0 for technical faults. If the arrays were found to be of good quality a 16-well module (Grace Bio-Lab) was affixed to the slide to create 16 independent array wells allowing 16 samples to be tested per array.

The slides were blocked overnight at 4°C with 3% BSA (w/v) in PBS and washed 6 times with PBST. Samples were diluted in PBST buffer containing 3% bovine serum albumin (BSA, Sigma) and 1% human serum albumin (HSA, Sigma). 50 μL of each sample was run in duplicate on two different wells on different slides. All samples were then incubated at 37°C with gentle agitation (100 RPM) for 4 h. After washing 3 times with PBST, bound antibodies were detected by incubating with DyLight™549-conjugated goat anti-human IgG (Jackson ImmunoResearch Laboratories, Inc.) diluted 1:500 in PBS containing 3% HSA and 1% BSA for 2 h at 37 °C with gentle agitation. After washing 7 times with PBST, slides were removed from modules, immersed in wash buffer for 5 min, and centrifuged at 1000 rpm for 5 min.

Seven well-characterised bNAbs were used to assess the ability of the glycan arrays to detect HIV-specific glycan-binding. PG16 and 2G12 were obtained from the NIH AIDS Reagent Program (NIH AIDS Reagent, 2013). PGT121, PGT125, PGT126, PGT128 and

PGT130 were provided by the IAVI's Neutralizing Antibody Center (Scripps Research Institute, La Jolla, CA (IAVI, 2013)).

Reference serum, consisting of pooled human sera, used as controls to assess reproducibility of the arrays, was purchased from Valley Biomedical Products and Services (Winchester, VA).

3.3.3. Image Analysis and Data Processing

Slides were scanned at 10μm resolution with a Genepix 4000A microarray scanner (Molecular Devices Corporation) and analysed with Genepix Pro 6.0 software as previously reported (Campbell*, et al.*, 2010). Spots were defined as circular features with a diameter of 90μm. The background-corrected median was used for data analysis, and technical faults (e.g. missing spots) were flagged and excluded from further analysis. To minimize the impact of noise in our comparisons, spots with intensities lower than 150 were considered too low to be measured accurately and were set to 150.

For human serum sample data, duplicate spot averages were calculated and normalized to the reference samples. Log-transformation (base 2) was applied for each slide, and the final data value was obtained from the normalized average of data from both slides. For HIV-1 bNAbs, the average of duplicate spots was calculated.

3.3.4. gp120 Competition Assay

To determine if glycan-binding profiles were HIV-specific, sera with high Man_8 binding levels were incubated with purified monomeric ConC gp120 (100 μg/mL) or PBS alone at room temperature for 15 minutes. Monomeric ConC gp120 was expressed as described in Gray et al., (Gray*, et al.*, 2011a) and purified using a *Galanthus Nivalis* lectin (Stratagene).

Sera incubated with or without gp120 were run on the same glycan array slide to minimize variability. A total of three wells (six replicate spots) were analysed.

Since samples incubated with or without gp120 could be compared directly to each other, a well-to-well signal comparison was performed during data analysis to account for batch-to-batch and slide-to-slide variability. After flagging and excluding data with technical faults (e.g. missing spots) from analysis, the average relative fluorescence unit (RFUs) for each individual well was calculated using the two duplicate spot signals. To adjust for printing defects, the maximum and minimum averaged values were determined. If the values for a particular array component differed by more than 20-fold, all data from that array component were removed from further analysis. From the remaining data, the median signal was calculated, adjusted for PMT (photomultiplier tube) voltage, and signals below threshold were set to 150 as described above.

3.3.5. Statistical and Data Analysis

Fold change differences between any two time points, calculated by subtracting one time point (e.g. 1 year) from another (e.g. year 0) and unlogging the data ($2^{(\text{logged value})}$), were used to assess the differences in glycan-binding over time for the HIV-negative samples. Two different techniques were created to assess the glycan-binding profiles during HIV infection: 1) comparing the levels of glycan-binding in the HIV-positive samples to HIV-negative samples and 2) comparing the levels of glycan-binding during HIV infection to matched pre-infection samples.

In Method 1, an HIV-negative range of IgG glycan-binding was determined for each glycan from all pre-infection and HIV-negative samples. Glycans with binding levels ≥2-fold than the top-binding observed in the HIV-negative range in two or more HIV-infected individuals were taken as significant. This method involved all time points from

the 20 HIV-negative individuals, 12 pre-infection samples and all post-infection samples from 27 HIV-positive individuals.

Method 2 focused on 11 HIV-positive individuals that had a pre-infection sample and two consecutive HIV-positive time points. All glycans that showed ≥2-fold increases in glycan-binding over time compared to pre-infection in at least two individuals were considered significant.

A Student's paired t-test was used to compare the binding profiles of three-year post-infection sera from eight individuals incubated with PBS or gp120 and fold-change differences were calculated between groups. Glycans with p-value ≤ 0.05 were considered significant. A Student's two-sampled unpaired t-test was used to calculate the differences between the BCN and non-BCN groups yearly post-infection. Bonferroni correction for multiple testing was applied to comparisons between BCN and non-BCN individuals. Glycans with p-values ≤ 0.05 were considered significant.

3.4. RESULTS

3.4.1. Detection of bNAb Binding on Glycan Arrays

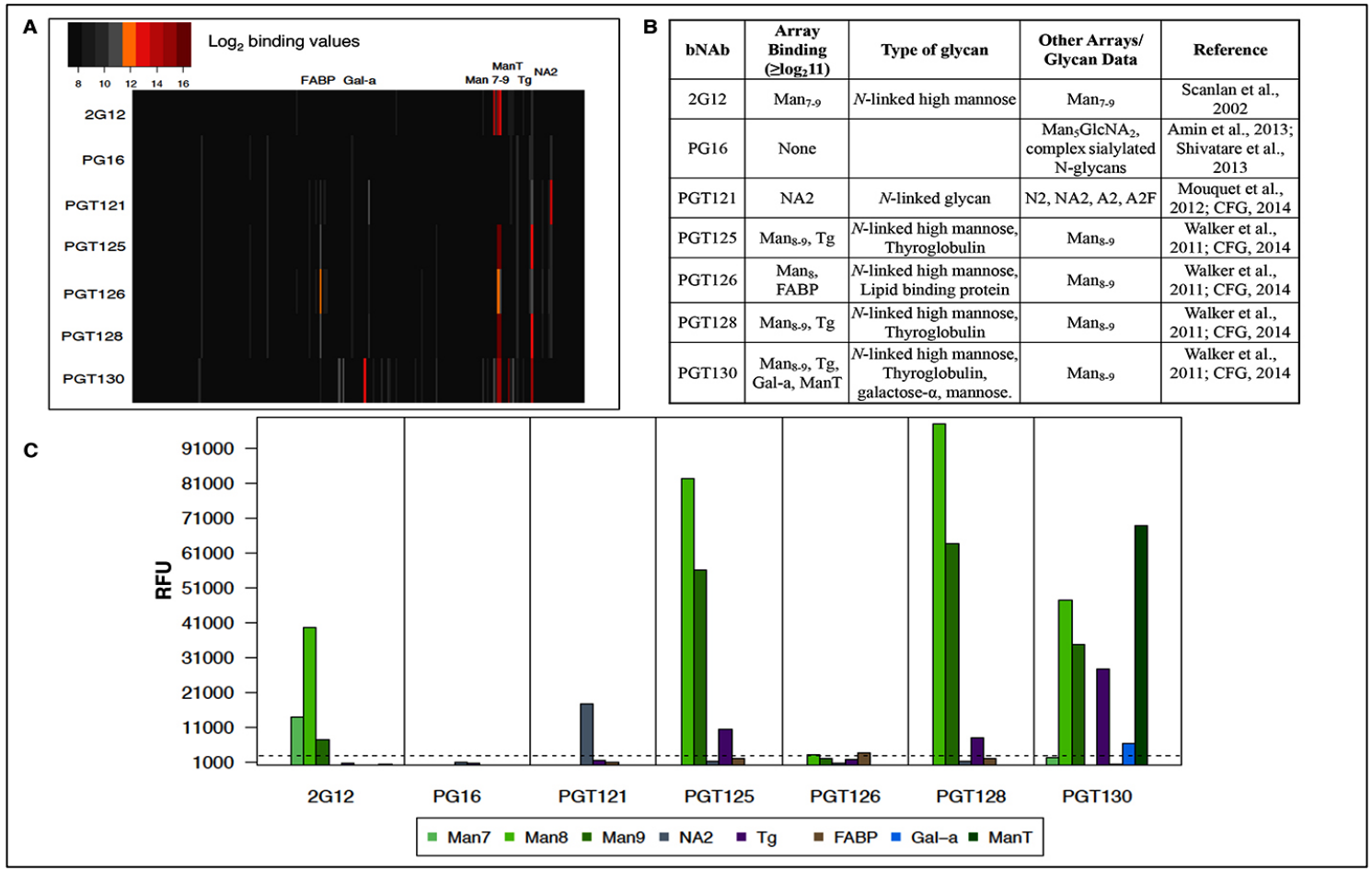
Glycan arrays consisting of 337 (of which 245 are unique) array components spotted onto a glass slide were used to study longitudinal serum anti-glycan IgG binding profiles from HIV-negative and HIV-positive individuals. The components on the array included *N*-linked glycans, Tn-peptides, glycans derived from glycolipids and glycoproteins. Ten controls (including BSA, HSA, the linkers used to attach the glycans to the glass slides and the fluorochromes (Cy3 and Cy5) used to detect binding) were removed from downstream analysis. These arrays were designed to capture predominant glycan antibodies and have previously been used to assess the anti-glycan antibody responses to SIV-vaccination and SIV-infection in non-human primates (Campbell*, et al.*, 2013). Prior to the analysis of serum anti-glycan IgG profiles we assessed binding of HIV-specific anti-glycan antibodies using well-characterised bNAbs. In order to compare the glycan-binding for these bNAbs to other glycan array studies only glycan hits with relative fluorescence values (RFU) ≥ 3000 ($\log_2 11$) were considered, since this was the cut-off used in the previous studies.

The seven bNAbs used were 2G12, PG16, PGT121, PGT125, PGT126, PGT128 and PGT130. The majority of these bNAbs (2G12, PGT125, PGT126, PGT128 and PGT130) target high mannose *N*-linked glycans on the V3 loop of gp120. PGT121 targets complex glycans on the V3 loop of gp120 and PG16 targets both complex and high mannose *N*-linked glycans on the V2 loop of gp120. Six of the seven bNAbs showed binding above $\log_2 11$ for least two glycans (Figures 3.1A and C). The hits observed for these bNAbs were similar to what has been reported in other glycan array experiments

(Walker, *et al.*, 2011; Mouquet, *et al.*, 2012; Shivatare, *et al.*, 2013; CFG, 2014), shown in Figure 3.1B. The only bNAb with no binding above 3000 ($log_2$11) was PG16, which is known to bind a complex trimer specific epitope. A previous study also showed no binding on an array due to the complexity of the epitope (Shivatare, *et al.*, 2013). We detected PGT121 binding to NA2 (A2, A2F and N2 previously associated with PGT121 were not on this array (Walker, *et al.*, 2011; Mouquet, *et al.*, 2012; CFG, 2014)). The majority (5/7) of the bNAbs bound high mannose *N*-linked glycans Man_8 and Man_9. In addition to the glycans previously described we found binding to (Tg (Thyroglobulin), FABP (fatty acid binding protein), ManT (branched Man_3 glycan) and Gal-α (galactose-α). Binding to Tg was observed for 3/7 bNAbs (PGT125, PGT128 and PGT130), whereas the remaining three were recognised by single bNAbs (Figure 3.1).

3.4.2. Natural Fluctuations in Anti-Glycan IgG Antibodies in HIV-Negative Individuals

Prior to the identification of HIV-specific glycan-binding responses, we aimed to understand natural serum IgG anti-glycan antibody variation over a three-year time frame. Previous studies have shown that fluctuations over 1-3 months in healthy individuals are small, and that changes greater than 2.6-fold occur infrequently (Oyelaran, *et al.*, 2009). However, variations in anti-glycan antibody levels over 1-3 year time frames have not been studied. Therefore, we profiled serum samples from 20 HIV-negative individuals (from the CAPRISA 002 and 004 cohorts) followed yearly for 3 years (time points 0, 1, 2, and 3 years).



bNAb	Array Binding (≥$log_2$11)	Type of glycan	Other Arrays/ Glycan Data	Reference
2G12	Man_{7-9}	*N*-linked high mannose	Man_{7-9}	Scanlan et al., 2002
PG16	None		Man_5GlcNA_2, complex sialylated N-glycans	Amin et al., 2013; Shivatare et al., 2013
PGT121	NA2	*N*-linked glycan	N2, NA2, A2, A2F	Mouquet et al., 2012; CFG, 2014
PGT125	Man_{8-9}, Tg	*N*-linked high mannose, Thyroglobulin	Man_{8-9}	Walker et al., 2011; CFG, 2014
PGT126	Man_8, FABP	*N*-linked high mannose, Lipid binding protein	Man_{8-9}	Walker et al., 2011; CFG, 2014
PGT128	Man_{8-9}, Tg	*N*-linked high mannose, Thyroglobulin	Man_{8-9}	Walker et al., 2011; CFG, 2014
PGT130	Man_{8-9}, Tg, Gal-a, ManT	*N*-linked high mannose, Thyroglobulin, galactose-α, mannose.	Man_{8-9}	Walker et al., 2011; CFG, 2014

Figure 3.1: bNAb glycan-binding profiles

A) Heatmap of bNAb glycan-binding on the array. Shown are the log_2 binding values for each of the glycans on the array for the seven bNAbs tested. **B)** Table comparing glycan-binding profiles from this array to previous studies. Listed are the glycans that had RFU binding values ≥ 3000 (≥ $log_2$11). **C)** Glycan-binding profiles of all seven bNAbs for all glycans with binding ≥ 3000. Shown are the RFU binding values for each of the bNAbs tested for all eight glycans observed with binding ≥ 3000 (the cut-off used in previous studies) on this array.

Overall, IgG antibody profiles from the same individual at different time points were highly correlated, with Pearson correlation coefficients typically ≥0.9. An example of such is shown in Figure 3.2 where the glycan-binding profiles of CAP39 during year 0 and year 1 have an r^2 value of 0.90. Although profiles for the same patient at different time points were highly correlated we observed fluctuations across all time points at levels greater than those seen over 1- 3 months. For example, across all HIV-negative individuals and all time points, 2.6-fold changes were observed for about 10% of the glycans, 4-fold changes were observed for about 3% of the glycans and 10-fold changes were observed for less than 0.5% of the glycans.

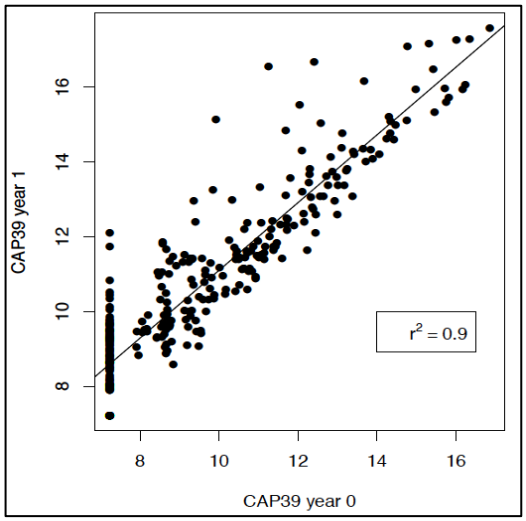


Figure 3.2: IgG antibody profiles within individuals

Shown are the $\log_2$ binding values (representing the RFU values) for CAP39 at year 0 (x-axis) and year 1 (y-axis).

Fluctuations were not consistent across individuals with some (CAP1 for example) having few glycan changes between time points and others (CAP39 and CAP54) having drastic changes in glycan-binding between time points (Figure 3.3). In addition some individuals (CAP11 and CAP17) had fewer than 10% of the glycans increasing or decreasing ≥2.6-fold, while others (CAP15 and CAP39 for example) had over 25% of the glycans either increasing or decreasing ≥2.6-fold between time points.

While substantial changes could be observed in certain individuals at certain time points, anti-glycan antibody signals oscillated throughout the three years, rather than continually increasing or decreasing. The results suggest a fluctuation about a steady state level for essentially all the antibody populations that we were measuring in the HIV-negative individuals. Thus we concluded that comparing a single time point between HIV-infected and pre-infection samples could give rise to a high proportion of false positives or perceived HIV-responses, especially if small numbers of HIV-infected individuals were tested.

3.4.3. Induction of IgG Anti-Glycan Antibodies During HIV infection

Based on the level of fluctuations observed within the HIV-negative samples over time we used two methods to identify responses to HIV-1 infection in the longitudinal HIV-positive samples. In the first method, an HIV-negative binding range was defined for each of the glycans on the array based on the levels of binding to these glycans for each of the HIV-negative samples. Glycans that were ≥2-fold higher in ≥ two individuals in the HIV-positive time points compared to the HIV-negative binding range were selected as having a potential HIV-response. The second method focused on 11 individuals with pre-infection and two or more consecutive post-infection time points. Glycans were selected if they were increased ≥2-fold in two consecutive HIV positive time points in ≥2 people compared to their matched pre-infection sample.

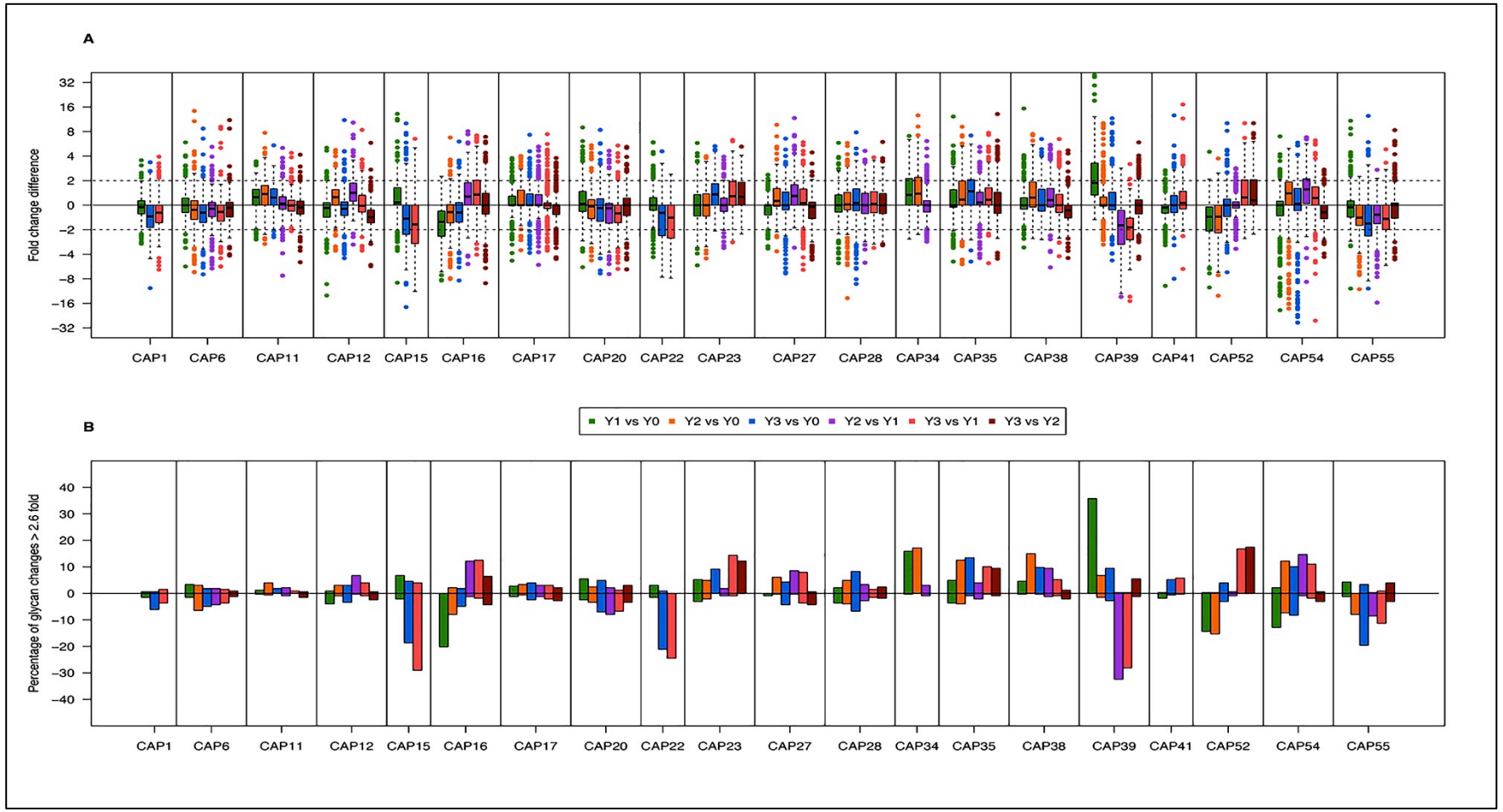


Figure 3.3: Fluctuations in anti-glycan IgG binding in HIV-negative longitudinal serum

A) Fold-change differences in longitudinal serum IgG antibodies. Shown are the differences between year 1 (Y1) and year 0 (Y0) in green, year 2 (Y2) and Y0 in orange, year 3 (Y3) and Y0 in blue, Y2 and Y1 in purple, Y3 and Y1 in pink and Y3 and Y2 in dark red. **B)** Percentage of glycans that had ≥2.6-fold changes for each comparison. Shown are the percentages of glycans with ≥2.6-fold change differences between any two time points.

Method 1: Glycans with Elevated Levels of Binding in HIV-Positive Compared to HIV-Negative Samples

All time points from the 27 HIV-positive individuals were compared to the level of binding in the HIV-negative samples (20 HIV-negative individuals over three years and 12 pre-infection samples). Fifteen glycans out of 327 (4.6%) were ≥2-fold higher in ≥ two HIV-positive individuals at any time point compared to the HIV-negative range. These 15 glycans included four *N*-linked glycans, two glycolipids, six Tn-peptides, one Lewis antigen, one glycoprotein and one non-human glycan, listed in Table 3.1. Of the four *N*-linked glycans three are high mannose glycans (Man_5, Man_8 and Man_9), all of which have been associated with bNAbs (Figure 3.1). Interestingly, the average fold-change for the *N*-linked mannose glycans increased with duration of infection, up to 14.9-fold at three years for Man5, as well as in the number of individuals with elevated binding during infection. Three glycans (GT3-Sp - 03, Ac-A-Tn(Thr)-S-G - 05 and Ac-S-Tn(Thr)-G-G - 03) had ≥2-fold increases throughout infection. The majority of these changes were seen in one or two individuals at one and three years post-infection.

Table 3.1: Glycans with increased binding in HIV-positive compared to HIV-negative samples

Glycan Type	Glycan Name	CAPRISA ID	One Year Post-Infection		Two Years Post-Infection		Three Years Post-Infection	
			n	Fold above HIV-negatives	n	Fold above HIV-negatives	n	Fold above HIV-negatives
***N*-linked**	Man_5	221, 295, 312			1	3.8	2	14.9
	Man_8	65, 137, 200, 206, 221, 244			3	2.5	4	6
	Man_9	8, 200, 221			1	3.3	3	5.1
	NA3	221, 310	1	2.3			1	4
Tn-peptides	Ac-A-Tn(Thr)-S-G - 05	206, 256, 332	1	2.8	1	2	2	2.2
	Ac-A-Tn(Thr)-S-G - 08	206, 256	1	3			1	2
	Ac-S-Tn(Thr)-Tn(Thr)-G - 05	8, 256	1	3			1	2.9
	Ac-S-Tn(Thr)-Tn(Thr)-G - 09	8, 283	1	2.3			1	2.3
	Ac-S-Tn(Thr)-G-G - 03	206, 256, 312, 332	1	2.1	2	2.2	2	2.4
	Ac-P-Tn(Thr)-T-G - 08	248, 256	2	2.1				
Glycolipids	GT2-Sp - 08	314, 322	1	2.8			1	2.1
	GT3-Sp - 03	88, 225, 314, 322	2	3	2	3.2	2	3.4
Glycoprotein	Hsp90 (heat shock protein)	8, 256	1	3.4			1	3.6
Lewis Antigen	LeX (monomeric)	229, 312	2	2.5				
Non-human	Fuc-b - 04	248, 314	2	2.6			1	2.2

Highlighted in blue are the glycans that were also bound by bNAbs on the array. "n" represents the number of individuals with elevated glycan-binding levels. Fold values represent the average fold above the HIV-negative samples for all individuals with ≥2-fold binding for each glycan.

Method 2: Glycans Elevated in HIV infection Compared to Matched Pre-Infection Samples

Eleven individuals (out of 27) had pre-infection samples and two consecutive post-infection time points and were used for this method. Based on these 11 individuals, 39 glycans appeared to increase with HIV-1 infection, and are listed in Table 3.2.

N-linked glycans account for 23% (9/39) of the selected glycans followed by Tn-peptides (21%, 8/39) and glycolipids and glycoproteins both accounting for 13% (5/39). Ten of the 39 glycans (26%), including Man_8, were also selected in Method 1 and/or bound by some of the bNAbs tested (Figure 3.1, highlighted in Table 3.2). Although many of these glycans were not selected in Method 1 they are related to those previously selected. For example, GT3-Sp - 07 has the glycan as GT3-Sp - 03 with a larger number of copies per molecule of BSA spotted onto the array (for higher density), and LacNac-Man_5 is related to Man_5. The r^2 values for each of these glycans ranged from 0.03 to 0.47, with NGA5B (*N*-linked glycan) having the lowest and Hep-5000 (Heparin polysaccharide) having the highest.

Table 3.2: List of glycans with elevated levels during HIV-1 infection compared to matched pre-infection samples

Type of Glycan	Number of Glycans	Name of Glycan
***N*-linked**	9	Man_1 - 04, Man_1 - 12, GlcNAc-Man_3, Man_3, LacNAc-Man_5, Man_8, NA4, NGA2B, NGA5B
Tn-peptides	8	Ac-A-Tn(Thr)-S-G - 05, Ac-P-Tn(Thr)-T-G - 05, Ac-S-Tn(Thr)-A-G - 04, Ac-S-Tn(Thr)-G-G - 03, Ac-S-Tn(Thr)-S-G - 24, Ac-S-Tn(Thr)-Tn(Thr)-G - 05, Ac-S-Tn(Thr)-Tn(Thr)-G - 09, Ac-Tn(Ser)Tn(Ser)Tn(Ser)-G - 03
glycolipids	5	GD1b, GQ2-Sp - 03, GT2-Sp - 08, GT3-Sp - 03, GT3-Sp - 07
glycoproteins	5	Alpha-1-acid glycoprotein, Fetuin (asialo), OSM (asialo), hsp90, Tg
TF-peptides	3	DTVPLPTAHGTSASSTG, DTVPLP-TF(Thr)-AHGTSASSTG, DTVPLPTAHGT-TF(Ser)-ASSTG
GAG (heparin)	2	Hep-5000 (Heparin), Hep-N-acetylated (Heparin-N-acetylated)
Lewis	2	6'-sulpho-LeA, Sialyl LeA
non-human	1	Fuc-b - 04
Sialylated carbohydrate	1	DSLNT
Type 2 carbohydrate	1	LNnH-11
GlcNAcb carbohydrate	1	Ac-S-Ser(GlcNAcB)-S-G-7
Glc carbohydrate	1	Maltopentaose

Shown in red are the glycans that were also selected using Method 1 (Table 3.1), in blue are the glycans that were also bound by bNAbs (Figure 3.1) and in pink if selected in Method 1 and bound by bNAbs.

Anti-Glycan Antibodies Capable of Binding gp120

Since some of the glycans that were elevated in Methods 1 and 2 might be unrelated to HIV-1, we used a gp120 competition assay to determine if antibodies that bound these glycans were HIV-1 specific. Sera from eight HIV-positive individuals (listed in Appendix, Supplementary Table 3.3), with high levels to Man_8, were incubated with the HIV glycoprotein gp120 prior to glycan array binding. PBS was used as a control and differences between PBS and gp120 incubated sera with p-values ≤ 0.05 were considered significant. Any reduction in glycan-binding after gp120 incubation was identified as HIV-specific.

We observed 48 glycans that showed a reduction in glycan-binding and three that had an increase in glycan-binding in the gp120-incubated sample compared to the PBS-incubated sample (listed in Table 3.3). Of the 48 glycans that had a reduction in glycan-binding following gp120 incubation, 16 (33%) were shown to be elevated in the HIV-positive individuals compared to the HIV-negative and/or during HIV infection compared to pre-infection (highlighted in red in Table 3.3). The majority of these were *N*-linked high mannose glycans and Tn-peptides (12/15, 80%). The p-values and fold-change differences observed for all 51 glycans (including gp120 as a control) can be found in Appendix, Supplementary Table 3.4.

gp120, used as a control on the array, showed a 21.4-fold reduction in median binding in the gp120 incubated sera compared to the PBS-incubated sera (Figure 3.4). Of the 51 glycans affected by the gp120 assay, 11 (22%) had a $\geq$2-fold difference between the PBS and gp120-incubated samples (Figure 3.4). These eleven glycans are made up of *N*-linked high mannose glycans ($Man_{7\text{-}9}$), Tn-peptides, a glycolipid, heparin, blood group antigens, a GlcNAc carbohydrate and a carcinoembryonic antigen (CEA). All glycans, except CEA, showed a reduction in binding in the gp120-incubated sample compared to

the PBS-incubated sample. The highest reduction in glycan-binding was seen in Man_8 (10.9-fold, p-value = 0.007).

Table 3.3: List of glycans significantly affected by the gp120 competition assay

Type	Number of Glycans	Name
N-linked	5	GlcNAc-Man3, LacNAc-Man5, Man7, Man8, Man9
Mannose	5	Man-a - 20, Mana1-6Man-a - 15, ManT, G2M4 - 07, Galb1-6Man-a - 13
Tn-peptide	15	Ac-A-Tn(Thr)-S-G - 05 & 08, Ac-P-Tn(Thr)-T-G - 05, Ac-S-Tn(Ser)-S-G - 04 & 22, Ac-S-Tn(Thr)-A-G - 04, Ac-S-Tn(Thr)-G-G - 03 & 07, Ac-S-Tn(Thr)-S-G - 04, Ac-Tn(Ser)Tn(Ser)Tn(Ser)-G - 03, Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 08, Ac-T-Tn(Thr)-P-G - 04 & 08, Ac-V-Tn(Thr)-S-G - 04 & 08
Glycolipid	3	GM2-Sp - 14, Gb4 - 09 & tetra
Glycoprotein	3	CEA, fetuin (asialo), hsp90
Heparin	1	Hep-5000
Blood Group	4	BG-A1, BG-H1, BG-H2, Globo H
Lewis Antigen	1	Sialyl LeX
Fucosylated	2	Fuc-a, Fuc-b
GalNAc carbohydrate	7	GalNAc-a - 04 & 22, GalNAca1-6Galb - 04, GlcNAca1-4Galb - 03 & 20, GalNAc-b - 21, GlcNAc-b - 21
Glc carbohydrate	1	Glc-b - 23
GlcNAc carbohydrate	1	3'GN-Di-LacNAc-Sp - 06
Type 2 carbohydrate	3	LacNAc - 22 & trimeric, LNnT - 04

Highlighted in red are the glycans that were selected in either Method 1, Method 2 and/or bound by bNAbs. Highlighted in green are the glycan with elevated binding after gp120 incubation, all other glycans showed a reduction in the gp120-incubated samples.

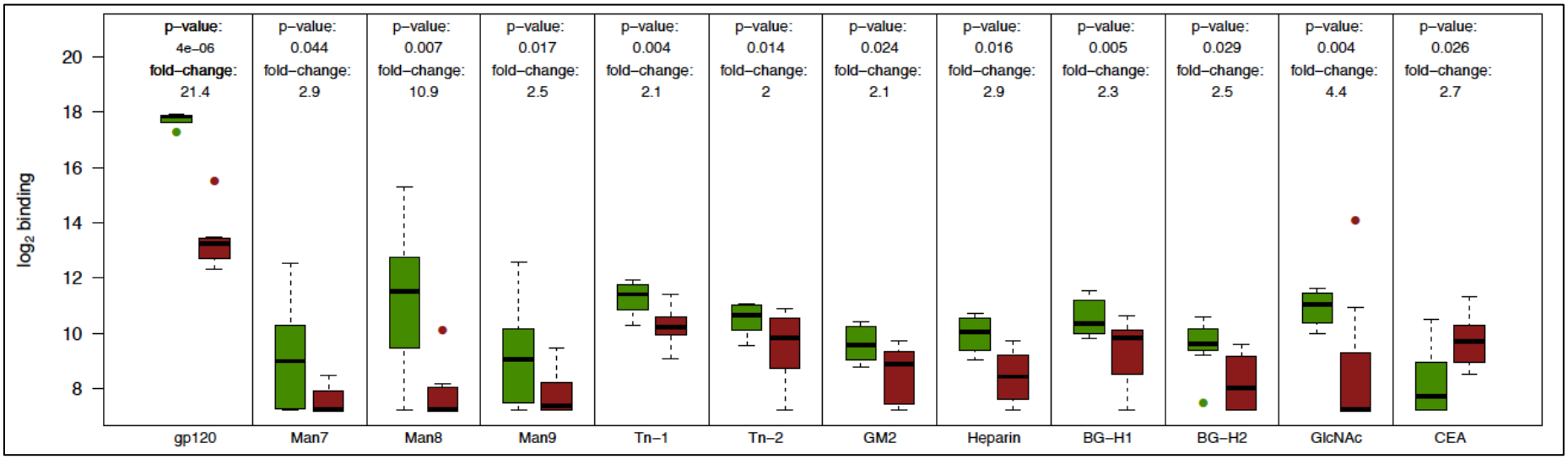


Figure 3.4: Difference in log_2 binding for PBS and gp120-incubated sera

Shown are log_2 binding values for each of the glycans that were significantly affected by the gp120 competition assay using the three-year post-infection time point for eight individuals. Shown in green are the PBS-incubated samples and the gp120-incubated samples in red. The p-values and fold change differences associated with each glycan are given above the plot. Tn-1 refers to Ac-S-Tn(Thr)-A-G - 04, Tn-2 refers to Ac-V-Tn(Thr)-S-G - 04, GM2 refers to GM2-Sp - 14, Heparin represents Hep-5000 and GlcNAc refers to GlcNAca1-4Galb.

Overall we identified several glycans that were increased in HIV-positive individuals compared to HIV-negative individuals and matched pre-infection samples, some of which were HIV specific (decreased when incubated with gp120 prior to glycan array binding) (listed in Table 3.4). Of these glycans *N*-linked mannose glycans and Heparin (Hep-5000) showed increased levels binding and number of individuals with the duration of infection (Figure 3.5). In contrast the Tn-peptides, glycolipids and fucosyl-b were elevated throughout HIV-1 infection (Figure 3.6).

Table 3.4: List of glycans that were elevated during HIV infection

Glycan Type	Glycan Name	bNAbs	Method 1	Method 2	gp120 assay
***N*-linked**	GlcNac-Man_3			✔	✔
	LacNac-Man_5			✔	✔
	Man_7	✔			✔
	Man_8	✔	✔	✔	✔
	Man_9	✔	✔		✔
	ManT	✔			✔
Tn-peptide	Ac-A-Tn(Thr)-S-G		✔	✔	✔
	Ac-P-Tn(Thr)-T-G			✔	✔
	Ac-S-Tn(Thr)-A-G			✔	✔
	Ac-S-Tn(Thr)-G-G		✔	✔	✔
	Ac-S-Tn(Thr)-S-G			✔	✔
	Ac-Tn(Ser)-Tn(Ser)-Tn(Ser)-G			✔	✔
	Ac-S-Tn(Thr)-Tn(Thr)-G		✔	✔	
Glycolipids	GT2		✔	✔	
	GT3		✔	✔	
Glycoprotein	hsp90 (heat shock protein)		✔	✔	✔
	Tg (thyroglobulin)	✔		✔	
Heparin (GAG)	Hep-5000 (Heparin)			✔	✔
Non-human	Fuc-b-04		✔	✔	✔

HIV-specificity was confirmed by a reduction in binding following gp120-incubation. Although ManT also showed reduced glycan-binding following gp120-incubation, the levels in the HIV-positive samples were not significantly elevated compared to the pre-infection samples (Appendix, Supplementary Figure 3.1A). We observed high levels of binding to some of the Tn-peptides in HIV-negative individuals (Appendix, Supplementary Figure 3.1 B-E). Likewise, one HIV-negative individual

(CAP39 at one year post-infection) also had elevated binding levels to Thyroglobulin (Tg), bound by some bNAbs (PGT125, PGT128 and PGT130, shown in Figure 3.1 and Appendix, Supplementary Figure 3.1G). Heat shock protein (hsp90) was higher in some HIV-positive samples at one year and three years but not at 2 years post-infection compared to the HIV-negative and pre-infection samples (Appendix, Supplementary Figure 3.1F).

3.4.4. BCN vs. non-BCN Glycan-Binding Profiles

Since the HIV envelope is covered with glycans and many well characterized bNAbs bind these glycans we aimed to determine if elevated levels of anti-glycan antibodies are a response to HIV-1 or unique to individuals who develop broadly neutralizing antibodies. Thus we compared the $\log_2$ binding values for each of the glycans on this array between the BCN and non-BCN individuals, across all time points (pre-infection, one, two and three years post-infection). Differences with p-values ≤ 0.05 after Bonferroni correction were considered significant.

We observed 17/327 glycans (5%) that were significantly lower in the BCN group compared to the non-BCN group for all time points (pre-infection and yearly for three years post-infection), (Table 3.5 and Figure 3.7). However, when comparing each time point separately there were no differences between the two groups. The majority of these glycans were sialylated carbohydrates and Lewis antigens. We observed no differences between the BCN and non-BCN individuals for the glycans that were elevated during HIV-1 infection (listed in Table 3.4).

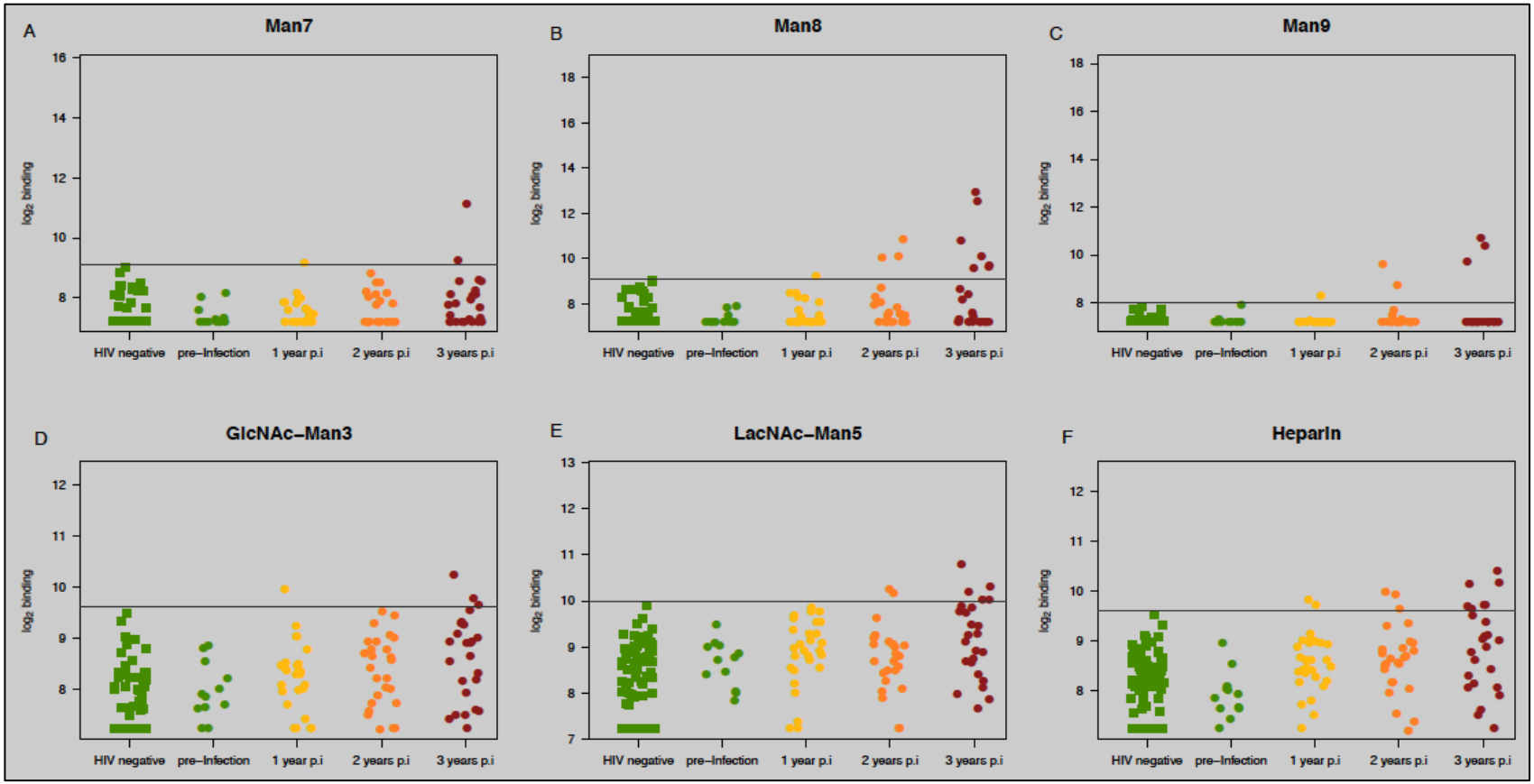


Figure 3.5: Longitudinal binding profiles of glycans that are elevated with the duration of HIV-1 infection

A) Man_7 longitudinal binding. Shown are the log_2 binding values for all serum samples. HIV-negative and pre-infection shown in green, HIV-positive samples at one year post-infection in yellow, HIV-positive samples at two years post-infection in orange and HIV-positive samples at three years post-infection in red. **B)** Man_8 longitudinal binding. **C)** Man_9 longitudinal binding. **D)** GlcNac-Man_3 longitudinal binding. **E)** LacNAc-Man_5 longitudinal binding. **F)** Heparin longitudinal binding.

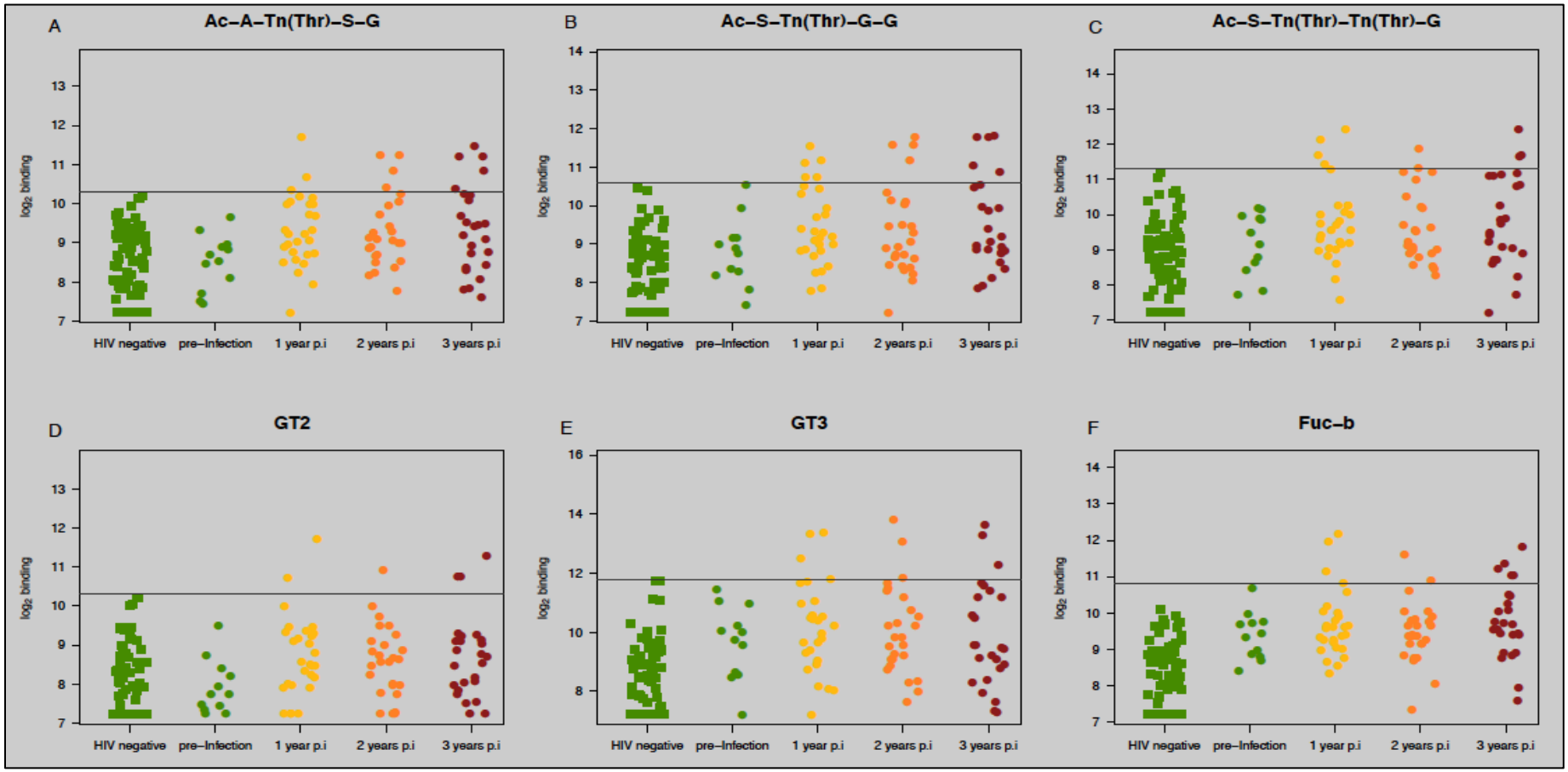


Figure 3.6: Longitudinal binding of glycans that were elevated throughout HIV-1 infection

A-C) Tn-peptide longitudinal binding. Shown are the log_2 binding values for all serum samples. HIV-negative and pre-infection shown in green, HIV-positive samples at one year post-infection in yellow, HIV-positive samples at two years post-infection in orange and HIV-positive samples at three years post-infection in red. **D-E)** Glycolipid longitudinal binding. **F)** Fucosyl-b longitudinal binding.

Table 3.5: List of glycans significantly lower in BCN vs. non-BCN individuals

Glycan Type	Glycan	Glycan Number in Figure 3.7
Sialylated carbohydrates	3'S(Gc)LeC−Sp − 05 & - 12	1 & 2
	3'SLeC−Sp − 05 & - 12	3 & 4
	9OAc3'SLeC−Sp − 05 & - 12	5 & 6
	3'−KDNLeC−Sp − 04 & - 12	7 & 8
Lewis antigens	Di−LeC−Sp − 06 & 16	9 & 10
	LeC−Sp − 06 & - 07 & - 15	11, 12 & 13
GlcNAc carbohydrate	3'GN type1−Sp − 04 & - 16	14 &15
Non-Human	Ara5	16
	Galb1−4Gal	17

Glycans shown are those that were lower in the BCN group compared to the non-BCN group with p-values (after Bonferroni correction) ≤0.05.

In summary elevated levels of glycan-binding antibodies were observed within HIV-infected individuals. These glycans were mostly N-linked high mannose glycans and Tn-peptides. Some glycans showed increased binding with the duration of infection, while others showed elevated levels of binding throughout infection. These elevated levels of glycan-binding were observed in both BCN and non-BCN individuals. However, BCN individuals showed lower levels of glycan-binding to some sialylated carbohydrates and Lewis antigens.

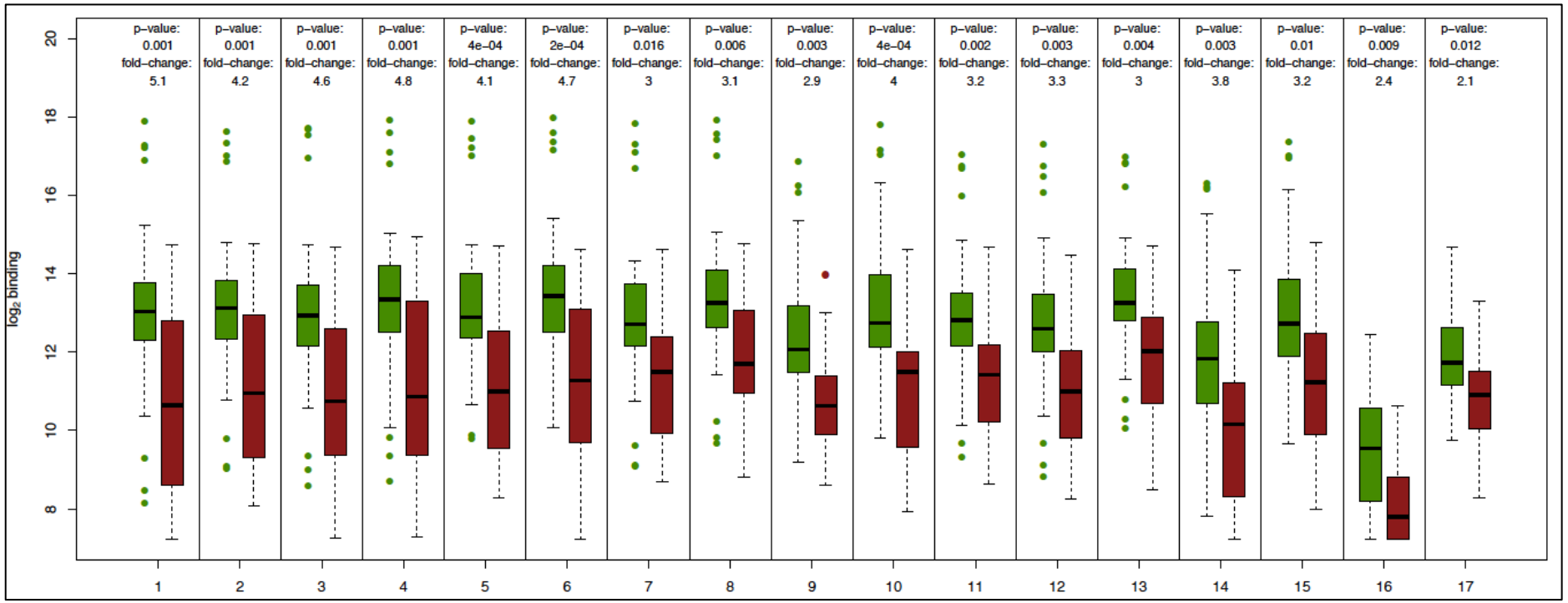


Figure 3.7: Glycans with significantly lower binding in the BCN individuals compared to the non-BCN individuals.

Shown are the $\log_2$ binding values for non-BCN individuals (green) for all serum samples (pre-infection and yearly post-infection for three years) and BCN individuals (red) for all serum samples (pre-infection and yearly post-infection for three years).

3.5. DISCUSSION

Using a glycan array we have identified glycan-binding antibodies that were elevated during HIV infection. These glycans were mostly *N*-linked mannose glycans, Tn-peptides and glycolipids. In particular we observed elevated levels of *N*-linked high mannose binding (Man_{7-9}) with the duration of infection with Man_8 having an average of 6-fold increases at 3 years post-infection. Tn-peptides and glycolipids were constantly elevated throughout infection compared to HIV-negative samples. We also observed lower levels of sialylated glycans and Lewis antigens in the BCN group compared to the non-BCN group, but no differences between the groups for the glycans that were elevated during HIV-1 infection. This suggests that elevated levels of glycan-binding antibodies were common in HIV-1 infection rather than specific to individuals who make broadly neutralizing antibodies.

We first used well-characterized bNAbs (2G12, PG16, PGT121, PGT125, PGT126, PGT128 and PGT130) to validate the array's ability to detect HIV-related anti-glycan responses. We observed the same glycan-binding profiles of these bNAbs on this array as those previously described (Scanlan*, et al.*, 2002; Walker*, et al.*, 2011; Mouquet*, et al.*, 2012; Amin*, et al.*, 2013; Shivatare*, et al.*, 2013; CFG, 2014). In addition we found high levels of binding to Tg by PGTs125, 128 and 130 and FABP by PGT126 that have not been previously described. We also found high levels of binding to a branched Man_3 (ManT) glycan, which is a fragment of many high mannose glycans.

This array has been used previously to assess glycan-binding antibodies following SIV-vaccination and SIV-infection, showing human and macaque anti-glycan antibody profiles to be similar (Campbell*, et al.*, 2013). In addition they observed few IgM and IgG

responses to mannose glycans (Manα1-6Manα and Man_6) but significantly elevated levels of IgM and IgG binding to Tn-peptides in the sera of macaques following vaccination and infection (Campbell*, et al.*, 2013). The lack of elevated binding to the high mannose *N*-linked glycans were likely the result of animals being followed for an insufficient length of time (just over a year) as high mannose responses in our study were typically seen later in infection with the greatest differences observed at 3 years post-infection. Tn-peptides are precursors of *O*-linked glycans and in normal mammalian cells the Tn-portion of the *O*-linked glycans are masked during the glycosylation pathway. However, in HIV infection the normal glycosylation pathways are altered revealing the Tn-portion of these *O*-linked glycans on the glycan shield (Campbell*, et al.*, 2013). In addition anti-Tn-peptide antibodies have been associated with the neutralization of HIV-1 by preventing fusion of the virus into infected cells (Hansen*, et al.*, 1991; Hansen*, et al.*, 1992). Other studies using these arrays have ruled out competition by lectins such as MBL (mannose binding lectin). All array experiments are done using 1:50 diluted sera and the concentration of MBL in the sera would be insufficient to bind the glycans on the array (unpublished data). In addition the binding of the lectins to the glycans on the array requires calcium and the assay is conducted in the absence of calcium (unpublished data). Thus it is unlikely that MBL would non-specifically interfere with this assay.

Many studies on glycan arrays using human sera have been limited to 1-3 months and since bNAbs against HIV-1 typically arise from 2 years post-infection it was important to understand natural fluctuations in glycan-binding over a similar time frame. By studying the levels of anti-glycan antibodies in HIV-negative individuals yearly for three years we observed fluctuations of IgG anti-glycan antibodies throughout the three years, at levels that were greater than those observed within 1-3 months (Oyelaran*, et al.*, 2009; Zhang*, et al.*, 2010). We also noted that the differences in glycan-binding tended to oscillate over

time, rather than a steady increase or decrease and differences were greater in some individuals. We note, however, that the HIV-negative individuals studied here are at a high risk of contracting infections (Van Loggerenberg*, et al.*, 2008; Abdool Karim*, et al.*, 2010) and thus may have higher levels of glycan-binding antibodies. Based on these findings we concluded that comparing a single post-infection time point to pre-infection might result in false positive associations or perceived HIV-responses and thus we developed two methods of identifying anti-glycan antibody responses to HIV.

In the first method we compared the level of glycan-binding in the HIV-negative and HIV-positive samples and observed 15 glycans that were ≥2-fold elevated in ≥2 individuals. These 15 glycans included Man_5, Man_8, Man_9, Tn-peptides and two glycolipids. Man_5, Man_8, Man_9 are found on the HIV envelope and have been associated with the epitopes of bNAbs against HIV (Scanlan*, et al.*, 2002; Walker*, et al.*, 2011; Mouquet*, et al.*, 2012; Amin*, et al.*, 2013; Shivatare*, et al.*, 2013; CFG, 2014). Tn-peptides, as previously mentioned, were also elevated in macaque sera following SIV-vaccination and infection and are able to prevent fusion of the virus into infected cells (Hansen*, et al.*, 1991; Hansen*, et al.*, 1992; Campbell*, et al.*, 2013). The glycolipids that were elevated in our study are ganglioside glycosphingolipids, which are found on the HIV-1 envelope and have been associated with the uptake of HIV-1 into mature dendritic cells (Hatch*, et al.*, 2009; Izquierdo-Useros*, et al.*, 2012; Puryear*, et al.*, 2012; Izquierdo-Useros*, et al.*, 2014). Ganglioside glycosphingolipids on the envelope of HIV-1 attach to the cellular receptor Siglec-1 found on mature dendritic cells enabling the incorporation of HIV-1 into the dendritic cell and the subsequent transfer to $CD4^+$ T-cells (Izquierdo-Useros*, et al.*, 2014).

Our second method compared the anti-glycan-binding levels of 11 individuals before and after HIV-1 infection. We found 39 glycans (compared to 15 glycans in the first method) that were elevated ≥2-fold in two consecutive HIV-infected samples and in ≥2

individuals. These 39 glycans included additional *N*-linked glycans, Tn-peptides and glycolipids as those seen in the first method. We also found elevated binding to Thyroglobulin that was also bound by bNAbs on the array. Thyroglobulin is a glycoprotein produced by the thyroid gland as part of thyroid hormone production (Hoffmann*, et al.*, 2007). Thyroid dysfunction has been widely associated with HIV-1 infection as well as elevated levels of anti-thyroglobulin autoantibodies in both infected children and adults as a result of the thyroid dysfunction (Fundarò*, et al.*, 1998; Hoffmann*, et al.*, 2007; Noureldeen*, et al.*, 2012; Ownby, 2012; De Carvalho*, et al.*, 2013).

Although eight glycans (18%, 8/45) were identified in both methods, we observed more anti-glycan responses in the second method compared to the first. Many of the elevated levels of anti-glycan antibodies observed in the second method are likely a consequence of HIV infection rather than antibodies binding HIV or may be unrelated to HIV since many of the glycans that were elevated in this method were seen at levels similar to the HIV-negative individuals. In addition, this method was limited to 11 individuals with pre-infection samples and likely missed some HIV-responses due to reduced sample size.

A gp120 competition assay was used confirm HIV-specific anti-glycan antibody responses. We observed decreases in binding following gp120-incubation compared to the PBS-incubated control sample for *N*-linked glycans (including Man_8), Tn-peptides and glycolipids. The greatest difference in binding between the PBS and gp120-incubated samples was observed for Man_8 with the gp120-incubated sample being 10.9-fold lower than the PBS-incubated sample. In addition we observed a 3.1-fold decrease in heparin following gp120 incubation. Heparin was also elevated in the HIV-positive samples compared to pre-infection and has been observed to inhibit viral infectivity including HIV-1 (Baba*, et al.*, 1988; Nagumo*, et al.*, 1988; Howell*, et al.*, 1996; Harrop*, et al.*, 1998).

Since monomeric gp120 was used for this assay we acknowledge that this would not detect antibodies that bind to epitopes in gp41 or those requiring the trimeric structure such as MPER, V1V2 and gp120-gp41 interface. A similar assay would need to be conducted using trimeric proteins.

Comparisons of anti-glycan antibody profiles between BCN and non-BCN individuals, highlighted 17 glycans that were lower in the BCN individuals compared to non-BCN individuals when all time points (including pre-infection) were grouped. These glycans were mostly sialylated carbohydrates and Lewis antigens. There were no differences between the groups when comparing each time point separately; suggesting that decrease in binding to these glycans was not a consequence of infection. There were no differences between the BCN and non-BCN groups for the glycans that were elevated during HIV-1 infection, suggesting that elevated levels of glycan-binding antibodies are a response to HIV-1 infection rather than specific to individuals who make broadly neutralizing antibodies.

The antibody specificities of eight BCN individuals in this study have previously been mapped to various parts of the HIV envelope glycoprotein (Moore*, et al.*, 2011; Moore*, et al.*, 2012; Moore*, et al.*, 2013; Wibmer*, et al.*, 2013). The plasma from six of these individuals bind glycans at either N332 on the V3 loop, N160 on the V2 loop or N276 on the CD4 binding site. Since antibody responses to the HIV envelope involve five different targets (CD4 binding site, MPER, C sheet of V1/V2, V3/C3 and the gp120/gp41 interface (Moore*, et al.*, 2015)), some of which do not involve glycans, it is likely that some of the BCN and non-BCN individuals used in this study target regions on the envelope that are not glycan related, and thus differences between the two groups would not be observed on the glycan array. In addition we used total serum IgG on the arrays,

thus binding and strain-specific antibody responses in these individuals might have diluted out the effect of the broadly neutralizing antibody responses.

Since some bNAbs target glycans on the glycan shield of gp120 and an effective HIV vaccine is likely to require bNAbs, many studies have focused on glycans as potential immunogens for vaccine design (reviewed in: Horiya*, et al.*, 2014). Most of these studies have focused on Man_9 binding of 2G12. Synthetic oligomannose constructs of Man_9 have been used to assess their affinity to 2G12 and whether they elicit an immunologic response as a potential immunogen for vaccine design (Wang, 2013; Horiya*, et al.*, 2014). These constructs include divalent cyclic peptides with Man_9, tetravalent Man_9 clusters, heteromultivalent clustering of high mannose glycans clustered with galactopyranoside scaffolds, cyclic peptide scaffolds, cholic acid scaffolds, PAMAM dendrons or bacteriophage Qβ scaffold (Wang, 2013; Horiya*, et al.*, 2014). Many of these constructs have shown affinity for 2G12 but mostly at levels lower than 2G12's affinity for gp120 (Wang, 2013; Horiya*, et al.*, 2014). Immunogenicity studies of these constructs have yielded anti-mannose antibodies but show either low or no cross reactivity with gp120 (Wang, 2013; Horiya*, et al.*, 2014). Other studies involving *Candida albicans* have been more successful at yielding high mannose binding antibodies (Doores*, et al.*, 2010a). Part of the failure of these synthetic glycans to elicit an appropriate cross-reactive response can be attributed to the scaffolds used, which affect the folding and orientation of the glycans and may impact antigenicity and immunogenicity (Wang, 2013; Horiya*, et al.*, 2014). Another explanation might be because high mannose *N*-linked glycan responses take longer to appear as seen in this study. Since responses to Tn-peptides and glycolipids were seen throughout infection, and these glycans are found on the HIV-1 envelope and are involved in the viral host cells entry (Hansen*, et al.*, 1991; Hansen*, et al.*, 1992; Izquierdo-Useros*, et al.*, 2014) they warrant further investigation.

3.6. CONCLUSIONS

We have identified an increase in IgG anti-glycan antibodies against high mannose *N*-linked glycans, Tn-peptides and glycolipids during HIV-1 infection. We noted that responses to high mannose *N*-linked glycans tend to occur during chronic infection with the highest levels of binding seen at 3 years post-infection, while elevated levels of binding to Tn-peptides and glycolipids occur constantly throughout HIV-1 infection. We found no differences in binding levels to these glycans between BCN and non-BCN individuals, suggesting that elevated levels of *N*-linked high mannose, Tn-peptides and glycolipids binding antibodies are a response to HIV-1 infection rather than specific to individuals who develop broadly neutralizing antibodies. *N*-linked high mannose glycans have been widely studied in HIV-vaccine immunogen design since they are bound by many bNAbs, but have showed little efficacy thus far. Tn-peptides and glycolipids, however, are also found on the HIV-1 envelope and although involved in HIV-infectivity they have not been studied in terms of immunogen design. Since responses to *N*-linked high mannose glycans are seen later in infection while responses to Tn-peptides and glycolipids were seen constantly throughout infection, further investigation into the role of these glycans during infection might be warranted.

3.7. REFERENCES

Abdool Karim, Q*, et al.* (2010) Effectiveness and safety of Tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science,* 329 (5996): 1168-1174

Amin, MN*, et al.* (2013) Synthetic glycopeptides reveal the glycan specificity of HIV-neutralizing antibodies. *Nature Chemical Biology,* 9 (8): 521-526

Astronomo, R*, et al.* (2010) Defining criteria for oligomannose immunogens for HIV using icosahedral virus capsid scaffolds. *Chemistry and Biology,* 17 (4): 357-370

Baba, M*, et al.* (1988) Mechanism of inhibitory effect of dextran sulfate and heparin on replication of human immunodeficiency virus in vitro. *Proceedings of the National Academy of Sciences,* 85: 6132-6136

Bonomelli, C*, et al.* (2011) The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One,* 6 (8): e23521

Campbell, C*, et al.* (2010) Construction and use of glycan microarrays. *Current Protocols in Chemical Biology,* 2: 37-53

Campbell, C*, et al.* (2013) High-throughput profiling of anti-glycan humoral responses to SIV vaccination and challenge. *PLoS One,* 8 (9): e75302

De Carvalho, LG*, et al.* (2013) Evaluation of thyroid function and autoimmunity in HIV-infected women. *Archives of Endocrinology and Metabolism,* 57 (6): 450-456

Doores, K*, et al.* (2010a) Antibody 2G12 recognizes di-mannose equivalently in domain- and nondomain-exchanged forms but only binds the HIV-1 glycan shield if domain exchanged. *Journal of Virology,* 84 (20): 10690-10699

Doores, KJ*, et al.* (2010) Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences,* 107 (31): 13800-13805

Dunlop, DC*, et al.* (2010) Polysaccharide mimicry of the epitope of the broadly neutralizing anti-HIV antibody, 2G12, induces enhanced antibody responses to self oligomannose glycans. *Glycobiology,* 20 (7): 812-823

Euler, Z and Schuitemaker, H (2012) Cross-reactive broadly neutralizing antibodies: timing is everything. *Frontiers in Immunology,* 3: 215

Fundarò, C*, et al.* (1998) Occurrence of anti-thyroid autoantibodies in children vertically infected with HIV-1. *Journal of Pediatric Endocrinology and Metabolism* 11 (6): 745-750

Go, EP*, et al.* (2013) Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *Journal of Proteome Research,* 12 (3): 1223-1234

Gray, ES*, et al.* (2011a) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology,* 85 (10): 4828-4840

Gray, ES*, et al.* (2011b) Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. *Journal of Virology,* 85 (15): 7719-7729

Hansen, JES*, et al.* (1991) Broadly neutralizing antibodies targeted to mucin-type carbohydrate epitopes of human immunodeficiency virus. *Journal of Virology,* 65 (12): 6461-6467

Hansen, JES*, et al.* (1992) An O-linked carbohydrate neutralization epitope of HIV-1 gp120 is expressed by HIV-1 env gene recombinant vaccinia virus. *Archives of Virology,* 126 (1-4): 11-20

Harrop, HA and Rider, CC (1998) Heparin and its derivatives bind to HIV-1 recombinant envelope glycoproteins, rather than to recombinant HIV-1 receptor, CD4. *Glycobiology,* 8 (2): 131-137

Hatch, SC*, et al.* (2009) Glycosphingolipid composition of human immunodeficiency virus type 1 (HIV-1) particles is a crucial determinant for dendritic cell-mediated HIV-1 trans-infection. *Journal of Virology,* 83 (8): 3496-3506

Hoffmann, CJ and Brown, TT (2007) Thyroid function abnormalities in HIV-infected patients. *Clinical Infectious Diseases,* 45 (4): 488-494

Horiya, S*, et al.* (2014) Recent strategies targeting HIV glycans in vaccine design. *Nature Chemical Biology,* 10 (12): 990-999

Howell, AL*, et al.* (1996) Inhibition of HIV-1 infectivity by low molecular weight heparin. *International Journal of Clinical and Laboratory Research,* 26: 124-131

Izquierdo-Useros, N*, et al.* (2012) Sialyllactose in viral membrane gangliosides is a novel molecular recognition pattern for mature dendritic cell capture of HIV-1. *PLoS Biology,* 10 (4): e1001315

Izquierdo-Useros, N*, et al.* (2014) HIV-1 capture and transmission by dendritic cells: the role of viral glycolipids and cellular receptor siglec-1. *PLoS Pathogens,* 10 (7): e1004146

Luallen, R*, et al.* (2008) An engineered Saccharomyces cerevisiae strain binds the broadly neutralizing human immunodeficiency virus type 1 antibody 2G12 and elicits mannose-specific gp120-binding antibodies. *Journal of virology,* 82 (13): 6447-6457

Manimala, J*, et al.* (2005) Carbohydrate array analysis of anti-Tn antibodies and lectins reveals unexpected specificities: Implications for diagnostic and vaccine development. *ChemBioChem,* 6 (12): 2229-2241

Manimala, JC*, et al.* (2006) High-throughput carbohydrate microarray analysis of 24 lectins. *Angewandte Chemie,* 45 (22): 3607-3610

Manimala, JC*, et al.* (2007) High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology,* 17 (8): 17C-23C

McLellan, J*, et al.* (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature,* 480 (7377): 336-343

Moir, S and Fauci, AS (2014) B-cell exhaustion in HIV infection: the role of immune activation. *Current Opinion in HIV and AIDS,* 9 (5): 472-477

Montefiori, DC (2009) Measuring HIV Neutralization in a Luciferase Reporter Gene Assay. *Methods in Molecular Biology,* 485: 395-405

Moore, PL*, et al.* (2011) Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *Journal of Virology,* 85 (7): 3128-3141

Moore, PL*, et al.* (2012) Evolution of an HIV glycan–dependent broadly neutralizing antibody epitope through immune escape. *Nature Medicine,* 18 (11): 1688-1692

Moore, PL*, et al.* (2013) Multiple pathways of escape from HIV broadly cross-neutralizing V2-dependent antibodies. *Journal of Virology,* 87 (9): 4882-4894

Moore, PL*, et al.* (2015) Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends in Microbiology,* 23 (4): 204-211

Morris, L*, et al.* (1998) HIV-1 antigen-specific and -nonspecific B cell responses are sensitive to combination antiretroviral therapy. *Journal of Experimental Medicine,* 188 (2): 233-245

Mouquet, H*, et al.* (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proceedings of the National Academy of Sciences,* 109 (47): E3268-E3277

Nagumo, T and Hoshino, H (1988) Heparin inhibits infectivity of human immunodeficiency virus in vitro. *Japanese Journal of Cancer Research,* 79: 9-11

Noureldeen, A*, et al.* (2012) Thyroid function in newly diagnosed HIV-infected patients. *Toxicology and Industrial Health,* 30 (10): 919-925

Ownby, RL (2012) Thyroid function and depression in HIV-1 infection. *World Journal of AIDS,* 02 (4): 279-285

Oyelaran, O*, et al.* (2009) Profiling human serum antibodies with a carbohydrate antigen microarray. *Journal of Proteome Research,* 8 (9): 4301-4310

Pancera, M, *et al.* (2013) Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nature Structural and Molecular Biology,* 20 (7): 804-813
Pancera, M, *et al.* (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature,* 514 (7523): 455-461
Puryear, WB, *et al.* (2012) HIV-1 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic cells. *Proceedings of the National Academy of Sciences,* 109 (19): 7475-7480
Scanlan, C, *et al.* (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of α 1→2 mannose residues on the outer face of gp120. *Journal of Virology,* 76 (14): 7306-7321
Shivatare, SS, *et al.* (2013) Efficient convergent synthesis of bi-, tri-, and tetra-antennary complex type N-glycans and their HIV-1 antigenicity. *Journal of the American Chemical Society,* 135 (41): 15382-15391
Tomaras, GD, *et al.* (2008) Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology,* 82 (24): 12449-12463
Trkola, A, *et al.* (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *Journal of Virology,* 70 (2): 1100-1108
Van Loggerenberg, F, *et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One,* 3 (4): e1954
Von Gunten, S, *et al.* (2009) Intravenous immunoglobulin contains a broad repertoire of anticarbohydrate antibodies that is not restricted to the IgG2 subclass. *Journal of Allergy and Clinical Immunology,* 123 (6): 1268-1276
Walker, LM, *et al.* (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477 (7365): 466-470
Wang, L, *et al.* (2014b) Cross-platform comparison of glycan microarray formats. *Glycobiology,* 24 (6): 507-517
Wang, L-X (2013) Synthetic carbohydrate antigens for HIV vaccine design. *Current Opinion in Chemical Biology,* 17 (6): 997-1005
Wibmer, CK, *et al.* (2013) Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. *PLoS Pathogens,* 9 (10): e1003738
Wibmer, CK, *et al.* (2015) HIV broadly neutralizing antibody targets. *Current Opinion in HIV and AIDS,* 10: 1-8
Zhang, Y, *et al.* (2010) Multidimensional glycan arrays for enhanced antibody profiling. *Molecular BioSystems,* 6 (9): 1583-1591
Zhu, X, *et al.* (2000) Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry,* 39 (37): 11194-11204

Online Databases

CFG - Consortium for Fuctional Glycomics [Online] Available: http://www.functionalglycomics.org/ [Accessed: December 2014]
IAVI's neutralizing antibody center [Online] Available: http://www.scripps.edu/research/nac/ [Accessed: February 2013]
NIH AIDS Reagent Program [Online] Available: www.aidsreagent.org [Accessed: February 2013]

GENERAL CONCLUSIONS

The best means of preventing new HIV-1 infections would be through a vaccine. Broadly neutralizing antibodies (bNAbs) have been a central focus in vaccine design because of their ability to neutralize numerous strains of HIV-1 and non-human primate studies using bNAbs have shown them to be protective (Sanders, *et al.*, 2002; Moldt, *et al.*, 2012; Barouch, *et al.*, 2013; Ko, *et al.*, 2014). Eliciting bNAbs by vaccination, however, has failed thus far. Studies have shown that approximately 15-30% of HIV-1 infected individuals develop bNAbs but after years of infection and that approximately 50% of HIV-infected individuals are capable of neutralizing ~50% of viruses (Stamatatos, *et al.*, 2009; Hraber, *et al.*, 2014). These antibodies tend to have unusual features such as high levels of somatic hypermutation probably accounting for the length of time required for them to develop (West, *et al.*, 2014) and/or long CDRH3s lengths present in rare populations of naive B cells (Briney, *et al.*, 2012; Doria-Rose, *et al.*, 2014). Understanding of the mechanisms by which bNAbs develop, however, remain limited, although we do know that in this particular cohort high viral loads and decreased CD4 T-cell counts contributed toward the development of bNAbs (Gray, *et al.*, 2011b). In addition the selection of B-cells following somatic hypermutation and class-switch recombination is mediated by T Follicular Helper (T_{FH}) cells (Heesters, *et al.*, 2014). The aim of this study was to understand host factors associated with broadly neutralizing antibodies against HIV-1 subtype C. It focused on three main aspects: 1) genetic variation in the heavy chain variable region genes (IGHV) responsible for encoding the majority of the antigen binding

part of antibodies, 2) evolution of a potent autologous anti-HIV IgA antibody during infection and 3) serum glycan-binding IgG antibodies during infection.

Previous studies have shown that CD4-binding site antibodies frequently use restricted germline IGHV genes such as IGHV1-2*02 and IGHV1-46*02 (Scheid*, et al.*, 2011; Kwong*, et al.*, 2012; West*, et al.*, 2012). This led to our hypothesis that individuals who develop bNAbs may have unique germline IGHV repertoires. However, after sequencing all seven IGHV subgroups, containing 48 functional IGHV genes and over 130 alleles, we observed no differences in the IGHV repertoire between individuals in the CAPRISA cohort who did and did not develop bNAbs during infection (Scheepers*, et al.*, 2015). This finding has positive implications for vaccine design, as it suggests that the development of bNAbs is not limited by the germline IGHV repertoire. In addition this study greatly expanded the number of germline IGHV sequences through the identification of 85 novel alleles and 38 alleles that have previously only been observed in rearranged antibody sequences. The germline IGHV gene repertoire is used as a basis from which germline gene usage is predicted from functional antibodies. These predictions are crucial when inferring intermediate sequences and the unmutated common ancestor (UCA) of antibody lineages. The inferred intermediates and UCA of lineages suggest a pathway of antibody evolution against a particular infection and is a current paradigm in HIV vaccine design. The deposition of sequences from this study into public databases (IgPdb, IMGT and CATNAP) will assist with more reliable germline assignments of isolated mAbs and better UCA predictions. Based on these data it is highly likely that other germline immunoglobulin genes including the heavy chain diversity (IGHD), heavy chain joining (IGHJ) and heavy chain constant region (Fc) genes will show a similar level of diversity. In addition immunoglobulin genes that make up the light chains, are also likely to contain

genetic variants that are not represented in the current germline immunoglobulin gene repertoire database (IMGT).

Since we revealed that the development of bNAbs is not reliant on germline IGHV repertoires, we next chose to investigate the evolutionary pathway of a potent strain-specific HIV-1 antibody in the hope of gaining insights into why some antibodies become broad and others not. Studies have shown that some bNAbs require long CDRH3 lengths and others high levels of somatic hypermutation (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014). Longitudinal studies have shown that strain-specific antibodies have the ability to develop breadth if the right viral epitopes are exposed through viral escape or they are able to drive the development of another broadly neutralizing antibody lineage (Moore*, et al.*, 2012; Wibmer*, et al.*, 2013; Gao*, et al.*, 2014). In this study we tracked the evolution of a strain-specific antibody (CAP88-CH06) that did not develop breadth despite persisting for years after viral escape. An IgA clade of this lineage with identical sequences to CAP88-CH06 was observed during early infection (5 weeks) and long after viral escape (2 years) and had very little evolution during infection. A transient clonally related IgG clade that developed after the IgA antibodies was also observed. We hypothesise that this antibody lineage was not able to develop breadth due to the limited evolution of the IgA antibodies and the disappearance of the IgG antibodies. Studies aimed at examining viral evolution in this donor (CAP88) over the same time period are underway. In particular we plan to determine if the lack of antibody evolution within the IgA clade of antibodies is due to limited evolution within the C3-V4 region (the target epitope of this mAb) and whether the disappearance of the IgG clade of antibodies was a result of complete viral escape. This is based on previous studies showing an accumulation of mutations occurring in the bNAb epitopes within the viral populations isolated from the respective donors prior to the development of the bNAbs (Liao*, et al.*, 2013; Doria-Rose*, et al.*, 2014). Understanding the

dynamics between the viral and antibody evolution in this donor may give greater insight as to why she did not develop neutralization breadth despite chronic infection.

Glycans on the HIV envelope are often targets of bNAbs, thus we aimed to understand the glycan-binding profile of serum IgG antibodies during HIV infection and whether glycan-binding was specific to individuals who develop bNAbs or just a consequence of infection. We found elevated levels of glycan-binding during HIV infection in individuals who did and did not develop broadly neutralizing antibodies. This data suggested that glycan-binding is a consequence of HIV-1 infection and that everyone is capable of making antibodies against glycans present on the HIV envelope. It is important to note that in this study total serum IgG rather than HIV-1 specific antibodies were used. Thus the abundant levels of binding antibodies in the serum could have overwhelmed the ability to identify bNAbs, many of which also target conformational epitopes that would not be detected on the arrays. In addition to detecting antibodies to high mannose N-linked glycans, which are known to be targets of bNAbs, we observed elevated levels of binding to Tn-peptides and glycolipids not previously associated with bNAbs. Isolating monoclonal antibodies against these glycans might provide insights into their role in HIV infection.

Overall this study has shown that bNAbs are not limited by individual germline IGHV repertoires, that strain-specific antibodies may require continued evolution to develop breadth and that glycan-binding antibodies are a consequence of HIV infection rather than specific to individuals who develop bNAbs. Such information contributes to a better understanding of how bNAbs develop which is hoped will help in the design of an HIV vaccine capable of inducing them.

FULL REFERENCE LIST

Abdool Karim, Q*, et al.* (2010) Effectiveness and safety of Tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science,* 329 (5996): 1168-1174

Altschul, SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25 (17): 3389-3402

Amin, MN*, et al.* (2013) Synthetic glycopeptides reveal the glycan specificity of HIV-neutralizing antibodies. *Nature Chemical Biology,* 9 (8): 521-526

An, P and Winkler, CA (2010) Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in Genetics,* 26 (3): 119-131

Arts, EJ and Hazuda, DJ (2012) HIV-1 antiretroviral drug therapy. *Cold Spring Harbor Perspectives in Medicine,* 2 (4): a007161

Astronomo, R*, et al.* (2008) A glycoconjugate antigen based on the recognition motif of a broadly neutralizing human immunodeficiency virus antibody, 2G12, is immunogenic but elicits antibodies unable to bind to the self glycans of gp120. *Journal of Virology,* 82 (13): 6359-6368

Astronomo, R*, et al.* (2010) Defining criteria for oligomannose immunogens for HIV using icosahedral virus capsid scaffolds. *Chemistry and Biology,* 17 (4): 357-370

Baba, M*, et al.* (1988) Mechanism of inhibitory effect of dextran sulfate and heparin on replication of human immunodeficiency virus in vitro. *Proceedings of the National Academy of Sciences,* 85: 6132-6136

Barouch, D*, et al.* (2013) Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature,* 503: 224-228

Barre-Sinoussi, F*, et al.* (2013) Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology,* 11 (12): 877-883

Bartesaghi, A*, et al.* (2013) Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nature Structural and Molecular Biology,* 20: 1352-1357

Bhiman, JN*, et al.* (2015) Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nature Medicine,* In Press

Binley, J*, et al.* (2000) The relationship between T cell proliferative responses and plasma viremia during treatment of Human Immunodeficiency Virus type 1 infection with combination antiretroviral therapy. *The Journal of Infectious Diseases,* 181: 1249-1263

Binley, J*, et al.* (2004) Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *Journal of Virology,* 78 (23): 13232-13252

Bonomelli, C*, et al.* (2011) The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One,* 6 (8): e23521

Boyd, SD*, et al.* (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel V-D-J pyrosequencing. *Science Translational Medicine,* 1 (12): 12ra23

Breden, F*, et al.* (2011) Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease. *PLoS One,* 6 (3): e16857

Brekke, OH and Sandlie, I (2003) Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nature Reviews Drug Discovery,* 2 (1): 52-62

Briney, BS, *et al.* (2012) Human peripheral blood antibodies with long HCDR3s are established primarily at original recombination using a limited subset of germline genes. *PloS One,* 7 (5): e36750+

Buchacher, A, *et al.* (1994) Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Research and Human Retroviruses,* 10 (4): 359-369

Buonaguro, L, *et al.* (2007) Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of Virology,* 81 (19): 10209-10219

Calarese, D, *et al.* (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science,* 300 (5628): 2065-2071

Cambier, JC, *et al.* (2007) B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nature Reviews Immunology,* 7 (8): 633-643

Campbell, C, *et al.* (2010) Construction and use of glycan microarrays. *Current Protocols in Chemical Biology,* 2: 37-53

Campbell, C, *et al.* (2013) High-throughput profiling of anti-glycan humoral responses to SIV vaccination and challenge. *PLoS One,* 8 (9): e75302

Casimiro, DR, *et al.* (2005) Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with DNA and recombinant adenoviral vaccine vectors expressing Gag. *Journal of Virology,* 79 (24): 15547-15555

Chun, T-W and Fauci, AC (1999) Latent reservoirs of HIV: obstacles to the eradication of virus. *Proceedings of the National Academy of Sciences,* 96: 10958-10961

Cohen, J (2013) Early treatment may have cured infant of HIV infection. *Science,* 339: 1134

Corti, D, *et al.* (2010) Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *Journal of Clinical Investigation,* 120 (5): 1663-1673

De Carvalho, LG, *et al.* (2013) Evaluation of thyroid function and autoimmunity in HIV-infected women. *Archives of Endocrinology and Metabolism,* 57 (6): 450-456

DeKosky, BJ, *et al.* (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology,* 31 (2): 166-169

Doores, K, *et al.* (2010a) Antibody 2G12 recognizes di-mannose equivalently in domain- and nondomain-exchanged forms but only binds the HIV-1 glycan shield if domain exchanged. *Journal of Virology,* 84 (20): 10690-10699

Doores, KJ, *et al.* (2010b) Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proceedings of the National Academy of Sciences,* 107 (31): 13800-13805

Doores, KJ, *et al.* (2015) Two classes of broadly neutralizing antibodies within a single lineage directed to the high-mannose patch of HIV envelope. *Journal of Virology,* 89 (2): 1105-1118

Doria-Rose, NA, *et al.* (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature,* 509 (7498): 55-62

Doria-Rose, NA, *et al.* (2015) A new member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. *Journal of Virology*: In Press

Dunlop, DC, *et al.* (2010) Polysaccharide mimicry of the epitope of the broadly neutralizing anti-HIV antibody, 2G12, induces enhanced antibody responses to self oligomannose glycans. *Glycobiology,* 20 (7): 812-823

Edgar, RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26 (19): 2460-2461

Engelman, A and Cherepanov, P (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology,* 10 (4): 279-290

Enriquez, M and McKinsey, DS (2011) Strategies to improve HIV treatment adherence in developed countries: clinical management at the individual level. *HIV/AIDS,* 3: 45-51

Euler, Z and Schuitemaker, H (2012) Cross-reactive broadly neutralizing antibodies: timing is everything. *Frontiers in Immunology,* 3: 215

Falkowska, E, *et al.* (2014) Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity,* 40 (5): 657-668
Fauci, AS, *et al.* (2013) HIV-AIDS: much accomplished, much to do. *Nature Immunology,* 14 (11): 1104-1107
Francica, JR, *et al.* (2015) Analysis of immunoglobulin transcripts and hypermutation following SHIVAD8 infection and protein-plus-adjuvant immunization. *Nature Communications,* 6: 6565
Frankel, AD and Young, JAT (1998) HIV-1: Fifteen proteins and an RNA. *Annual Reviews in Biochemistry,* 67: 1-25
Fundarò, C, *et al.* (1998) Occurrence of anti-thyroid autoantibodies in children vertically infected with HIV-1. *Journal of Pediatric Endocrinology and Metabolism* 11 (6): 745-750
Galtier, N, *et al.* (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computational Applied Biosciences,* 12 (6): 543-548
Gao, F, *et al.* (2014) Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell,* 158 (3): 481-491
Go, EP, *et al.* (2013) Characterization of host-cell line specific glycosylation profiles of early transmitted/founder HIV-1 gp120 envelope proteins. *Journal of Proteome Research,* 12 (3): 1223-1234
Gonzalez, SF, *et al.* (2011) Trafficking of B cell antigen in lymph nodes. *Annual Review of Immunology,* 29: 215-233
Gorny, MK, *et al.* (2009) Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Molecular Immunology,* 46 (5): 917-926
Gorny, MK, *et al.* (2011) Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PLoS One,* 6 (12): e27780
Gorny, MK, *et al.* (2012) Functional and immunochemical cross-reactivity of V2-specific monoclonal antibodies from HIV-1-infected individuals. *Virology,* 427 (2): 198-207
Gray, ES, *et al.* (2007) Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *Journal of Virology,* 81 (12): 6187-6196
Gray, ES, *et al.* (2011a) Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. *Journal of Virology,* 85 (15): 7719-7729
Gray, ES, *et al.* (2011b) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *Journal of Virology,* 85 (10): 4828-4840
Hansen, JES, *et al.* (1991) Broadly neutralizing antibodies targeted to mucin-type carbohydrate epitopes of human immunodeficiency virus. *Journal of Virology,* 65 (12): 6461-6467
Hansen, JES, *et al.* (1992) An O-linked carbohydrate neutralization epitope of HIV-1 gp120 is expressed by HIV-1 env gene recombinant vaccinia virus. *Archives of Virology,* 126 (1-4): 11-20
Harrop, HA and Rider, CC (1998) Heparin and its derivatives bind to HIV-1 recombinant envelope glycoproteins, rather than to recombinant HIV-1 receptor, CD4. *Glycobiology,* 8 (2): 131-137
Hatch, SC, *et al.* (2009) Glycosphingolipid composition of human immunodeficiency virus type 1 (HIV-1) particles is a crucial determinant for dendritic cell-mediated HIV-1 trans-infection. *Journal of Virology,* 83 (8): 3496-3506
Haynes, BF, *et al.* (2012a) Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine,* 366 (14): 1275-1286
Haynes, BF, *et al.* (2012b) B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnology,* 30 (5): 423-433
He, L, *et al.* (2014) Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Scientific Reports,* 4: 6778

Heesters, BA, *et al.* (2014) Follicular dendritic cells: dynamic antigen libraries. *Nature Reviews Immunology,* 14 (7): 495-504
Hirsch, VM, *et al.* (1994) Prolonged clinical latency and survival of macaques given a whole inactivated simian immunodefiniciency virus vaccine. *Journal of Infectious Diseases,* 170 (1): 51-59
Hoffmann, CJ and Brown, TT (2007) Thyroid function abnormalities in HIV-infected patients. *Clinical Infectious Diseases,* 45 (4): 488-494
Horiya, S, *et al.* (2014) Recent strategies targeting HIV glycans in vaccine design. *Nature Chemical Biology,* 10 (12): 990-999
Howell, AL, *et al.* (1996) Inhibition of HIV-1 infectivity by low molecular weight heparin. *International Journal of Clinical and Laboratory Research,* 26: 124-131
Hraber, P, *et al.* (2014) Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS,* 28 (2): 163-169
Huang, J, *et al.* (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature,* 491 (7424): 406-412
Huang, J, *et al.* (2014) Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature,* 515 (7525): 138-142
Huson, DH and Scornavacca, C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology,* 61 (6): 1061-1067
Izquierdo-Useros, N, *et al.* (2012) Sialyllactose in viral membrane gangliosides is a novel molecular recognition pattern for mature dendritic cell capture of HIV-1. *PLoS Biology,* 10 (4): e1001315
Izquierdo-Useros, N, *et al.* (2014) HIV-1 capture and transmission by dendritic cells: the role of viral glycolipids and cellular receptor siglec-1. *PLoS Pathogens,* 10 (7): e1004146
Jackson, K, *et al.* (2014) Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host and Microbe,* 16 (1): 105-114
Jardine, J, *et al.* (2013) Rational HIV immunogen design to target specific germline B cell receptors. *Science,* 340 (6133): 711-716
Julien, J-P, *et al.* (2013) Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science,* 342 (6165): 1477-1483
Kepler, Thomas B, *et al.* (2014) Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host & Microbe,* 16 (3): 304-313
King, C (2009) New insights into the differentiation and function of T follicular helper cells. *Nature Reviews Immunology,* 9 (11): 757-766
Kinoshita, K and Honjo, T (2001) Linking class-switch recombination with somatic hypermutation. *Nature Reviews Immunology,* 2: 493-503
Klein, F, *et al.* (2013) Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell,* 153 (1): 126-138
Klein, U and Dalla-Favera, R (2008) Germinal centres: role in B-cell physiology and malignancy. *Nature Reviews Immunology,* 8 (1): 22-33
Ko, S-Y, *et al.* (2014) Enhanced neonatal Fc receptor function improves protection against primate SHIV infection. *Nature,* 514 (7524): 642-645
Kong, L, *et al.* (2013) Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nature Structural and Molecular Biology,* 20 (7): 796-803
Kwong, PD and Mascola, JR (2012) Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity,* 37 (3): 412-425
Lavinder, JJ, *et al.* (2014) Systematic characterization and comparative analysis of the rabbit immunoglobulin repertoire. *PLoS One,* 9 (6): e101322
Lefranc, M-P, *et al.* (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research,* 37 (Database Issue): D1006-D1012
Liao, H-X, *et al.* (2011) Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *Journal of Experimental Medicine,* 208 (11): 2237-2249

Liao, H-X, *et al.* (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature,* 496 (7446): 469-476
Lingwood, D, *et al.* (2012) Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature,* 489 (7417): 566-570
Liu, R, *et al.* (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell,* 86: 367-377
Luallen, R, *et al.* (2008) An engineered Saccharomyces cerevisiae strain binds the broadly neutralizing human immunodeficiency virus type 1 antibody 2G12 and elicits mannose-specific gp120-binding antibodies. *Journal of Virology,* 82 (13): 6447-6457
Lyumkis, D, *et al.* (2013) Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science,* 342 (6165): 1484-1490
Manimala, J, *et al.* (2005) Carbohydrate array analysis of anti-Tn antibodies and lectins reveals unexpected specificities: Implications for diagnostic and vaccine development. *ChemBioChem,* 6 (12): 2229-2241
Manimala, JC, *et al.* (2006) High-throughput carbohydrate microarray analysis of 24 lectins. *Angewandte Chemie,* 45 (22): 3607-3610
Manimala, JC, *et al.* (2007) High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology,* 17 (8): 17C-23C
Market, E and Papavasiliou, FN (2003) V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biology,* 1 (1): E16
McGuire, A, *et al.* (2014) Diverse recombinant HIV-1 envs fail to activate B cells expressing the germline B cell receptors of the broadly neutralizing anti-HIV-1 antibodies PG9 and 447-52D. *Journal of Virology,* 88 (5): 2645-2657
McLellan, J, *et al.* (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature,* 480 (7377): 336-343
Mlisana, K, *et al.* (2014) Rapid disease progression in HIV-1 subtype C-infected South African women. *Clinical Infectious Diseases,* 59 (9): 1322-1331
Moir, S and Fauci, AS (2009) B cells in HIV infection and disease. *Nature Reviews Immunology,* 9 (4): 235-245
Moir, S and Fauci, AS (2014) B-cell exhaustion in HIV infection: the role of immune activation. *Current Opinion in HIV and AIDS,* 9 (5): 472-477
Moldt, B, *et al.* (2012) Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. *Proceedings of the National Academy of Sciences,* 109 (46): 18921-18925
Montefiori, DC (2009) Measuring HIV Neutralization in a Luciferase Reporter Gene Assay. *Methods in Molecular Biology,* 485: 395-405
Moore, PL, *et al.* (2008) The C3-V4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *Journal of Virology,* 82 (4): 1860-1869
Moore, PL, *et al.* (2009a) Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathogens,* 5 (9): e1000598
Moore, PL, *et al.* (2009b) Specificity of the autologous neutralizing antibody response. *Current Opinion in HIV and AIDS,* 4 (5): 358-363
Moore, PL, *et al.* (2011) Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *Journal of Virology,* 85 (7): 3128-3141
Moore, PL, *et al.* (2012) Evolution of an HIV glycan–dependent broadly neutralizing antibody epitope through immune escape. *Nature Medicine,* 18 (11): 1688-1692
Moore, PL, *et al.* (2013) Multiple pathways of escape from HIV broadly cross-neutralizing V2-dependent antibodies. *Journal of Virology,* 87 (9): 4882-4894
Moore, PL, *et al.* (2015) Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends in Microbiology,* 23 (4): 204-211
Morris, L, *et al.* (1998) HIV-1 antigen-specific and -nonspecific B cell responses are sensitive to combination antiretroviral therapy. *Journal of Experimental Medicine,* 188 (2): 233-245

Mouquet, H, *et al.* (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proceedings of the National Academy of Sciences,* 109 (47): E3268-E3277
Nagumo, T and Hoshino, H (1988) Heparin inhibits infectivity of human immunodeficiency virus in vitro. *Japanese Journal of Cancer Research,* 79: 9-11
Noureldeen, A, *et al.* (2012) Thyroid function in newly diagnosed HIV-infected patients. *Toxicology and Industrial Health,* 30 (10): 919-925
Ownby, RL (2012) Thyroid function and depression in HIV-1 infection. *World Journal of AIDS,* 02 (4): 279-285
Oyelaran, O, *et al.* (2009) Profiling human serum antibodies with a carbohydrate antigen microarray. *Journal of Proteome Research,* 8 (9): 4301-4310
Pancera, M, *et al.* (2013) Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nature Structural and Molecular Biology,* 20 (7): 804-813
Pancera, M, *et al.* (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature,* 514 (7523): 455-461
Parameswaran, P, *et al.* (2014) Convergent antibody signatures in human dengue. *Cell Host and Microbe,* 13 (6): 691-700
Pieper, K, *et al.* (2013) B-cell biology and development. *Journal of Allergy and Clinical Immunology,* 131 (4): 959-971
Pone, EJ, *et al.* (2010) Toll-like receptors and B-cell receptors synergize to induce immunoglobulin class-switch DNA recombination: relevance to microbial antibody responses. *Critical Reviews in Immunology,* 30 (1): 1-29
Puryear, WB, *et al.* (2012) HIV-1 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic cells. *Proceedings of the National Academy of Sciences,* 109 (19): 7475-7480
Rerks-Ngarm, S, *et al.* (2009) Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine,* 361 (23): 2209-2220
Robinson, WH (2014) Sequencing the functional antibody repertoire-diagnostic and therapeutic discovery. *Nature Reviews Rheumatology,* 11: 171-182
Samson, M, *et al.* (1996) Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature,* 382 (6593): 722-725
Sanders, RW, *et al.* (2002) The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *Journal of Virology,* 76 (14): 7293-7305
Saunders, KO, *et al.* (2015) Sustained delivery of a broadly neutralizing antibody in nonhuman primates confers long-term protection against simian/human immunodeficiency virus infection. *Journal of Virology,* 89 (11): 5895-5903
Scanlan, C, *et al.* (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of α 1→2 mannose residues on the outer face of gp120. *Journal of Virology,* 76 (14): 7306-7321
Scharf, L, *et al.* (2014) Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports,* 7 (3): 785-795
Scheepers, C, *et al.* (2015) Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. *The Journal of Immunology,* 194 (9): 4371-4378
Scheid, JF, *et al.* (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science,* 333 (6049): 1633-1637
Schirmer, M, *et al.* (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research,* 43 (6): e37
Schroeder, HW, Jr. and Cavacini, L (2010) Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology,* 125 (2 Suppl 2): S41-S52
Sharp, PM and Hahn, BH (2011) Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine,* 1 (1): a006841
Shivatare, SS, *et al.* (2013) Efficient convergent synthesis of bi-, tri-, and tetra-antennary complex type N-glycans and their HIV-1 antigenicity. *Journal of the American Chemical Society,* 135 (41): 15382-15391

Shiver, JW, *et al.* (2002) Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature,* 415: 331-335
Shrestha, R, *et al.* (2014) QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BioMedCentral Bioinformatics,* 15 (1): 33
Sievers, F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7: 539
Simek, MD, *et al.* (2009) Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *Journal of Virology,* 83 (14): 7337-7348
Smith, K, *et al.* (2013) Fully human monoclonal antibodies from antibody secreting cells after vaccination with Pneumovax®23 are serotype specific and facilitate opsonophagocytosis. *Immunobiology,* 218 (5): 745-754
Stamatatos, L, *et al.* (2009) Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature Medicine,* 15 (8): 866-870
Stone, KD, *et al.* (2010) IgE, mast cells, basophils, and eosinophils. *Journal of Allergy and Clinical Immunology,* 125 (2 Suppl 2): S73-S80
Tangye, SG, *et al.* (2002) Isotype switching by human B cells is division-associated and regulated by cytokines. *The Journal of Immunology,* 169 (8): 4298-4306
Taylor, BS, *et al.* (2008) The challenge of HIV-1 subtype diversity. *New England Journal of Medicine,* 358 (15): 1590-1602
Tobon, GJ, *et al.* (2013) B lymphocytes: development, tolerance, and their role in autoimmunity-focus on systemic lupus erythematosus. *Autoimmune Diseases,* 2013: 827254
Tomaras, GD, *et al.* (2008) Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology,* 82 (24): 12449-12463
Trkola, A, *et al.* (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *Journal of Virology,* 70 (2): 1100-1108
Van Gent, D, *et al.* (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews Genetics,* 2: 196-206
Van Loggerenberg, F, *et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One,* 3 (4): e1954
Vladutiu, AO (2000) Immunoglobulin D: properties, measurement and clinical relevance *Clinical and Diagnostic Laboratory Immunology,* 7 (2): 131-140
Von Gunten, S, *et al.* (2009) Intravenous immunoglobulin contains a broad repertoire of anticarbohydrate antibodies that is not restricted to the IgG2 subclass. *Journal of Allergy and Clinical Immunology,* 123 (6): 1268-1276
Walker, LM, *et al.* (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science,* 326 (5950): 285-289
Walker, LM, *et al.* (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature,* 477 (7365): 466-470
Wang, C, *et al.* (2014a) Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *Journal of Immunology* 192 (2): 603-611
Wang, HB, *et al.* (2015a) HIV vaccine research: the challenge and the way forward. *Journal of Immunology Research,* 2015: 503978
Wang, L, *et al.* (2014b) Cross-platform comparison of glycan microarray formats. *Glycobiology,* 24 (6): 507-517
Wang, L-X (2013) Synthetic carbohydrate antigens for HIV vaccine design. *Current Opinion in Chemical Biology,* 17 (6): 997-1005
Wang, S, *et al.* (2015b) Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies. *Cell,* 160 (4): 785-797

Wang, Y, *et al.* (2011) Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics,* 63 (5): 259-265
Wang, Y, *et al.* (2014c) IgE-associated IGHV genes from venom and peanut allergic individuals lack mutational evidence of antigen selection. *PLoS One,* 9 (2): e89730
Watson, CT and Breden, F (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes and Immunity,* 13 (5): 363-373
Watson, CT, *et al.* (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *American Journal of Human Genetics,* 92 (4): 530-546
West, AP, *et al.* (2012) Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proceedings of the National Academy of Sciences,* 109 (30): E2083–E2090
West, AP, *et al.* (2014) Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell,* 156 (4): 633-648
Wibmer, CK, *et al.* (2013) Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. *PLoS Pathogens,* 9 (10): e1003738
Wibmer, CK, *et al.* (2015) HIV broadly neutralizing antibody targets. *Current Opinion in HIV and AIDS,* 10: 1-8
Woof, JM and Mestecky, J (2005) Mucosal immunoglobulins. *Immunological Reviews,* 206: 64-82
Wrammert, J, *et al.* (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature,* 453 (7195): 667-671
Wright, I and Travers, S (2014) RAMICS: trainable, high-speed and biologically relevant alignment of high-throughput sequencing reads to coding DNA. *Nucleic Acids Research,* 42 (13): e106
Wu, TT, *et al.* (1993) Length distribution of CDRH3 in antibodies. *Proteins: Structure, Function and Genetics,* 16: 1-7
Wu, X, *et al.* (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science,* 329 (5993): 856-861
Wu, X, *et al.* (2015) Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell,* 161: 1-16
Wyatt, R and Sodroski, J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science,* 280: 1884-1888
Xiao, X, *et al.* (2009) Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochemical and Biophysical Research Communications,* 390 (3): 404-409
Zhang, J, *et al.* (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics,* 30 (5): 614-620
Zhang, Y, *et al.* (2010) Multidimensional glycan arrays for enhanced antibody profiling. *Molecular BioSystems,* 6 (9): 1583-1591
Zhu, J (2012) Somatic populations of PGT135–137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatic. *Frontiers in Microbiology,* 3: 315
Zhu, J, *et al.* (2013a) De novo identification of VRC01 class HIV-1–neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proceedings of the National Academy of Sciences,* 110 (43): E4088-E4097
Zhu, J, *et al.* (2013b) Mining the antibodyome for HIV-1–neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proceedings of the National Academy of Sciences,* 110 (16): 6470-6475
Zhu, X, *et al.* (2000) Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry,* 39 (37): 11194-11204
Zolla-Pazner, S (2014) A critical question for HIV vaccine development: which antibodies to induce? *Science,* 345 (6193): 167-168

Zwick, MB*, et al.* (2001) Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *Journal of Virology,* 75 (22): 10892-10905

Books

Janeway, C*, et al.* (2001) Immunobiology: the immune system in health and disease. New York: Garland Science. ISBN: 081533642
Lefranc, M-P (2001) The immunoglobulin factsbook. Academic Press. ISBN: 012441351
Parham, P (2005) The immune system. Garland Science. ISBN: 0815340931
Parham, P (2009) The immune system. Garland Science. ISBN: 9780815341468

Online Databases

NCBI's BLAST Alignment Tool [Online] Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi [Accessed: January 2012]
UCSC Genome Bioinformatics, BLAT [Online] Available: https://genome.ucsc.edu/cgi-bin/hgBlat?command=start [Accessed: January 2012]
CATNAP, HIV sequence database [Online] Available: http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp [Accessed: June 2014]
CFG - Consortium for Fuctional Glycomics [Online] Available: http://www.functionalglycomics.org/ [Accessed: December 2014]
NCBI's dbSNP short genetic variations tool. [Online] Available: http://www.ncbi.nlm.nih.gov/SNP/ [Accessed: November 2013]
ENSEMBL [Online] Available: www.ensembl.org [Accessed: November 2013]
NCBI's Genetic Sequence Database, GenBank [Online] Available: http://www.ncbi.nlm.nih.gov/genbank/ [Accessed: November 2013]
IAVI's neutralizing antibody center [Online] Available: http://www.scripps.edu/research/nac/ [Accessed: February 2013]
NCBI's IgBLAST tool [Online] Available: http://www.ncbi.nlm.nih.gov/igblast/ [Accessed: June 2014]
IgPdb: Immunoglobulin Polymorphism Database [Online] Available: http://cgi.cse.unsw.edu.au/~ihmmune/IgPdb/display_ref.php?ref=2 [Accessed: June 2014]
IMGT - The international immunogenetics information system [Online] Available: http://www.imgt.org/ [Accessed: June 2014]
Medscape: HIV-1 Clades [Online] Available: http://img.medscape.com/article/708/915/708915-fig1.jpg
NIH AIDS Reagent Program [Online] Available: www.aidsreagent.org [Accessed: February 2013]
UNAIDS global report on the global AIDS epidemic 2013 [Online] Available: http://www.unaids.org/
World Health Organization (WHO) HIV/AIDS factsheet [Online] Available: http://www.who.int/mediacentre/factsheets/fs360/en/

APPENDIX

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Miss Catherine Mitchell

CLEARANCE CERTIFICATE	**M111104**
PROJECT	Host Factors and Broadly Neutralizing Antibodies in South Africa
INVESTIGATORS	Miss Catherine Mitchell.
DEPARTMENT	School of Pathology/Human Genetics
DATE CONSIDERED	25/11/2011
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 25/11/2011 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable
cc: Supervisor : Prof Lynn Morris

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

First page of the Turnitin Originality Report

Turnitin Originality Report

ScheepersThesis22June2015.pdf by Cathrine Scheepers

From Research 1 (1Uz7XMbJSkrX4UYNV9nCKsnzXduuEhxikCnm5LuM808xeVF7BbMs5duCWMStkMoxfwMDPinDlJbxMHo8

- Processed on 23-Jun-2015 1:24 PM SAST
- ID: 552257179
- Word Count: 62084

Similarity Index
20%
Similarity by Source

Internet Sources:
18%
Publications:
13%
Student Papers:
9%

sources:

1 2% match (Internet from 23-Jun-2015)
http://www.jimmunol.org/content/194/9/4371

2 1% match (publications)
Sok, Devin, Marit J. van Gils, Matthias Pauthner, Jean-Philippe Julien, Karen L. Saye-Francisco, Jessica Hsueh, Bryan Briney, Jeong Hyun Lee, Khoa M. Le, Peter S. Lee, Yuanzi Hua, Michael S. Seaman, John P. Moore, Andrew B. Ward, Ian A. Wilson, Rogier W. Sanders, and Dennis R. Burton. "Recombinant HIV envelope trimer selects for quaternary-dependent antibodies targeting the trimer apex", Proceedings of the National Academy of Sciences, 2014.

3 < 1% match (publications)
Mouquet, H., L. Scharf, Z. Euler, Y. Liu, C. Eden, J. F. Scheid, A. Halper-Stromberg, P. N. P. Gnanapragasam, D. I. R. Spencer, M. S. Seaman, H. Schuitemaker, T. Feizi, M. C. Nussenzweig, and P. J. Bjorkman. "PNAS Plus: Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies", Proceedings of the National Academy of Sciences, 2012.

4 < 1% match (student papers from 06-Jul-2012)
Submitted to Indian Institute of Science, Bangalore on 2012-07-06

5 < 1% match (Internet from 16-Apr-2015)
http://edoc.rki.de/documents/dissertationen/strasz-nicola-2013-06-01/PDF/strasz.pdf

6 < 1% match (Internet from 09-Jun-2014)
http://dhvi.duke.edu/faculty/details/0114780

7 < 1% match (publications)
Zhu, J., X. Wu, B. Zhang, K. McKee, S. O'Dell, C. Soto, T. Zhou, J. P. Casazza, J. C. Mullikin, P. D. Kwong, J. R. Mascola, L. Shapiro, J. Becker, B. Benjamin, R. Blakesley, G. Bouffard, S. Brooks, H. Coleman, M. Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove, S.-l. Ho, T. Johnson, R. Legaspi, S. Lovett, Q. Maduro, C. Masiello, B. Maskeri, J. McDowell, C. Montemayor, J. Mullikin, M. Park, N. Riebow, K. Schandler, B. Schmidt, C. Sison, M. Stantripop, J. Thomas, P. Thomas, M. Vemulapalli, and A. Young. "De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts", Proceedings of the National Academy of Sciences, 2013.

Supplementary Information for Section B, Chapter 1:

Supplementary Table 1.1: Viral loads and neutralization breadth from all CAPRISA participants

Shown are all of the BCN individuals, all non-BCN individuals and the two intermediate individuals. Viral loads were determined at 6 months post-infection. The breadth shown is the peak percentage of breadth obtained for each individual during infection. For the BCN individuals the time at peak breadth is the earliest time point that peak breadth was obtained if the same level of breadth remained over time, whereas the time at peak breadth for the non-BCN individuals is the latest time point at which peak breadth was obtained if the same level of breadth was maintained over-time. Samples with an * are on antiretroviral treatment (ART) or were lost to follow up following peak breadth. The "+" highlights participants that were only sequenced using 454 technology.

BCN				Non-BCN							
PTID	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**	**PTID**	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**	**PTID**	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**
CAP008*	3	98 400	39%	CAP65*	5	47 300	0%	CAP244	4	47 300	22%
CAP177*	3	32 300	56%	CAP88*+	5	21 100	0%	CAP332*	3	21 100	22%
CAP206*	3	156 000	44%	CAP137*	3	1 500	0%	**Median**	3,5	34200	22%
CAP248*	5	1 290	72%	CAP200*	3	349 000	6%				
CAP255*+	3	59 500	61%	CAP221*	3	3 830	0%				
CAP256*+	3	750 000	83%	CAP225*	5	59 700	6%				
CAP257*+	4	16 900	94%	CAP229	5	8 470	6%				
CAP287*	4	39 500	50%	CAP268*	4	15 100	0%				
CAP292	5	134 000	83%	CAP283*	2	66 300	0%				
CAP310*	2	171 000	33%	CAP289*	3	62 300	6%				
CAP312	3	40 400	56%	CAP290*	2	72 200	6%				
CAP314*	2	215 000	44%	CAP295	4	19 100	11%				
CAP322*	2	92 200	39%	CAP305*	4	67 200	0%				
Median	**3**	**92 200**	**56%**	**Median**	**4**	**47 300**	**0%**				

Supplementary Table 1.2: Germline IGHV primers for 454 sequencing

Four different MID (MID1-4) sequences were used to allow pooling of four samples per run. Two reverse primers (3a and 3b) were used for IGHV3 to ensure all IGHV3 genes would be amplified.

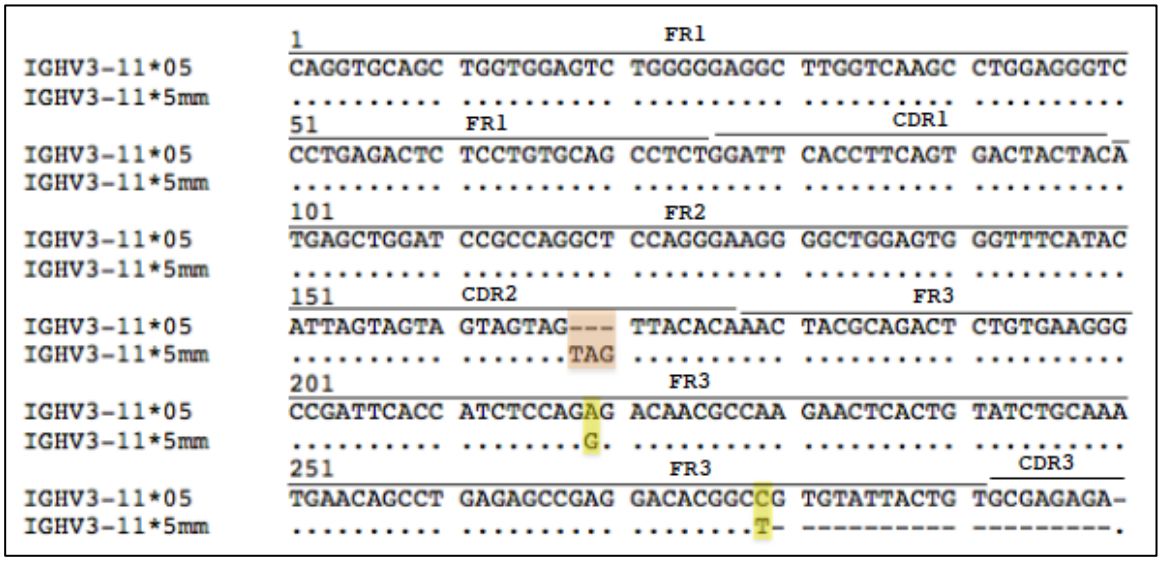
Primer ID	Adapter, Tag and MID sequence	IGHV sequence
MID1-VH1-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	CAGCTKGTRCARTCTGGRSCTG
MID1-VH1-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	AGGATGTGGKTTYTCACACTGTG
MID1-VH2-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	TTCTCCACAGGGGTCTTRTC
MID1-VH2-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	CCTGGGCTGTGTCTYTGTGGT
MID1-VH3-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	TGTTTGCAGRTGTCCARTGT
MID1-VH3a-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	TRRCYTCYCCTCRCTGTG
MID1-VH3b-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	GTGTYTCTMRYACAGTAATACA
MID1-VH4-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	GGCTCACTGTGKGTYTYTCTGTT
MID1-VH4-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	CTCACACTCACCTCCCCTCA
MID1-VH5-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	CCMYACAGGAGTCTGTKCCG
MID1-VH5-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	GGTTTCTCTCACTGTGTGTCT
MID1-VH6-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	TTTTGTCTCCAGGTGTCCT
MID1-VH6-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	GACTTCCCCTCACTGTGTCT
MID1-VH7-F	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT	CTCTCCACAGGTRCCYACTCC
MID1-VH7-R	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT	TCAGGATGTGGGTTTCCACA
MID2-VH1-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	CAGCTKGTRCARTCTGGRSCTG
MID2-VH1-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	AGGATGTGGKTTYTCACACTGTG
MID2-VH2-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	TTCTCCACAGGGGTCTTRTC
MID2-VH2-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	CCTGGGCTGTGTCTYTGTGGT
MID2-VH3-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	TGTTTGCAGRTGTCCARTGT
MID2-VH3a-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	TRRCYTCYCCTCRCTGTG
MID2-VH3b-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	GTGTYTCTMRYACAGTAATACA
MID2-VH4-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	GGCTCACTGTGKGTYTYTCTGTT
MID2-VH4-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	CTCACACTCACCTCCCCTCA
MID2-VH5-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	CCMYACAGGAGTCTGTKCCG
MID2-VH5-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	GGTTTCTCTCACTGTGTGTCT
MID2-VH6-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	TTTTGTCTCCAGGTGTCCT
MID2-VH6-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	GACTTCCCCTCACTGTGTCT
MID2-VH7-F	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA	CTCTCCACAGGTRCCYACTCC
MID2-VH7-R	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA	TCAGGATGTGGGTTTCCACA
MID3-VH1-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	CAGCTKGTRCARTCTGGRSCTG
MID3-VH1-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	AGGATGTGGKTTYTCACACTGTG
MID3-VH2-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	TTCTCCACAGGGGTCTTRTC
MID3-VH2-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	CCTGGGCTGTGTCTYTGTGGT
MID3-VH3-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	TGTTTGCAGRTGTCCARTGT
MID3-VH3a-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	TRRCYTCYCCTCRCTGTG
MID3-VH3b-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	GTGTYTCTMRYACAGTAATACA
MID3-VH4-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	GGCTCACTGTGKGTYTYTCTGTT
MID3-VH4-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	CTCACACTCACCTCCCCTCA
MID3-VH5-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	CCMYACAGGAGTCTGTKCCG
MID3-VH5-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	GGTTTCTCTCACTGTGTGTCT
MID3-VH6-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	TTTTGTCTCCAGGTGTCCT
MID3-VH6-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	GACTTCCCCTCACTGTGTCT
MID3-VH7-F	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC	CTCTCCACAGGTRCCYACTCC
MID3-VH7-R	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTC	TCAGGATGTGGGTTTCCACA
MID4-VH1-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	CAGCTKGTRCARTCTGGRSCTG
MID4-VH1-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	AGGATGTGGKTTYTCACACTGTG
MID4-VH2-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	TTCTCCACAGGGGTCTTRTC
MID4-VH2-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	CCTGGGCTGTGTCTYTGTGGT
MID4-VH3-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	TGTTTGCAGRTGTCCARTGT

MID4-VH3a-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	TRRCYTCYCCTCRCTGTG
MID4-VH3b-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	GTGTYTCTMRYACAGTAATACA
MID4-VH4-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	GGCTCACTGTGKGTYTYTCTGTT
MID4-VH4-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	CTCACACTCACCTCCCCTCA
MID4-VH5-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	CCMYACAGGAGTCTGTKCCG
MID4-VH5-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	GGTTTCTCTCACTGTGTGTCT
MID4-VH6-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	TTTTGTCTCCAGGTGTCCT
MID4-VH6-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	GACTTCCCCTCACTGTGTCT
MID4-VH7-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAG	CTCTCCACAGGTRCCYACTCC
MID4-VH7-R	CTATGCGCCTTGCCAGCCCGCTCAGAGCACTGTAG	TCAGGATGTGGGTTTCCACA

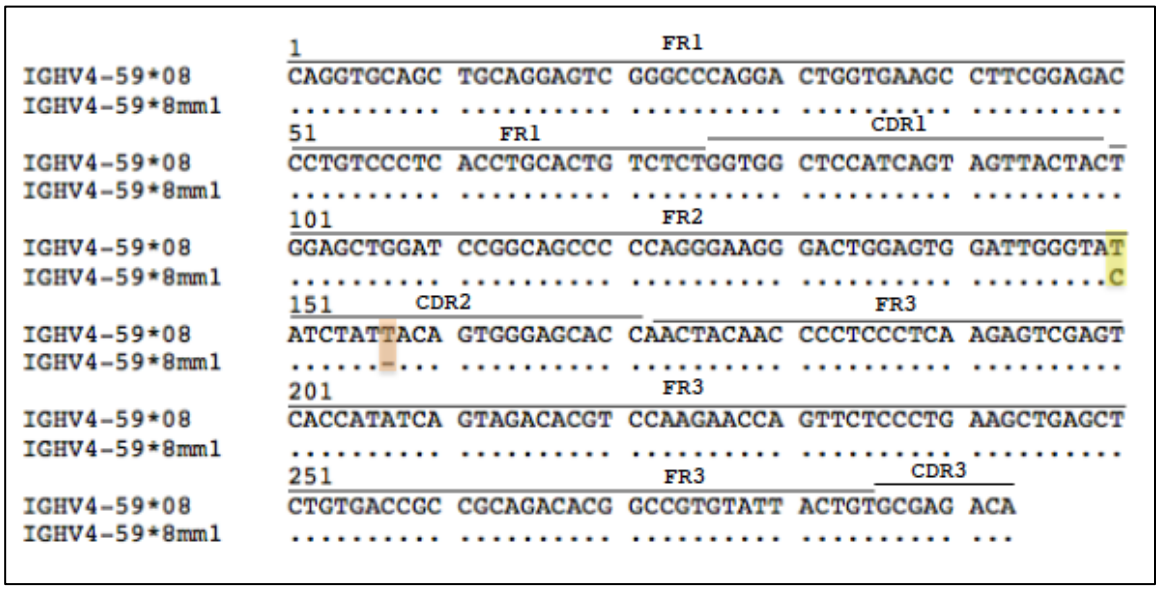
Supplementary Table 1.3: Germline IGHV primers for Illumina sequencing

Read 1 and Read 2 sequences were used for MiSeq 2x250bp paired-end sequencing. As with the 454 sequencing two IGHV3 reverse primers (IGHV3a and 3b) were used to ensure full coverage of the IGHV3 subgroup.

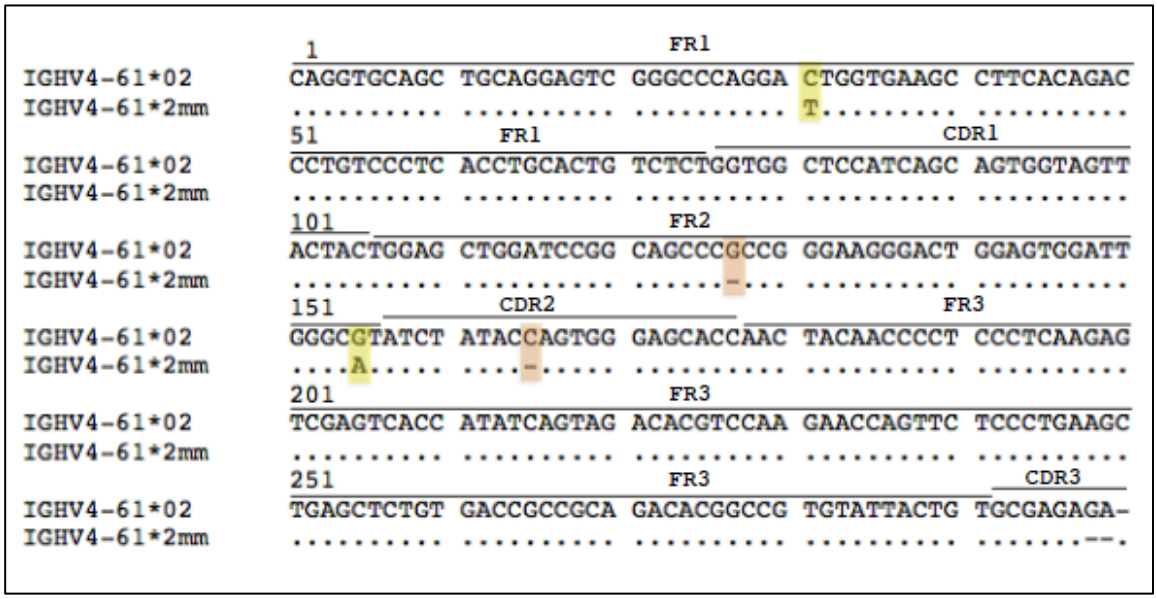
Name	Index Binding Site	Read 1/Read2	Gene Specific Primer
IGHV1 -Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	CAGCTKGTRCARTCTGGRSCTG
IGHV-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	AGGATGTGGKTTYTCACACTGTG
IGHV2-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	TTCTCCACAGGGGTCTTRTC
IGHV2-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	CCTGGGCTGTGTCTYTGTGGT
IGHV3-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	TGTTTGCAGRTGTCCARTGT
IGHV3a-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	TRRCYTCYCCTCRCTGTG
IGHV3b-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	GTGTYTCTMRYACAGTAATACA
IGHV4-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	GGCTCACTGTGKGTYTYTCTGTT
IGHV4-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	CTCACACTCACCTCCCCTCA
IGHV5-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	CCMYACAGGAGTCTGTKCCG
IGHV5-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	GGTTTCTCTCACTGTGTGTCT
IGHV6-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	TTTTGTCTCCAGGTGTCCT
IGHV6-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	GACTTCCCCTCACTGTGTCT
IGHV7-Forward	TCGTCGGCAGCGTC	AGATGTGTATAAGAGACAG	CTCTCCACAGGTRCCYACTCC
IGHV7-Reverse	GTCTCGTGGGCTCGG	AGATGTGTATAAGAGACAG	TCAGGATGTGGGTTTCCACA



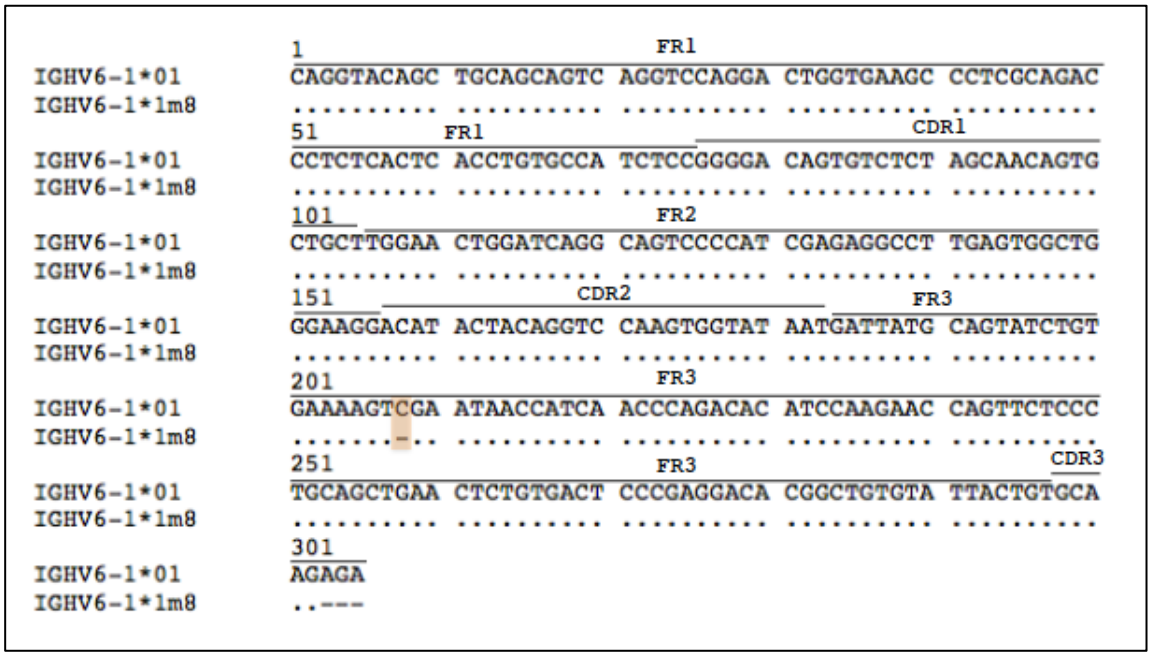
Supplementary Figure 1.1: Nucleotide sequence alignment for IGHV3-11*5mm
Shown is an alignment of IGHV3-11*5mm and IGHV3-11*05 sequences with a 3 nucleotide insertion (highlighted in orange) in the complementarity determining region 2 (CDR2) and 2 SNPs (highlighted in yellow) within framework 3 (FR3). Missing data in the sequence is represented by "-".



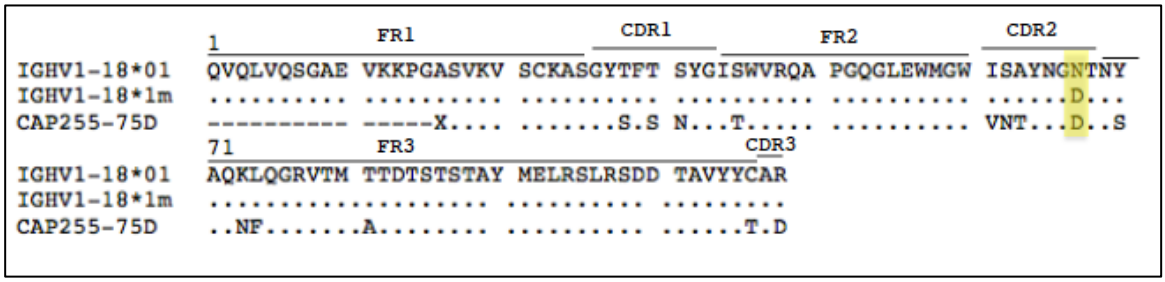
Supplementary Figure 1.2: Nucleotide sequence alignment for IGHV4-59*8mm1
Shown is an alignment of IGHV4-59*8mm1 (likely to be a pseudogene) and IGHV4-59*08 sequences with a single nucleotide deletion (highlighted in orange) in the complementarity determining region 2 (CDR2) and a SNP (highlighted in yellow) in the framework 2 region (FR2). Missing data in the sequence is represented by "-".



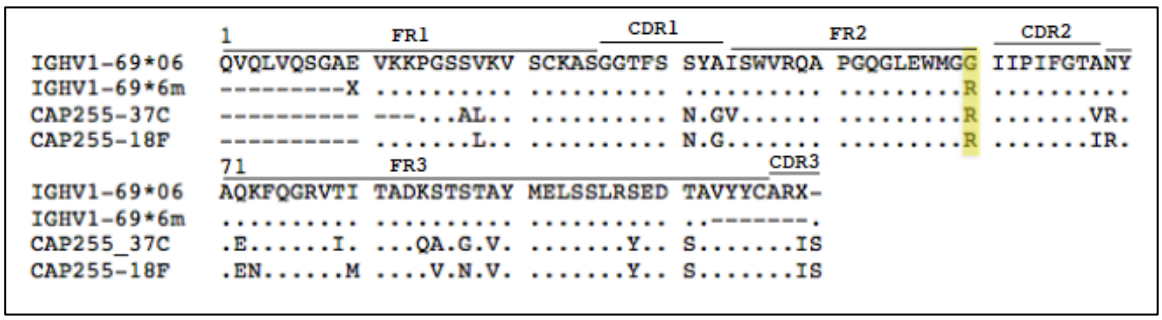
Supplementary Figure 1.3: Nucleotide sequence alignment for IGHV4-61*2mm
Shown is an alignment of IGHV4-61*2mm (likely to be a pseudogene) and IGHV4-61*02 with two single nucleotide deletions (highlighted in orange) in the framework 2 region (FR2) and complementarity determining region 2 (CDR2) and two SNPs (highlighted in yellow) in framework 1 (FR1) and framework 2 (FR2) regions. Missing data in the sequence is represented by "-".



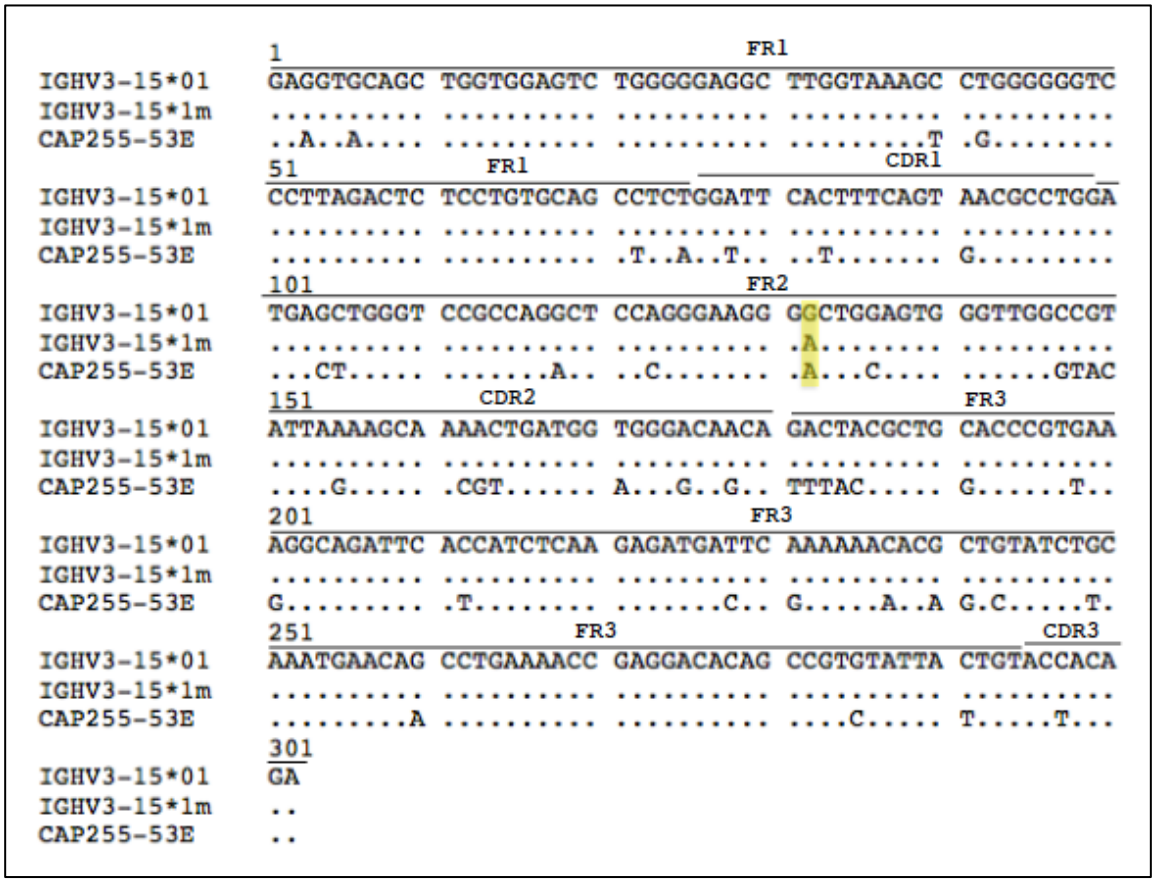
Supplementary Figure 1.4: Nucleotide sequence alignment for IGHV6-1*1m8
Shown is an alignment for IGHV6-1*1m8 (likely to be a pseudogene) and IGHV6-1*01 with a single nucleotide deletion (highlighted in orange) in the framework 3 (FR3) region. Missing data in the sequence is represented by "-".



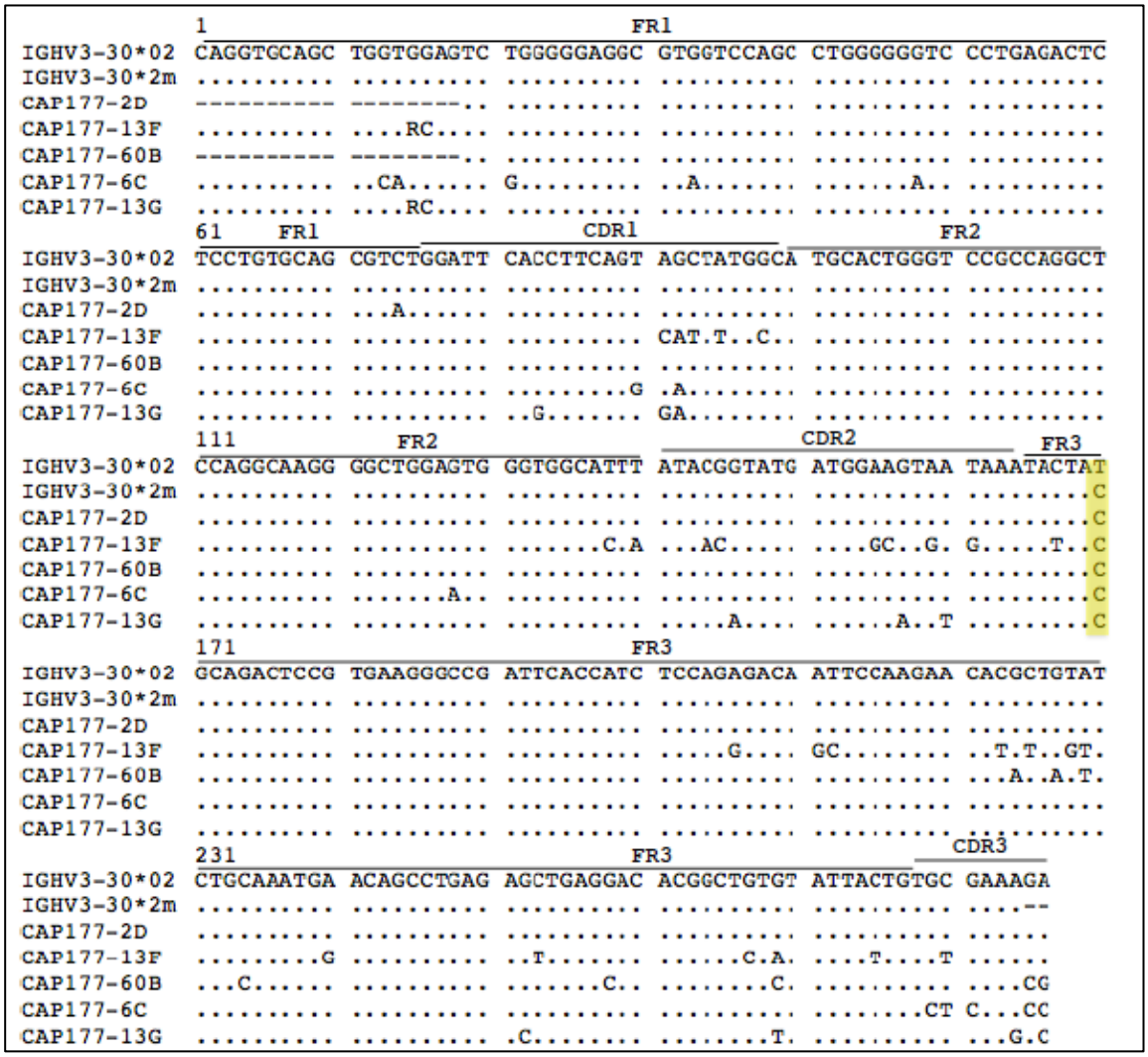
Supplementary Figure 1.5: Germline IGHV gene usage for CAP255-75D
CAP255-75D is predicted to use IGHV1-18*1m. The closest matching IMGT allele is IGHV1-18*01. Both CAP255-75D and IGHV1-18*1m have an Aspartic Acid (D) at position 66, in the complementarity determining region 2 (CDR2), compared to Asparagine (N), highlighted in yellow, as seen in IGHV1-18*01 at the same position. All other changes in the antibody sequence are likely due to somatic hypermutation. Missing data in the sequence is represented by "-".



Supplementary Figure 1.6: Germline IGHV gene usage for CAP255-37C and CAP255-18F
CAP255-37C and CAP255-18F are predicted to use IGHV1-69*6m. The closest matching IMGT allele is IGHV1-69*06. CAP255-37C, CAP255-18F and IGHV1-69*6m have an Arginine (R) at position 59, in framework 2 (FR2) compared to Glycine (G), highlighted in yellow, as seen in IGHV1-69*06 at the same position. All other changes in the antibody sequence are likely due to somatic hypermutation. Missing data in the sequence is represented by "-".

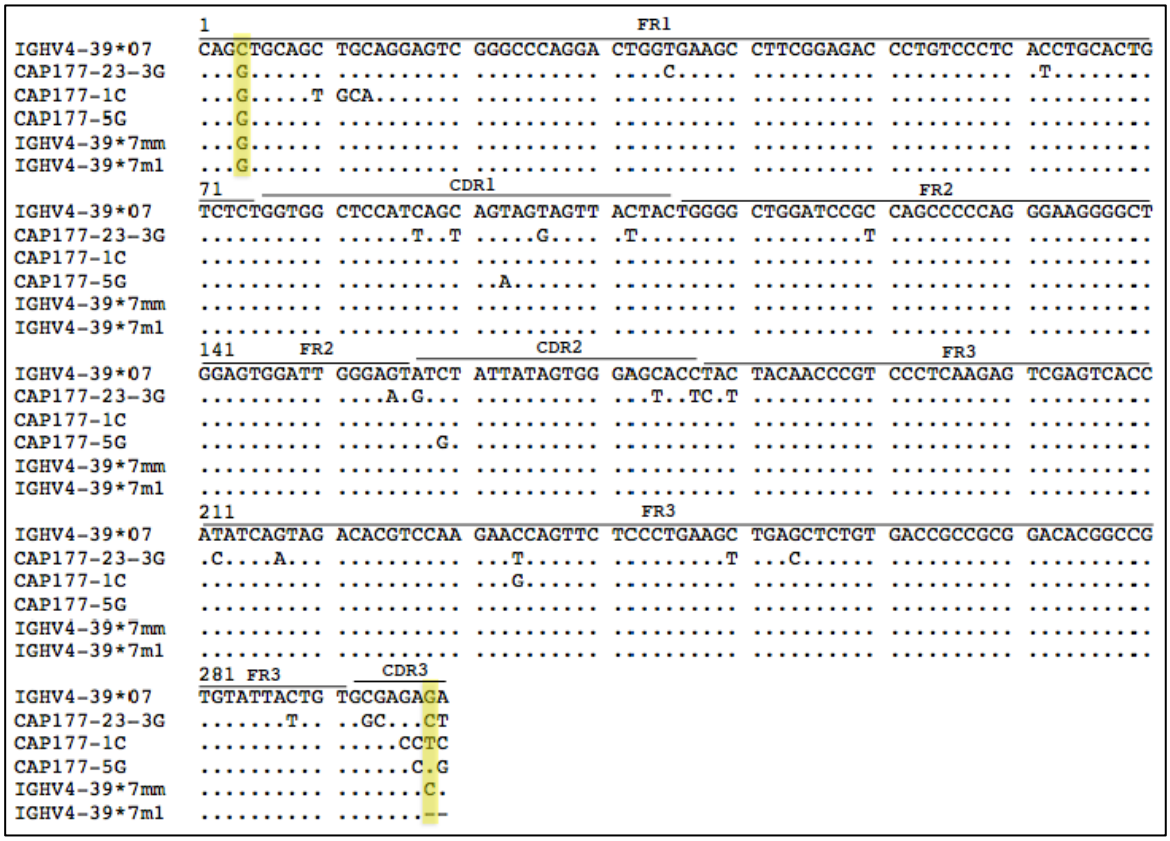


Supplementary Figure 1.7: Germline IGHV gene usage for CAP255-53E
CAP255-53E is predicted to use IGHV3-15*1m. The closest matching IMGT allele is IGHV3-15*01. Both CAP255-53E and IGHV3-15*1m have an Adenine (A) at position 165, in the framework 2 (FR2), compared to Guanine (G), highlighted in yellow, as seen in IGHV3-15*01 at the same position. This change is synonymous and therefore doesn't change the amino acid sequence. All other changes in the antibody sequence are likely due to somatic hypermutation. Missing data in the sequence is represented by "-".



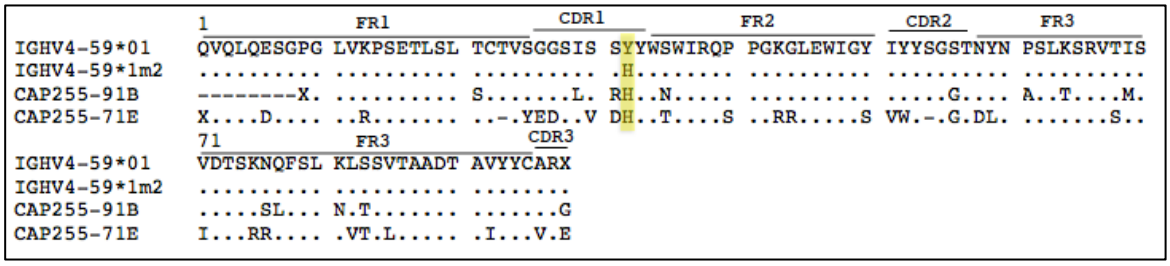
Supplementary Figure 1.8: Germline IGHV gene usage for CAP177-2D, 13F, 60B, 6C and 13G

Five mAbs (2D, 13F, 60B, 6C and 13G) isolated from CAP177 were predicted to use IGHV3-30*2m. The closest matching IMGT allele is IGHV3-30*02. All five mAbs and IGHV3-30*2m have a Cytosine (C) at position 192 in the framework 1 (FR1), compared to Thymine (T), highlighted in yellow, as seen in IGHV3-30*02 at the same position. This change is synonymous and therefore doesn't change the amino acid sequence. All other changes in the antibody sequence are likely due to somatic hypermutation. Missing data in the sequence is represented by "-".

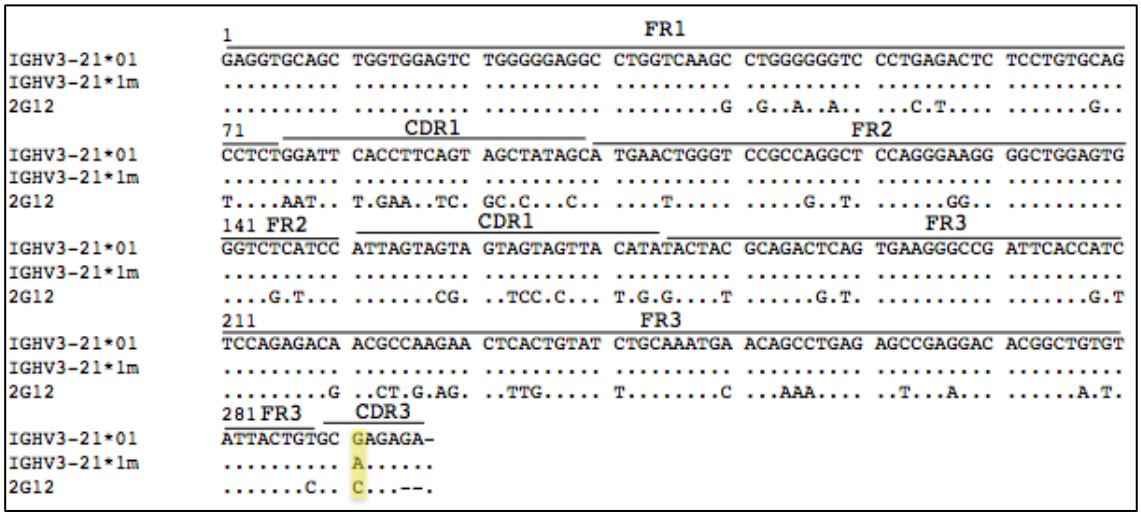


Supplementary Figure 1.9: Germline IGHV gene usage for CAP177-23-3G, 1C and 5G

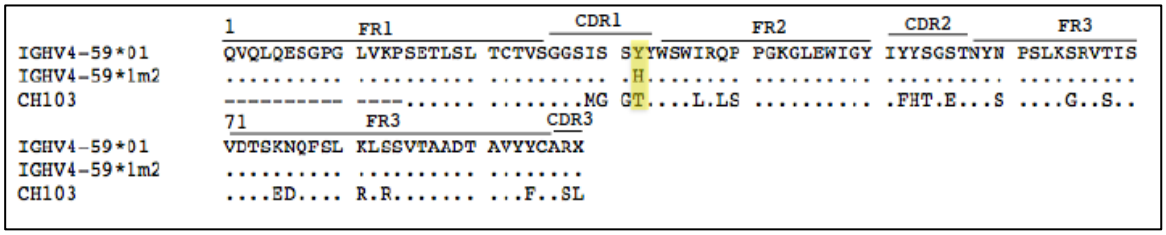
CAP177-23-3G was predicted to use IGHV4-39*7mm, which has a Guanine (G) at position 4 (which causes a Leucine (L) to Valine (V) amino acid change at position 2) within framework 1 (FR1), and Cytosine (C) at position 298 (no amino acid change) in the complementarity determining region 3 (CDR3), highlighted in yellow, compared to a Cytosine (C) and Guanine (G) at the same positions, respectively. The other two mAbs (CAP177-1C and CAP177-5G) were predicted to use either IGHV4-39*7mm or IGHV4-39*7m1. The difference between IGHV4-39*7mm and IGHV4-39*7m1 is the G298C which is present on IGHV4-39*7mm but absent on IGHV4-39*7m1. All other changes in the antibody sequence are likely due to somatic hypermutation. Missing data in the sequence is represented by "-".



Supplementary Figure 1.10: Germline IGHV gene usage for CAP255-91B and 71E
CAP255-91B and CAP255-71E were predicted to use IGHV4-59*1m2, all of which have a Histidine (H) at position 32, within the complementarity determining region 1 (CDR1), compared to the closest matching IGHV4-59*01 which has a Tyrosine (Y) at the same position (highlighted in yellow). All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".

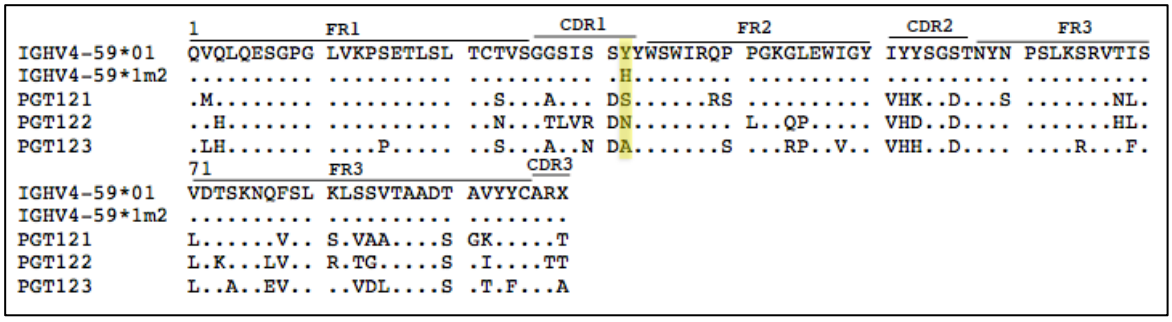


Supplementary Figure 1.11: Germline IGHV gene usage for 2G12
2G12 has been reported to use IGHV3-21*01 (Xiao*, et al.*, 2009). However, when compared to our database of IMGT, novel and non-IMGT alleles, 2G12 is just as likely to be using IGHV3-21*1m. The difference between IGHV3-21*01 and IGHV3-21*1m is highlighted in yellow, where IGHV3-21*01 has a Guanine (G) at position 291, in the complementarity determining region 3 (CDR3), IGHV3-21*1m has an Adenine (A) and 2G12 has a Cytosine (C) at the same position. The difference between IGHV3-21*01 and IGHV3-21*1m doesn't change the amino acid sequence. All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".



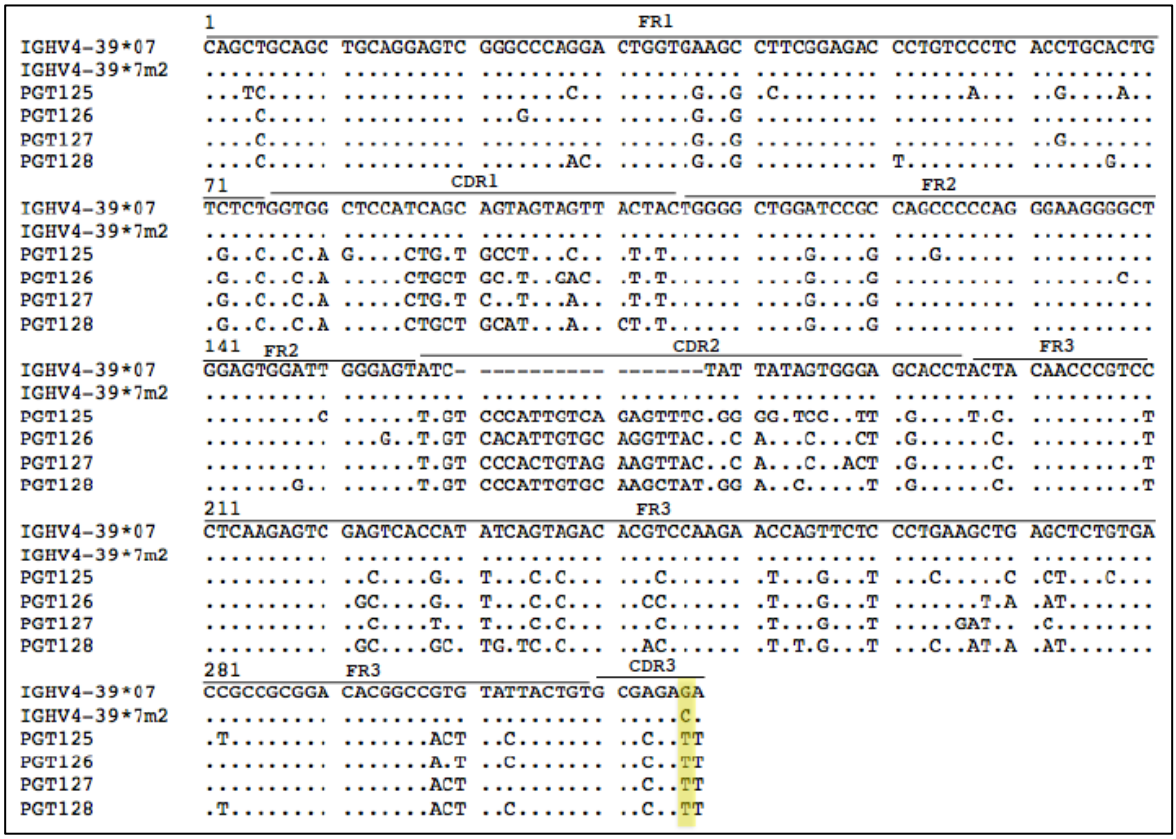
Supplementary Figure 1.12: Germline IGHV gene usage for CH103

CH103 has been reported to use IGHV4-59*01 (Liao*, et al.*, 2013). However, when compared to our database of IMGT, novel and non-IMGT alleles, CH103 is just as likely to be using IGHV4-59*1m2. The difference between IGHV4-59*01 and IGHV4-59*1m2 is highlighted in yellow, where IGHV4-59*01 has a Tyrosine (Y) at position 32, in the complementarity determining region 1 (CDRH1), IGHV4-59*1m2 has a Histidine (H) and CH103 has a Threonine (T) at the same position. All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".



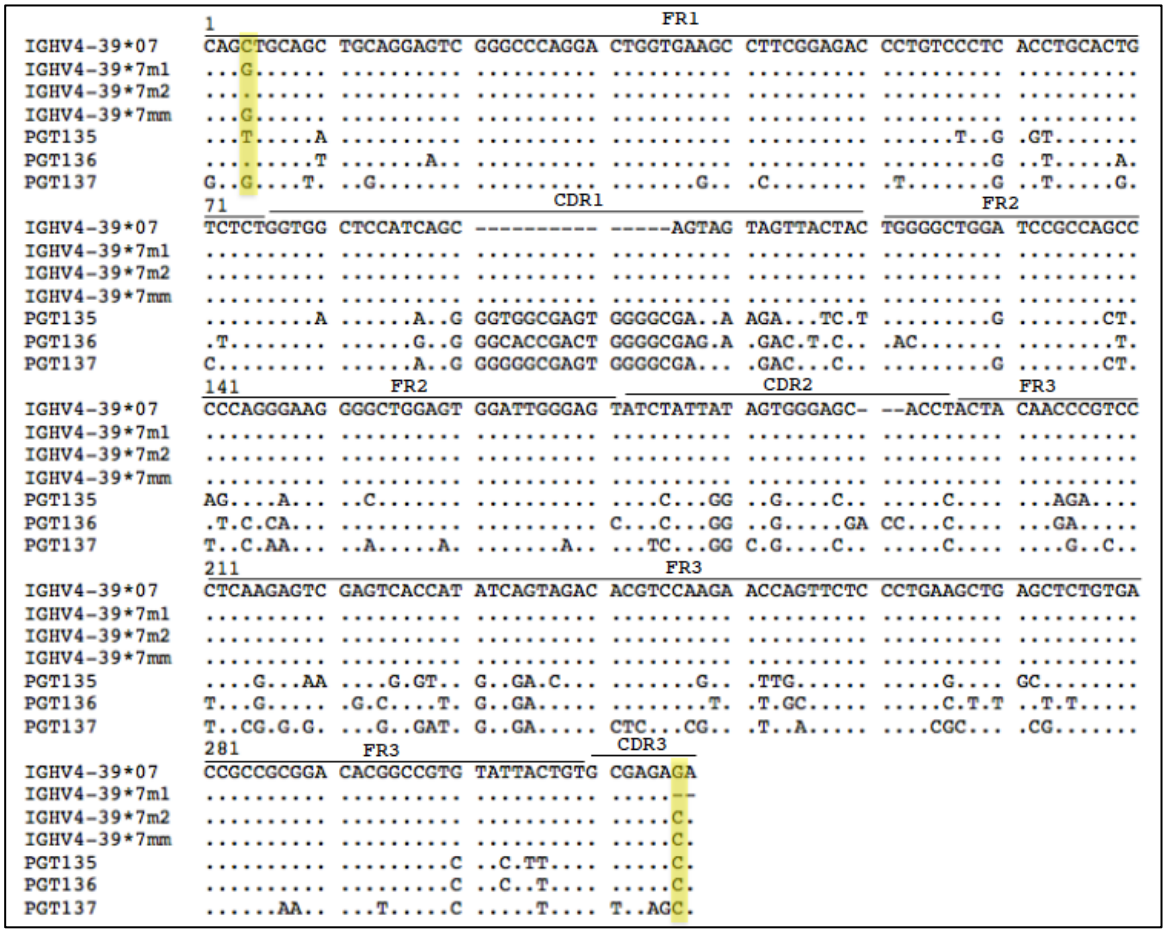
Supplementary Figure 1.13: Germline IGHV gene usage for PGTs121-123

PGTS121-123 have been reported to use IGHV4-59*01 (Walker*, et al.*, 2011). However, when compared to our database of IMGT, novel and non-IMGT alleles, PGT121-123 are just as likely to be using IGHV4-59*1m2. The difference between IGHV4-59*01 and IGHV4-59*1m2 is highlighted in yellow, where IGHV4-59*01 has a Tyrosine (Y) at position 32, in the complementarity determining region 1 (CDR1), IGHV4-59*1m2 has a Histidine (H) and the PGTs have a Serine (S), Asparagine (N) or Alanine (A) at the same position. All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".



Supplementary Figure 1.14: Germline IGHV gene usage for PGTs125-128

PGTS125-128 have been reported to use IGHV4-39*07 (Walker*, et al.*, 2011). However, when compared to our database of IMGT, novel and non-IMGT alleles, PGT125-128 are just as likely to be using IGHV4-39*7m2. The difference between IGHV4-59*07 and IGHV4-39*7m2 is highlighted in yellow, where IGHV4-39*07 has a Guanine (G) at position 316, within the complementarity determining region 3 (CDR3) IGHV4-39*7m2 has a Cytosine (C) and the PGTs have a Thymine (T) at the same position. All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".



Supplementary Figure 1.15: Germline IGHV gene usage for PGTs135-137

PGTS135-137 have been reported to use IGHV4-39*07 (Walker, *et al.*, 2011). However, when compared to our database of IMGT, novel and non-IMGT alleles, PGT125-128 are just as likely to be using IGHV4-39*7m1/7m2/7mm. The differences between IGHV4-59*07 and IGHV4-39*7m1/7m2/7mm are highlighted in yellow, where IGHV4-39*07 has a Cytosine (C) at position 4 in framework 1 (FR1) and Guanine (G) at position 316 in the complementarity determining region 3 (CDR3). IGHV4-39*7m1 and IGHV4-39*7mm have a Guanine, while IGHV4-39*7m2 has a Cytosine (C) at position 4 and the PGTs have a either a Guanine (G), Thymine (T) or Cytosine at the same position. At position 316 IGHV4-39*7m2 and IGHV4-39*7mm have a Cytosine as do all three PGTs. All other changes are likely due somatic hypermutation. Missing data in the sequence is represented by "-".

Ability To Develop Broadly Neutralizing HIV-1 Antibodies Is Not Restricted by the Germline Ig Gene Repertoire

Cathrine Scheepers,[*,†] Ram K. Shrestha,[‡] Bronwen E. Lambson,[*,†] Katherine J. L. Jackson,[§,¶] Imogen A. Wright,[‡] Dshanta Naicker,[*] Mark Goosen,[*] Leigh Berrie,[*] Arshad Ismail,[*] Nigel Garrett,[‖,#] Quarraisha Abdool Karim,[‖,] Salim S. Abdool Karim,[‖,**] Penny L. Moore,[*,†,‖] Simon A. Travers,[‡] and Lynn Morris[*,†,‖]**

The human Ig repertoire is vast, producing billions of unique Abs from a limited number of germline Ig genes. The IgH V region (IGHV) is central to Ag binding and consists of 48 functional genes. In this study, we analyzed whether HIV-1–infected individuals who develop broadly neutralizing Abs show a distinctive germline IGHV profile. Using both 454 and Illumina technologies, we sequenced the IGHV repertoire of 28 HIV-infected South African women from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 and 004 cohorts, 13 of whom developed broadly neutralizing Abs. Of the 259 IGHV alleles identified in this study, approximately half were not found in the International Immunogenetics Database (IMGT). This included 85 entirely novel alleles and 38 alleles that matched rearranged sequences in non-IMGT databases. Analysis of the rearranged H chain V region genes of mAbs isolated from seven of these women, as well as previously isolated broadly neutralizing Abs from other donors, provided evidence that at least eight novel or non-IMGT alleles contributed to functional Abs. Importantly, we found that, despite a wide range in the number of IGHV alleles in each individual, including alleles used by known broadly neutralizing Abs, there were no significant differences in germline IGHV repertoires between individuals who do and do not develop broadly neutralizing Abs. This study reports novel IGHV repertoires and highlights the importance of a fully comprehensive Ig database for germline gene usage prediction. Furthermore, these data suggest a lack of genetic bias in broadly neutralizing Ab development in HIV-1 infection, with positive implications for HIV vaccine design. ***The Journal of Immunology*****, 2015, 194: 4371–4378.**

The induction of broadly neutralizing Abs (bNAbs) is likely to be crucial for an efficacious HIV vaccine. Although the majority of chronically HIV-infected individuals develop some level of cross-neutralizing activity (1), bNAbs are generally found in <20% of HIV-infected individuals (2). The mechanisms underlying bNAb emergence are largely unknown, but a better understanding of how these Abs arise in natural infection would provide a blueprint for a vaccine designed to elicit them. Over the last few years a large number of potent and broad bNAbs have been isolated from selected HIV-infected donors. These bNAbs target conserved epitopes on the HIV envelope, including the membrane proximal external region (MPER), V2 glycans, V3 glycans, CD4 binding site (CD4bs), and the gp120/gp41 interface (3, 4). However, most bNAbs have unusual genetic features, including high levels of somatic hypermutation (SHM), long CDRH3s (the complementary determining region 3 on the H chain) (3, 5), and, for some classes, biases in germline IgH V region (IGHV) gene usage (3, 5–9).

The VRC01 class of Abs to the CD4bs (including NIH45-46, 12A12, 3BNC117, VRC-PG04, and VRC-CH31 isolated from

*Centre for HIV and Sexually Transmitted Infections, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg 2131, South Africa; [†]Division of Virology and Communicable Disease Surveillance, School of Pathology, University of the Witwatersrand, Johannesburg 2050, South Africa; [‡]South African National Bioinformatics Institute, South African Medical Research Council Bioinformatics Unit, University of the Western Cape, Bellville 7535, South Africa; [§]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia; [¶]Department of Pathology, School of Medicine, Stanford University, Stanford, CA 94305; [‖]Centre for the AIDS Programme of Research in South Africa, KwaZulu-Natal 4013, South Africa; [#]Department of Infectious Diseases, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, 4041 Durban, South Africa; and **Department of Epidemiology, Columbia University, New York, NY 10032

Received for publication January 16, 2015. Accepted for publication February 24, 2015.

This work was supported by the Poliomyelitis Research Foundation, the University of the Witwatersrand Health Sciences Faculty Research Council, and the National Research Foundation. The Centre for the AIDS Programme of Research in South Africa is funded by the National Institute of Allergy and Infectious Diseases/National Institutes of Health and the U.S. Department of Health and Human Services (Grant AI51794). The South African National Bioinformatics Institute received funding from the South African Department of Science and Technology, South African National Research Foundation Study Bursary, Atlantic Philanthropies, and the South African Medical Research Council. C.S. was supported by the Columbia University–Southern African Fogarty AIDS International Training and Research Program through the Fogarty International Center, National Institutes of Health (Grant 5 D43 TW000231). P.L.M. is a Wellcome Trust Intermediate Fellow in Public Health and Tropical Medicine (Grant 089933/Z/09/Z).

Address correspondence and reprint requests to Dr. Lynn Morris, National Institute for Communicable Diseases, a Division of the National Health Laboratory Service, Centre for HIV and Sexually Transmitted Infections: HIV Virology Section (Morris Laboratory), 1 Modderfontein Road, Sandringham 2131, South Africa. E-mail address: lynnm@nicd.ac.za

The online version of this article contains supplemental material.

Abbreviations used in this article: BCN, broadly cross-neutralizing; BLAST, basic local alignment search tool; bNAb, broadly neutralizing Ab; CAPRISA, Centre for the AIDS Programme of Research in South Africa; CD4bs, CD4 binding site; FR, framework; IGHV, IgH V region; IMGT, International Immunogenetics Database; MPER, membrane proximal external region; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; non-BCN, non–broadly cross-neutralizing; SHM, somatic hypermutation; SNP, single nucleotide polymorphism; UCA, unmutated common ancestor; VH, H chain V region.

Copyright © 2015 by The American Association of Immunologists, Inc. 0022-1767/15/$25.00

www.jimmunol.org/cgi/doi/10.4049/jimmunol.1500118

multiple donors) was shown to preferentially use either IGHV1-2*02 or IGHV1-46*02 germline alleles (3, 10, 11). This preference is thought to be due to the electrostatic and hydrophobic contacts afforded by conserved residues in framework (FR)1, CDRH2, and FR3 of this allele (some of these features are conserved in IGHV1-46*02 and IGHV1-3*01, which are also used by CD4bs Abs) (9, 10). Anti-HIV Abs frequently use IGHV1 genes compared with non-HIV Abs (5, 6). In particular, IGHV1-69 is used by mAbs that target V2, the CD4 induced site, and gp41 in HIV infection, as well as other viral infections, such as influenza (8, 12). The preference for this particular gene can be attributed to the interaction of hydrophobic residues in the CDRH2 with helical elements or hydrophobic β-sheets like those found on the hemagglutinin of influenza and gp41 and gp120 of the HIV envelope (8, 13). The use of IGHV5-51*01 and IGHV5-51*03, which are underrepresented in mature Abs, also was reported to be favored by anti-V3 HIV Abs (6, 7).

The human germline IGHV repertoire consists of seven IGHV subgroups, which are described in the International Immunogenetics Database (IMGT; www.imgt.org). These IGHV1–7 subgroups include functional genes, open-reading frames, and pseudogenes with only functional genes being involved in Ab production (14). IGHV3 is the largest of the IGHV subgroups, with 21 functional genes, followed by IGHV1 and IGHV4 (both with 10 functional genes), and IGHV2, IGHV5, IGHV6, and IGHV7 (with three or fewer genes) (14). Most of these genes have multiple alleles, including functional and nonfunctional alleles that differ by either a single nucleotide polymorphism (SNP) or by multiple SNPs, which can be either synonymous or nonsynonymous, or frameshift mutations caused by indels that contribute diversity to the Ig gene repertoire. Furthermore, whole IGHV genes were reported to have been duplicated or deleted from the germline repertoire of some individuals, resulting in varied gene copy numbers (15, 16). IGHV genes make up the majority of the H chain V region (VH) of mature Abs and are central to Ag binding. Differences in germline IGHV repertoires between different populations were highlighted recently in an extensive study of the human Ig gene locus (16), in which African individuals were found to be particularly diverse.

Given the propensity of some HIV mAbs to use restricted IGHV genes, we examined whether HIV-infected individuals who develop bNAbs have unique IGHV repertoires compared with individuals who do not, to determine whether the ability to develop these types of Abs is genetically restricted.

Materials and Methods

Samples and ethics statement

This study involved 28 adult (>18 y) women of African ancestry (mostly Zulu speaking) with HIV-1 subtype C infection from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 (17) and 004 (18) cohorts who were being followed at urban and rural clinics in KwaZulu-Natal, South Africa. Of the 28 women, 13 developed broadly neutralizing anti-HIV Abs, 13 did not develop broadly neutralizing Abs despite chronic HIV infection, and 2 were intermediate neutralizers (Supplemental Table I). Throughout this article, "BCN" is used to refer to the HIV-infected women who develop broadly neutralizing Abs, and "non-BCN" refers to women who did not develop broadly neutralizing Abs. BCN individuals were defined as those whose sera from 2 y postinfection were able to neutralize 33–94% (median 56%) of a panel of 18 viruses, made up of 6 subtype A, 6 subtype B, and 6 subtype C viruses, of which 2 viruses were isolated from the women in the cohort. Non-BCN individuals neutralized 0–11% of the 18 viruses and had the same viral loads as the BCN individuals at 6 mo postinfection [to remove viral load biases associated with the development of bNAbs (19)].

This study was given ethics clearance from the Human Research Ethics Committee for Medical Research in Johannesburg, South Africa (clearance number M111104 for this particular study and M080470 for the CAPRISA parent study).

DNA extraction

Genomic DNA was extracted from PBMCs from each individual. Prior to extraction, PBMCs were thawed (at 37°C) and washed with 10 ml RPMI 1640 with 10% FBS. Genomic DNA was extracted from the pellet from all 28 individuals using a Promega Wizard Genomic purification kit.

Primer design and amplicon library construction

Alignments were created for each IGHV subgroup using sequences obtained from ENSEMBL (http://www.ensembl.org/index.html) and IMGT (http://www.imgt.org/). Forward primers (with the exception of IGHV1-F) were mapped to the intron in leader sequence. The reverse primers were mapped to the intron after the CDR3 region. IGHV1, IGHV3, and IGHV4 primers were designed based on previously published primer sets (20). New primers for IGHV2, IGHV5, IGHV6, and IGHV7 were designed based on related sequences for each subgroup. A total of eight primer sets was designed to amplify all seven subgroups. The binding properties of the primer sets were determined using University of California Santa Cruz's BLAT database (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) and the National Center for Biotechnology Information (NCBI)'s basic local alignment search tool (BLAST). Additional next-generation sequencing (NGS)-specific sequences were added to the gene-specific primers to allow sequencing of the Roche 454 and Illumina MiSeq. The 454 primers included the 454 adapter sequence, key sequence, and 10-bp MID sequences (four unique sequences in total; primers are listed in Supplemental Table II). The Illumina MiSeq primers included a MiSeq index-binding tag and the read 1 or read 2 tags (used for paired-end reads; primers are listed in Supplemental Table III). Nextera XT Indexing tags, which allow pooling of samples on the MiSeq, were added during library construction rather than included into the primer.

Each IGHV subgroup was amplified three times for each individual to ensure adequate coverage of the subgroup and minimize PCR bias. The PCR conditions for all eight amplicons and both 454 and Illumina primers were the same, with the exception of the annealing temperatures: 56°C for IGHV1, IGHV3a, and IGHV5; 59°C for IGHV2, IGHV6, and IGHV7; and 55°C for IGHV3b and IGHV4. The PCR conditions were as follows: initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 15 s, annealing for 45 s, extension for 1 min at 72°C, final extension at 72°C for 8 min, and held at 4°C. Each PCR contained 17.5 μl dH_2O, 2.5 μl Roche FastStart High Fidelity 10× buffer with 18 mM $MgCl_2$, 0.5 μl deoxyribonucleotide triphosphates mix (10 mM each), 1 μl each primer (10 μM), 0.25 μl Roche FastStart High Fidelity Enzyme (5 U/μl), and 1 μl 10 ng/μl DNA. The PCR amplicon lengths ranged from ~390 to 440 bp. All replicates of the eight amplicons from each individual were pooled to create a full IGHV repertoire for each individual.

Next-generation sequencing

Fifteen individuals were sequenced on the Roche 454 GS Junior, with four individuals pooled per sequencing run. All clean-up procedures and sequencing on the Roche 454 GS Junior (Titanium) was done as per the manufacturer's recommendations.

To confirm rare alleles that were only found in single individuals, 11 of the individuals sequenced on the GS Junior were resequenced on the Illumina MiSeq along with 13 additional individuals. A custom amplicon-sequencing approach was used for the MiSeq sequencing. Nextera XT Indexing tags were added to the pooled MiSeq amplicon libraries for each individual using 5 μl each Nextera XT index (two per sample), 1 μl cleaned PCR product (amplicon library prep), 0.5 μl Epicentre FailSafe enzyme, 25 μl Epicentre FailSafe PCR PreMix, and 13.5 μl dH_2O. Thermocycler conditions were 72°C for 3 min, 95°C for 30 s, 12 cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 30 s, followed by a final extension at 72°C for 5 min. All products were checked on an Agilent Bioanalyzer and cleaned up using 0.75× AMPure Beads, according to the manufacturer's protocol. Each sample was quantified on a Qubit and diluted to 8 nM. A single 8-nM pooled library was created by pooling 4 μl each diluted sample, 5 μl of which was denatured using 5 μl 0.2 N NaOH, according to the MiSeq protocol. A final concentration of 12 pM denatured DNA library with 15% PhiX control was run onto the Illumina MiSeq, using the MiSeq reagent kit (version 2) with 2 × 250 paired-end reads.

Sequence data analysis

FASTQ files were extracted from the raw GS Junior output (.sff files) or automatically generated on the MiSeq. The resulting sequences were quality trimmed using QTrim (21). Sequences with a minimum read length of 260 bp and a mean quality score ≥ 23 (≥ 99% confidence) were considered high quality and used for downstream analyses. Paired-end MiSeq reads were merged using PEAR (22), after quality trimming. All identical sequences

were collapsed to single unique reads. Unique reads compiled from six or more identical sequences were used for further analysis, whereas all unique reads compiled from less than six sequences were discarded from further analysis. The unique sequences were compared with a database of IGHV alleles downloaded from IMGT using a custom BLAST approach with a match reward score of 2, a mismatch penalty of 5, a gap penalty for insertions of 16, and a gap extension for deletions with a penalty of 4. All related sequences for each IGHV gene were aligned using RAMICS (23).

Sequences with 100% matches to functional IMGT alleles were assigned as that particular IMGT allele. Sequences matching open-reading frames or pseudogenes were removed from further analysis. Sequences with non-synonymous and/or synonymous SNPs, insertions, or deletions compared with their top matched functional IMGT sequence were given the name of the top-scoring IMGT match with the suffix "m" to denote a single mismatch (nonsynonymous and synonymous SNPs) (e.g., IGHV1-2*5m) and "mm" for multiple SNPs, insertions, or deletions (e.g., IGHV1-2*5mm). If more than one sequence with a single mismatch or multiple mismatches to the same top-scoring IMGT functional sequence was observed, an additional numerical identifier was given to the name (e.g., IGHV1-2*5m2 and IGHV1-2*5mm2). These sequences were compared with those listed in the Ig Polymorphism Database, GenBank, NCBI's dbSNP, NCBI's IgBLAST, and other published IGHV data (24–28). If 100% matches were found in any of these databases or publications, these sequences are described as non-IMGT alleles. If no match was found, and sequences were observed in multiple individuals or in both sequencing platforms in a single individual, we report these as novel alleles.

bNAb sequences were obtained from the HIV Sequence Database - CATNAP (http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp) or GenBank. The germline IGHV gene usage for these sequences also was obtained from CATNAP or the relevant publication. The bNAb sequences were compared with our IMGT, non-IMGT, and novel allele sequences, using blastn from BLAST 2.2.29+ with a match reward of 1, a mismatch penalty of −1, gap deletions of 5, a gap extension of 2, a word size of 7, and an E value threshold of $1e^{-10}$. mAbs were isolated previously from seven of the CAPRISA participants examined in this study using single-cell sorting and PCR amplification of the H and L chains (N. Mkhize, E. Gray, and L. Morris, unpublished observations). The majority of these mAbs were not HIV specific. The sequences obtained from these mAbs were used to assign germline IGHV gene usage in each participant.

Statistical analysis

The two-tailed Fishers exact test with 95% confidence intervals was used to assess the difference in germline gene repertoires between the BCN and non-BCN sample groups.

Results

Novel germline IGHV alleles in South African women

The germline IGHV gene repertoire of 28 HIV-infected women in the CAPRISA 002 and 004 cohorts, based in KwaZulu-Natal, South Africa, were sequenced using both the 454 and Illumina NGS platforms. Sequences with perfect matches to alleles listed in IMGT (29) were given the IMGT allele name and are referred to as

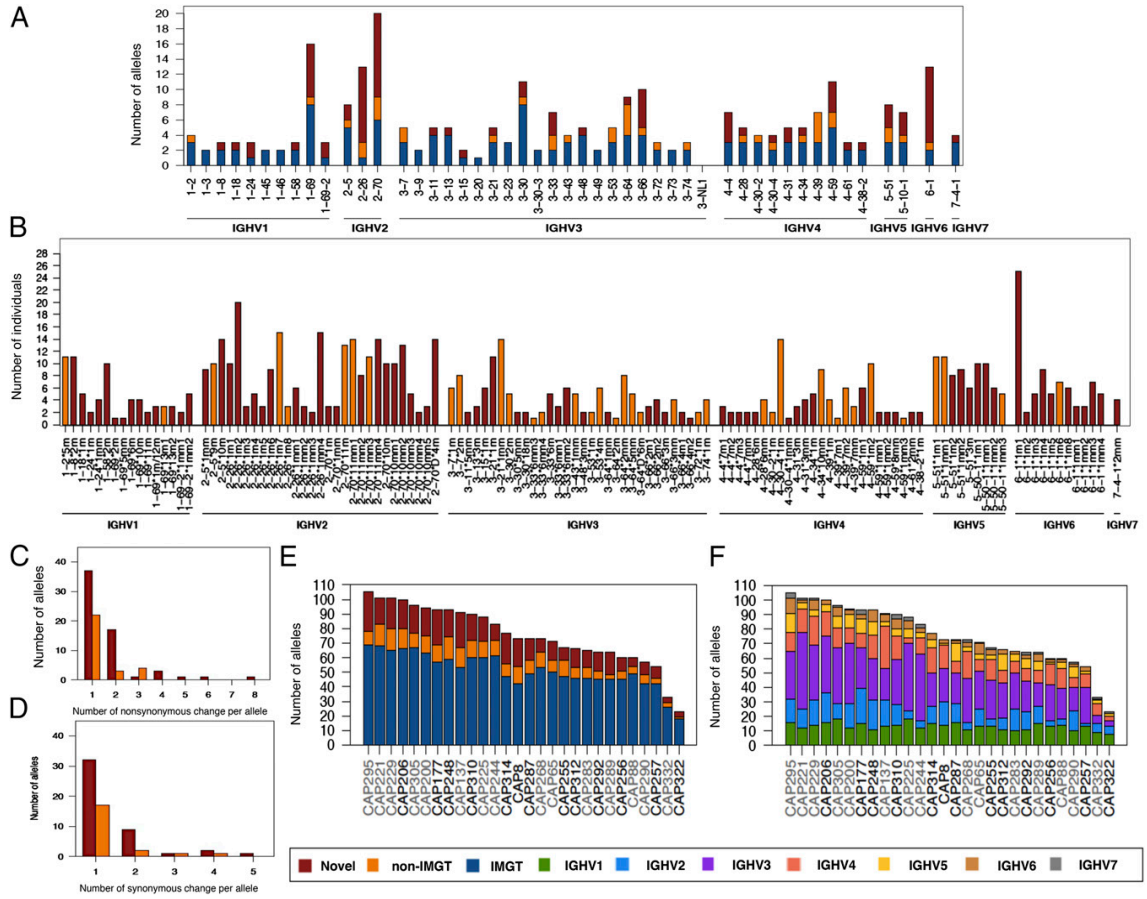


FIGURE 1. Novel germline IGHV alleles in 28 South African individuals. (**A**) Number of alleles observed for each IGHV gene. Shown are the numbers of novel (red), non-IMGT (orange), and IMGT (blue) alleles. (**B**) Prevalence of novel and non-IMGT germline IGHV alleles. (**C**) Number of non-synonymous changes in each allele compared with top-matched IMGT allele. (**D**) Number of synonymous changes in each allele compared with top-matched IMGT allele. (**E**) Total number of IMGT, novel, and non-IMGT germline IGHV alleles. BCN individuals are highlighted in black, and non-BCN individuals are highlighted in gray. (**F**) Total number of germline alleles from each IGHV subgroup. Shown are the total numbers of alleles colored according to IGHV subgroup.

IMGT alleles. Sequences that did not match IMGT alleles as a result of synonymous or nonsynonymous mutations or indels, but were matches to non-IMGT Ig gene databases (30–33) or Ig gene publications (24–28), were given the published name and are referred to as non-IMGT alleles. Sequences with no matches to either IMGT or non-IMGT alleles and were observed in multiple individuals or across both technologies in one individual are referred to as "novel" alleles.

A total of 47 functional IGHV genes representing all seven subgroups was found in our cohort (Fig. 1A). Only one published gene, IGHV3-NL1, first identified in Papua New Guinea (20), was not detected. There was a wide range in the number of alleles associated with each gene, with 20 alleles in the IGHV2-70 gene compared with only a single allele for the IGHV3-20 gene. Of the 259 alleles identified in this population, just over half (~52%, n = 136) had an exact match to those listed in IMGT, whereas ~15% (n = 38) were non-IMGT alleles, and ~33% (n = 85) were novel alleles. Eighty-one of the eighty-five novel alleles (95%) were observed in more than one individual (Fig. 1B). The four novel alleles observed in single individuals were confirmed with both 454 and Illumina sequences. The most commonly observed novel alleles were IGHV6-1*1m1 found in 25 individuals (~89%) and IGHV2-26*1m2 found in 20 individuals (~71%), both of which contained a single SNP compared with the known alleles. Overall, 48% (n = 60) of the novel and non-IMGT alleles identified in this study were found in at least four individuals, indicating that many are fairly common. Furthermore, this study significantly expanded the number of alleles for certain genes. For example, IGHV6-1, which only had 2 alleles reported in IMGT was found to have 10 novel alleles and 1 non-IMGT allele (Fig. 1A, 1B). Similarly, IGHV2-26 was found to have 13 alleles (including 10 novel alleles and 2 non-IMGT alleles) compared with 1 recorded in IMGT. Other significant expansions were in the IGHV1-69 and IGHV2-70 genes.

The majority (~97%, n = 119) of novel and non-IMGT alleles had single or multiple SNPs (synonymous and nonsynonymous),

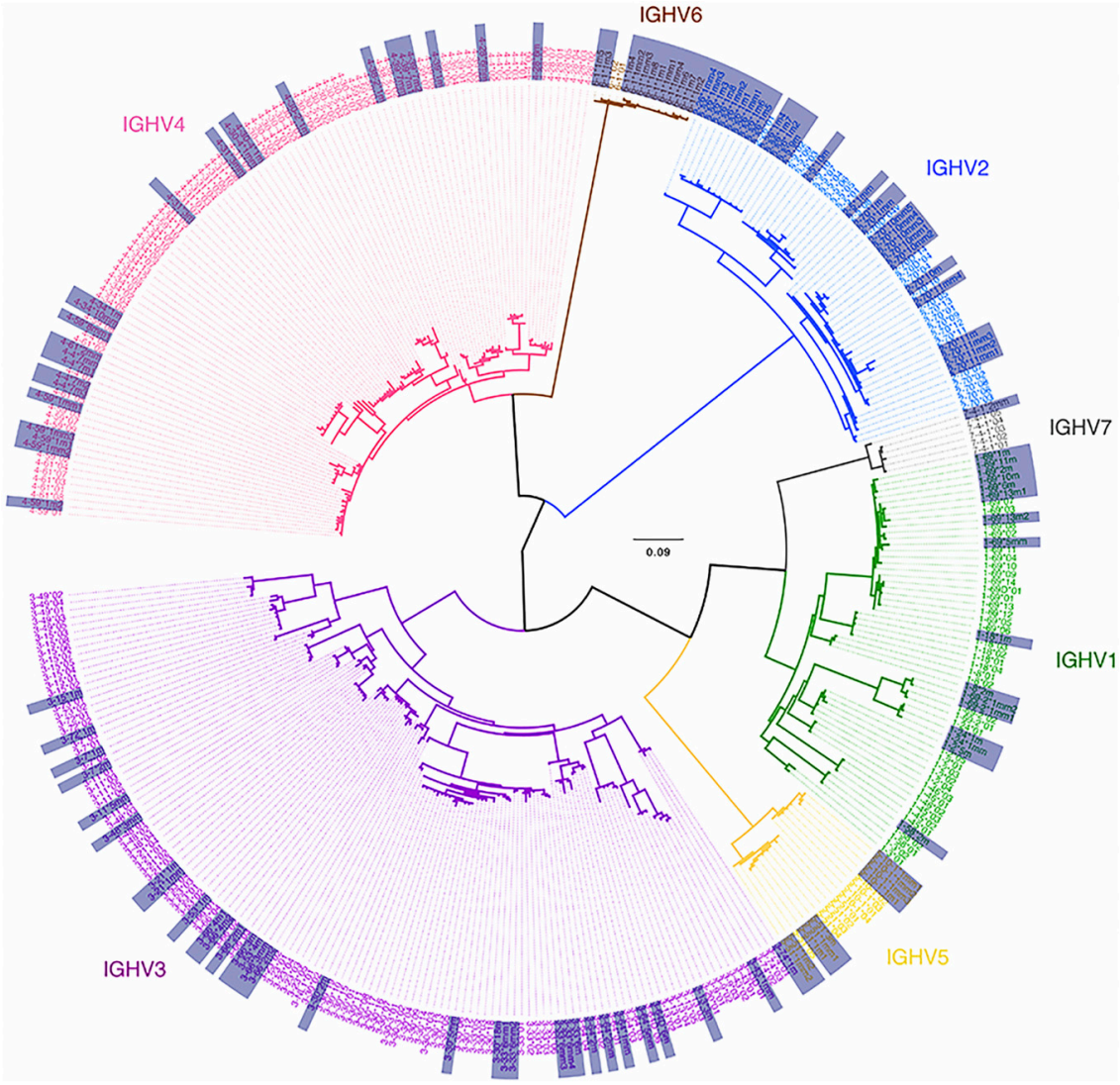


FIGURE 2. Phylogenetic tree of all germline IGHV alleles observed in 28 South African individuals. Shown are all IMGT, non-IMGT, and novel alleles observed in this study. Alleles are colored according to IGHV subgroup, and the non-IMGT and novel alleles are highlighted in the gray boxes.

with single nonsynonymous SNPs being the most common (Fig. 1C, 1D). One novel allele (IGHV2-26*1mm4, observed in 15 individuals) had eight nonsynonymous SNPs (Fig. 1B, 1C). Four novel alleles had indels relative to their most closely matched IMGT allele, and three of the four had additional SNPs. IGHV3-11*5mm, observed in two individuals, had a full codon insertion, whereas frameshifts as a result of single nucleotide deletions were found in IGHV4-59*8mm and IGHV4-61*2mm, each observed in two individuals, and IGHV6-1*1m8, which was found in six individuals. It is likely that these three alleles with the frameshift mutations are not functional, but rather are novel pseudogenes.

When the repertoires for each of the 28 individuals were analyzed separately, it was apparent that there was a wide range in the number of alleles/individual (Fig. 1E, 1F). CAP295 had the largest number of alleles (n = 105), whereas CAP322 had the fewest (n = 23). Each individual's repertoire included IMGT, non-IMGT, and novel alleles. There was variation in the number of alleles/IGHV family, because some individuals lacked alleles for IGHV6 or IGHV7, whereas others had smaller numbers of IGHV2, IGHV4, or IGHV5 alleles. A phylogenetic tree of all 259 alleles, including IMGT, non-IMGT, and novel alleles, sequenced in this study revealed separate clustering of each of the IGHV subgroups, with all of the novel and non-IMGT alleles clustering with their respective IMGT genes (Fig. 2).

Novel germline IGHV allele usage in bNAbs

To determine whether these novel and non-IMGT germline alleles are being used by the human immune system to generate functional Abs, we analyzed the germline gene usage of mAbs isolated from seven of the CAPRISA participants in this study. We found 15 mAbs that made use of seven novel or non-IMGT alleles identified from NGS of the 28 individuals (Fig. 3A). Of these, two clonally related mAbs were HIV-1 specific (CAP255-37C and CAP255-18F). An example of one of the non-IMGT alleles used by two mAbs isolated from two individuals (CAP177 and CAP248) is shown in Fig. 3B. Based on the IGHV repertoires from CAP177 and CAP248, the germline IGHV gene predicted to be used by these two mAbs is IGHV2-5*9m, which has a serine (S) at position 62 rather than a glycine (G) present in the closest related IMGT allele IGHV2-5*09. Compared with the germline allele, all other SNPs found in these two mAbs are likely the results of SHM. Of the remaining seven novel and non-IMGT alleles used by these mAbs, four (57%) had a better match at the amino acid level and three (43%) had a better match at the nucleotide level compared with the closest matching IMGT allele (Fig. 3A). We also analyzed the IGHV gene usage of 57 well-known HIV-1 bNAbs and found 14 instances in which the novel or non-IMGT alleles, identified in this study, provided the same or a better match than did their currently predicted IMGT allele (Fig. 3C). Two of the predicted alleles (IGHV4-39*7m2 and IGHV4-59*1m2) also were observed among the mAbs isolated from the CAPRISA participants. An example of alternate IGHV allele usage for the bNAbs is shown in Fig. 3D: PGT130 and PGT131 are more likely to be using IGHV4-39*7m1 or IGHV4-39*7mm, which like the mAbs have a valine (V) at the second amino acid, rather than a leucine (L), as seen in IGHV4-39*07.

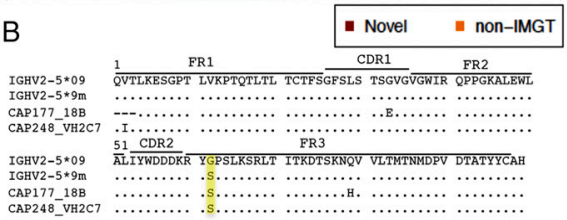
Comparison of germline IGHV repertoires in BCN and non-BCN individuals

Given that some HIV broadly neutralizing mAbs show biased variable H chain gene usage, we analyzed the germline IGHV repertoires of BCN and non-BCN individuals to determine whether the ability to develop bNAbs was associated with a particular genotypic profile. Among the 28 CAPRISA women, 13 developed bNAbs in their plasma after 2–4 y of HIV infection; 13 women did not develop bNAbs, despite matching viral loads; and 2 were intermediate neutralizers [Supplemental Table I (19) and N. Mkhize, E. Gray, and L. Morris, unpublished observations]. Comparison of the BCN and non-BCN individuals revealed that they had similar profiles of IGHV genes (Fig. 4A), approximately the same number of alleles for each of those genes (Fig. 4B), and the same number of potential alleles (Fig. 1E, 1F). The only gene that showed

A

Participant	mAb	Predicted Allele	Corresponding IMGT Allele	Nucleotide Change	Amino Acid Change
CAP255	75D	IGHV1-18*1m	IGHV1-18*01	A169G	N57D
CAP255	37C, 18F	IGHV1-69*6m	IGHV1-69*06	G148A	G50R
CAP177, CAP248	18B, VH2C7	IGHV2-5*9m	IGHV2-5*09	G184A	G62S
CAP255	53E	IGHV3-15*1m	IGHV3-15*01	G132A	none
CAP177	2D, 13F, 60B, 6C, 13G	IGHV3-30*2m	IGHV3-30*02	T180C	none
CAP177	23-3G, 1C, 5G	IGHV4-39*7m2	IGHV4-39*07	G298C	none
CAP255	71E, 91B	IGHV4-59*1m2	IGHV4-59*01	T94C	Y32H

B



C

bNab	IMGT Allele	Alternate Allele(s)	Nucleotide Change	Amino Acid Change
2G12	IGHV3-21*01	IGHV3-21*1m	G291A	none
CH103	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT121-123	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT125-128	IGHV4-39*07	IGHV4-39*7m2	G298C	none
PGT130-131	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7mm	C4G C4G, G298C	L2V L2V
PGT135-137	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7m2 IGHV4-39*7mm	C4G G298C C4G, G298C	L2V none L2V

D

FIGURE 3. Novel IGHV alleles used by isolated mAbs and some bNAbs. (**A**) Germline IGHV allele usage for isolated mAbs from study participants. The germline IGHV alleles used, as well as their corresponding IMGT allele, nucleotide, and amino acid substitutions between the two alleles, are shown. (**B**) Alignment of germline IGHV allele usage for isolated mAbs from two study participants. Highlighted in yellow are the differences in amino acids at position 62: G (glycine) and S (serine). Amino acid changes not highlighted in the figure are likely a result of SHM during Ab maturation. (**C**) Germline IGHV allele usage for selected bNAbs. The published IMGT predicted allele and alternate allele(s) predicted from IGHV alleles in 28 South African individuals are shown. (**D**) Alignment of predicted germline IGHV allele usage for PGT130 and PGT131. Highlighted in yellow are the differences at the second amino acid: V (valine) and L (leucine). Amino acid changes not highlighted in the figure are likely a result of SHM during Ab maturation.

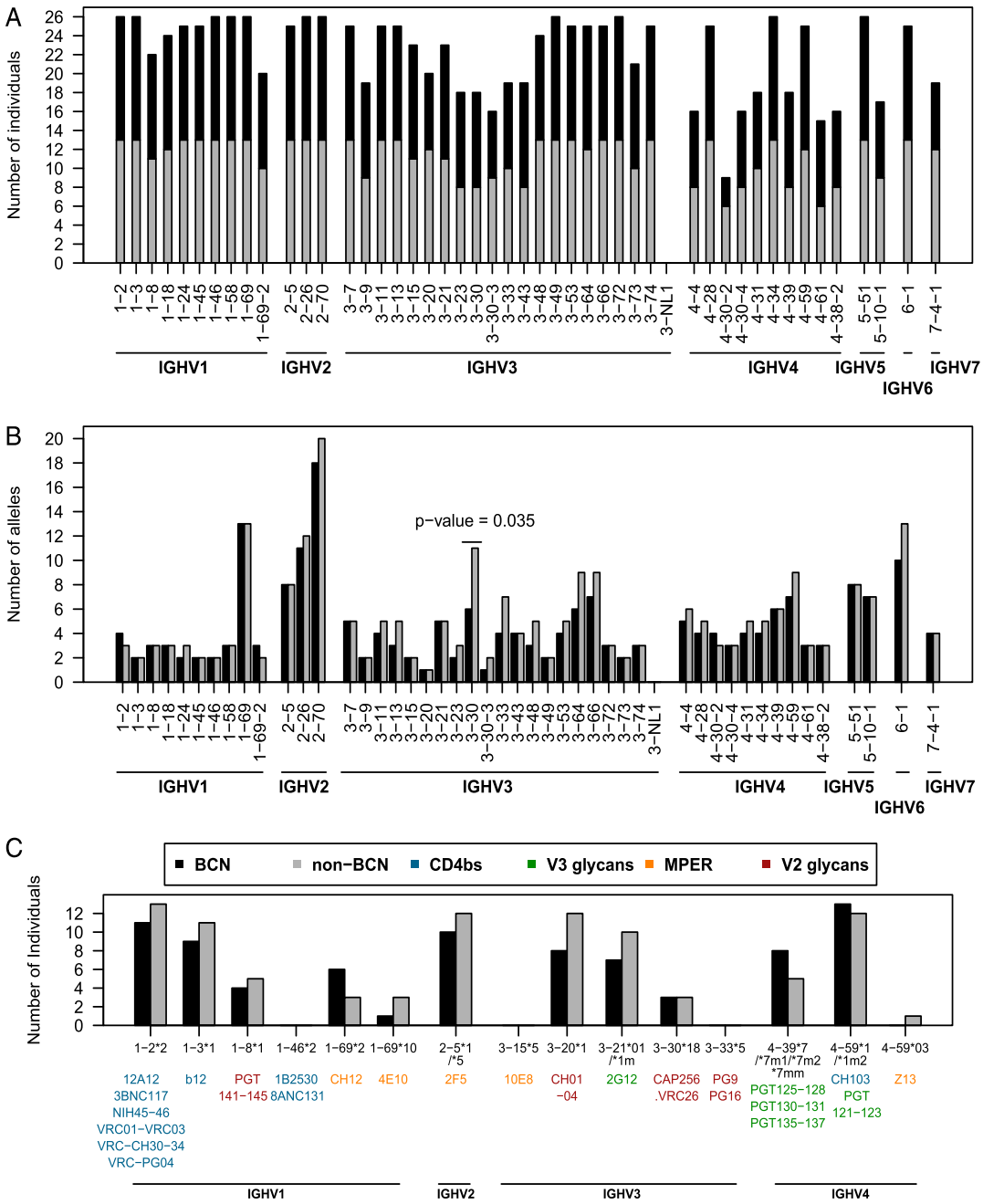


FIGURE 4. Comparison of germline IGHV repertoires between BCN and non-BCN individuals. (**A**) Number of BCN and non-BCN individuals with each of the functional IGHV genes in their IGHV repertoire. (**B**) Number of alleles observed for each of the functional IGHV genes in BCN and non-BCN individuals. (**C**) Number of BCN and non-BCN individuals with the same IGHV alleles in their repertoire as those used by monoclonal bNAbs.

statistical significance between the two groups was IGHV3-30 (p = 0.035, although significance was lost following Bonferroni multiple testing correction); non-BCN individuals had five more alleles (n = 11) compared with BCN individuals (n = 6) (Fig. 4B).

We also compared the frequency of the specific germline IGHV alleles used by broadly neutralizing mAbs between the non-BCN and BCN groups. Of the 57 monoclonal bNAbs targeting one of four different epitopes on the HIV envelope glycoprotein, most made use of IGHV1, IGHV3, and IGHV4 germline genes (Fig. 4C), and, as reported above, some may use novel/non-IMGT alleles (Fig. 3C). This included IGHV1-46*02, used by some CD4bs Abs; however, there were two other alleles for this gene, including IGHV1-46*01, which were found in all individuals. Similarly, IGHV3-15*05 (used by bNAb 10E8) and IGHV3-33*05 (used by bNAbs PG9 and PG16) were not observed in any of the individuals in this study, but other alleles were more commonly observed,

including some novel/non-IMGT alleles. The IGHV1-2*02 allele, which is also used preferentially by CD4bs Abs (shown in blue in Fig. 4C), occurred at high frequency in both the non-BCN and BCN groups. The most prevalent IGHV alleles in both groups were IGHV4-59*01 (used by the CD4bs bNAb CH103 and the V3/glycan bNAbs PGT121–123), IGHV1-2*02 (used by CD4bs mAbs), IGHV1-3*01 (used by the CD4bs bNAb IgG1b12), and IGHV2-5*01/05 (used by bNAb 2F5 that targets the MPER). Importantly, there was no difference in the frequency of any of these alleles between the 13 BCN and 13 non-BCN individuals. Thus, the inability of all HIV-infected individuals to develop bNAbs is not due to a paucity of the relevant alleles in their IGHV germline gene repertoires.

Discussion

We analyzed the germline IgH variable gene repertoire encoded in the genomic DNA of individuals from KwaZulu-Natal, South Africa, and noted that ~48% of the alleles seen in this population are not reported in IMGT. Some of these alleles (non-IMGT alleles) were described in published studies of rearranged Abs, although most alleles were novel and are described in this article for the first time, to our knowledge. Further analysis of these IGHV repertoires revealed there to be no differences between individuals who developed bNAbs to HIV and those who did not, despite equivalent antigenic load. Because the induction of these types of Abs is considered essential for an effective HIV vaccine, these data suggest that the ability to develop bNAbs is not restricted by the IGHV repertoire.

Previous studies reported differences in the frequencies of IGHV genes between different populations, with Africans showing particularly unique profiles (16). The presence of IGHV3-64D, IGHV5-10-1, and IGHV7-4-1 was reported to be lower in African (Luyha, Maasai, and Yoruba) populations compared with Asian and European groups (16). We found a similarly low frequency of IGHV3-64D (14%) in this Zulu-speaking South African population. However, IGHV5-10-1 was observed at a higher prevalence (64%) compared with the studied African, Asian, and European groups (0.03–16, 20–21, and 34–48%, respectively) (16), whereas IGHV7-4-1 frequencies were similar (75%) to those seen in Asian (~78%) and European (54–75%) groups (16). IGHV1-69-2, IGHV3-43D, and IGHV4-38-2 genes were reported to be common in African populations (16), which we corroborated in our study for IGHV1-69-2 and IGHV4-38-2 (75 and 61%, respectively). However, we observed IGHV3-43D*01 in only 29% (8/28) of individuals studied, which is lower than the prevalence reported in other African groups (45–65%) (16).

We showed that both novel and non-IMGT alleles are being used by mAbs isolated from individuals in this study, as well as some well-characterized anti-HIV bNAbs. This included CH103 and 12 Abs in the PGT121-137 series, all of which were isolated from African donors (34–36). Given that three of the novel and non-IMGT alleles (IGHV3-21*1m, IGHV4-39*1m2, and IGHV4-39*7m2) potentially being used by these bNAbs were fairly common in the South African individuals (39, 36 and 21%, respectively), it is perhaps not surprising that these were found to contribute to functional Abs in other African individuals. IGHV4-39*7m1 and IGHV4-39*7mm were less common in our study group (4 and 11%, respectively). The other novel and non-IMGT alleles used by mAbs isolated from the study participants also were fairly commonly observed, ranging from 14 to 36%. Only 2 of the 15 mAbs isolated from the study participants were able to bind HIV, highlighting that the use of novel and non-IMGT alleles by functional Abs is not HIV specific and, thus, could play a role in immune responses to other diseases or infections.

Despite the wide range in the number of IGHV alleles present in each individual, there were no differences in the overall germline IGHV repertoires between BCN and non-BCN individuals. This extended to a subanalysis of the genes and alleles used by known broadly neutralizing mAbs whose frequency differed within the cohort but not between BCN and non-BCN groups. The only gene that showed a difference between the two groups was IGHV3-30: non-BCN individuals had more alleles than did BCN individuals. IGHV3-30 is used by the CAP256-VRC26 family of broad and potent V1V2 Abs, which were isolated from CAPRISA donor CAP256, a participant in this study (37). We also noted that alleles IGHV1-46*02, IGHV3-15*05, and IGHV3-33*05, used by CD4bs, MPER, and V2 glycan bNAbs, respectively, were not found in this South African cohort. However, other alleles of these genes were observed in this cohort, as well as allelic variants used by other bNAbs targeting the same epitopes. Thus, although the specific alleles used by these bNAbs were not present, it did not preclude these individuals from generating Abs to these epitopes. In addition, CH103 and PGT121-123, which both use IGHV4-59*01 (or the non-IMGT allele IGHV4-59*1m2), target different epitopes (CD4bs and V3 glycans, respectively), demonstrating that a single germline allele can be used by Abs against numerous HIV envelope targets.

The germline IGHV gene makes the greatest contribution to the unmutated common ancestor (UCA) Ab encoding the entire CDRH1, CDRH2, FR1, FR2, and FR3 regions. Elegant studies on the UCAs of the VRC01 class of Abs, which use the VH1-2*02 allele, showed how two glycans in the V5 region of the HIV envelope obstruct binding of the UCA (9, 38). Such studies guide the design of suitable envelope immunogens able to trigger the VRC01 class of Abs (9). However, other studies aimed at identifying UCA-binding envelopes for PG9 and other bNAb lineages were less successful, primarily because the UCA approximations used in such studies were predicted from highly mutated Abs and using incomplete germline databases. The public availability of a comprehensive database of germline Ig genes, including data from this study (available from the authors), will significantly enhance the accuracy with which UCAs can be inferred.

By studying an underrepresented population group in southern Africa, this investigation has greatly expanded the repertoire of germline IGHV genes. We further hypothesize that the IGHD and IGHJ germline genes, which make up the remainder of the VH region of functional Abs, are also likely to be highly variable within this population and, thus, warrant investigation. Furthermore, we showed that some of these novel and non-IMGT alleles are functionally active in both HIV and non-HIV Abs. This knowledge will contribute to a better understanding of the Ab response to HIV, as well as other infections, immunizations, and B cell pathologies. Importantly, we also showed that the development of bNAbs against HIV is not restricted by the germline IGHV repertoire, which is significant for vaccine development, because it suggests that everyone has the potential to make Abs capable of neutralizing all strains and subtypes of HIV.

Acknowledgments

We thank the participants of the CAPRISA 002 and 004 studies from whom these samples were obtained and the staff involved in all of the sample collection and processing. We also thank Dr. Mashudu Madzivhandila and Tandile Hermanus for generating the neutralization data and Dr. Nono Mkhize and Dr. Elin Gray for generating mAbs from some of these participants.

Disclosures

The authors have no financial conflicts of interest.

References

1. Hraber, P., M. S. Seaman, R. T. Bailer, J. R. Mascola, D. C. Montefiori, and B. T. Korber. 2014. Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* 28: 163–169.
2. Stamatatos, L., L. Morris, D. R. Burton, and J. R. Mascola. 2009. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat. Med.* 15: 866–870.
3. Kwong, P. D., and J. R. Mascola. 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* 37: 412–425.
4. Scharf, L., J. F. Scheid, J. H. Lee, A. P. West, Jr., C. Chen, H. Gao, P. N. Gnanapragasam, R. Mares, M. S. Seaman, A. B. Ward, et al. 2014. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports* 7: 785–795.
5. Breden, F., C. Lepik, N. S. Longo, M. Montero, P. E. Lipsky, and J. K. Scott. 2011. Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease. *PLoS ONE* 6: e16857.
6. Gorny, M. K., X. H. Wang, C. Williams, B. Volsky, K. Revesz, B. Witover, S. Burda, M. Urbanski, P. Nyambi, C. Krachmarov, et al. 2009. Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Mol. Immunol.* 46: 917–926.
7. Gorny, M. K., J. Sampson, H. Li, X. Jiang, M. Totrov, X. H. Wang, C. Williams, T. O'Neal, B. Volsky, L. Li, et al. 2011. Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PLoS ONE* 6: e27780.
8. Gorny, M. K., R. Pan, C. Williams, X. H. Wang, B. Volsky, T. O'Neal, B. Spurrier, J. M. Sampson, L. Li, M. S. Seaman, et al. 2012. Functional and immunochemical cross-reactivity of V2-specific monoclonal antibodies from HIV-1-infected individuals. *Virology* 427: 198–207.
9. Jardine, J., J.-P. Julien, S. Menis, T. Ota, O. Kalyuzhniy, A. McGuire, D. Sok, P.-S. Huang, S. MacPherson, M. Jones, et al. 2013. Rational HIV immunogen design to target specific germline B cell receptors. *Science* 340: 711–716.
10. Scheid, J. F., H. Mouquet, B. Ueberheide, R. Diskin, F. Klein, T. Y. Oliveira, J. Pietzsch, D. Fenyo, A. Abadir, K. Velinzon, et al. 2011. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333: 1633–1637.
11. West, A. P., Jr., R. Diskin, M. C. Nussenzweig, and P. J. Bjorkman. 2012. Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc. Natl. Acad. Sci. USA* 109: E2083–E2090.
12. Corti, D., A. L. Suguitan, Jr., D. Pinna, C. Silacci, B. M. Fernandez-Rodriguez, F. Vanzetta, C. Santos, C. J. Luke, F. J. Torres-Velez, N. J. Temperton, et al. 2010. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *J. Clin. Invest.* 120: 1663–1673.
13. Lingwood, D., P. M. McTamney, H. M. Yassine, J. R. Whittle, X. Guo, J. C. Boyington, C. J. Wei, and G. J. Nabel. 2012. Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489: 566–570.
14. Lefranc, M. P., and G. Lefranc. 2001. *The Immunoglobulin FactsBook*. Academic Press: London.
15. Watson, C. T., and F. Breden. 2012. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun.* 13: 363–373.
16. Watson, C. T., K. M. Steinberg, J. Huddleston, R. L. Warren, M. Malig, J. Schein, A. J. Willsey, J. B. Joy, J. K. Scott, T. A. Graves, et al. 2013. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am. J. Hum. Genet.* 92: 530–546.
17. van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. Abdool Karim, A. Grobler, N. Barnabas, I. Iriogbe, S. S. Abdool Karim; CAPRISA 002 Acute Infection Study Team. 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS ONE* 3: e1954.
18. Abdool Karim, Q., S. S. Abdool Karim, J. A. Frohlich, A. C. Grobler, C. Baxter, L. E. Mansoor, A. B. Kharsany, S. Sibeko, K. P. Mlisana, Z. Omar, et al; CAPRISA 004 Trial Group. 2010. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* 329: 1168–1174.
19. Gray, E. S., M. C. Madiga, T. Hermanus, P. L. Moore, C. K. Wibmer, N. L. Tumba, L. Werner, K. Mlisana, S. Sibeko, C. Williamson, et al; CAPRISA002 Study Team. 2011. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J. Virol.* 85: 4828–4840.
20. Wang, Y., K. J. Jackson, B. Gäeta, W. Pomat, P. Siba, W. A. Sewell, and A. M. Collins. 2011. Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics* 63: 259–265.
21. Shrestha, R. K., B. Lubinsky, V. B. Bansode, M. B. Moinz, G. P. McCormack, and S. A. Travers. 2014. QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BMC Bioinformatics* 15: 33.
22. Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.
23. Wright, I. A., and S. A. Travers. 2014. RAMICS: trainable, high-speed and biologically relevant alignment of high-throughput sequencing reads to coding DNA. *Nucleic Acids Res.* 42: e106.
24. Boyd, S. D., E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, et al. 2009. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* 1: 12ra23.
25. Jackson, K. J., Y. Liu, K. M. Roskin, J. Glanville, R. A. Hoh, K. Seo, E. L. Marshall, T. C. Gurley, M. A. Moody, B. F. Haynes, et al. 2014. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe* 16: 105–114.
26. Parameswaran, P., Y. Liu, K. M. Roskin, K. K. Jackson, V. P. Dixit, J.-Y. Lee, K. L. Artiles, S. Zompi, M. J. Vargas, B. B. Simen, et al. 2013. Convergent antibody signatures in human dengue. *Cell Host Microbe* 13: 691–700.
27. Wang, C., Y. Liu, L. T. Xu, K. J. Jackson, K. M. Roskin, T. D. Pham, J. Laserson, E. L. Marshall, K. Seo, J.-Y. Lee, et al. 2014. Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *J. Immunol.* 192: 603–611.
28. Wang, Y., K. J. Jackson, J. Davies, Z. Chen, B. A. Gaeta, J. Rimmer, W. A. Sewell, and A. M. Collins. 2014. IgE-associated IGHV genes from venom and peanut allergic individuals lack mutational evidence of antigen selection. *PLoS ONE* 9: e89730.
29. The International Immunogenetics Information System Database. Available at: http://www.imgt.org/. Accessed: June 2014.
30. Immunoglobulin Polymorphism Database. Available at: http://cgi.cse.unsw.edu.au/~ihmmune/IgPdb/. Accessed: June 2014.
31. National Center for Biotechnology Information dbSNP - Short Genetic Variation Database. Available at: http://www.ncbi.nlm.nih.gov/SNP/. Accessed: June 2014.
32. National Center for Biotechnology Information GenBank. Available at: http://www.ncbi.nlm.nih.gov/genbank/. Accessed: June 2014.
33. National Center for Biotechnology Information. Immunoglobulin BLAST Tool. Available at: http://www.ncbi.nlm.nih.gov/igblast/. Accessed: June 2014.
34. Simek, M. D., W. Rida, F. H. Priddy, P. Pung, E. Carrow, D. S. Laufer, J. K. Lehrman, M. Boaz, T. Tarragona-Fiol, G. Miiro, et al. 2009. Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J. Virol.* 83: 7337–7348.
35. Walker, L. M., M. Huber, K. J. Doores, E. Falkowska, R. Pejchal, J.-P. Julien, S.-K. Wang, A. Ramos, P.-Y. Chan-Hui, M. Moyle, et al; Protocol G Principal Investigators. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477: 466–470.
36. Liao, H.-X., R. Lynch, T. Zhou, F. Gao, S. M. Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A. Schramm, Z. Zhang, et al; NISC Comparative Sequencing Program. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496: 469–476.
37. Doria-Rose, N. A., C. A. Schramm, J. Gorman, P. L. Moore, J. N. Bhiman, B. J. DeKosky, M. J. Ernandes, I. S. Georgiev, H. J. Kim, M. Pancera, et al; NISC Comparative Sequencing Program. 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* 509: 55–62.
38. McGuire, A. T., J. A. Glenn, A. Lippy, and L. Stamatatos. 2014. Diverse recombinant HIV-1 Envs fail to activate B cells expressing the germline B cell receptors of the broadly neutralizing anti-HIV-1 antibodies PG9 and 447-52D. *J. Virol.* 88: 2645–2657.

Supplementary Information for Section B, Chapter 3:

Supplementary Table 3.1: Plasma neutralization breadth, viral loads and duration of infection for HIV-positive samples

Shown are all of the BCN individuals, all non-BCN individuals and the two intermediate individuals. Viral loads were determined at 6 months post-infection. The breadth shown is the peak percentage of breadth obtained for each individual during infection. For the BCN individuals the time at peak breadth is the earliest time point that peak breadth was obtained if the same level of breadth remained over time. Whereas the time at peak breadth for the non-BCN individuals is the latest time point at which peak breadth was obtained if the same level of breadth was maintained over-time. Samples with an * are on antiretroviral treatment (ART) or were lost to follow up.

BCN				Non-BCN							
PTID	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**	**PTID**	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**	**PTID**	**Time at Peak Breadth (years)**	**Viral Load (6 months)**	**Peak Breadth (%)**
CAP008*	3	98 400	39%	CAP65*	5	47 300	0%	CAP244	4	47 300	22%
CAP177*	3	32 300	56%	CAP88*	5	21 100	0%	CAP332*	3	21 100	22%
CAP206*	3	156 000	44%	CAP137*	3	1 500	0%	**Median**	3,5	34200	22%
CAP248*	5	1 290	72%	CAP200*	3	349 000	6%				
CAP255*	3	59 500	61%	CAP221*	3	3 830	0%				
CAP256*	3	750 000	83%	CAP225*	5	59 700	6%				
CAP257*	4	16 900	94%	CAP229	5	8 470	6%				
CAP287*	4	39 500	50%	CAP268*	4	15 100	0%				
CAP310*	2	171 000	33%	CAP283*	2	66 300	0%				
CAP312	3	40 400	56%	CAP289*	3	62 300	6%				
CAP314*	2	215 000	44%	CAP290*	2	72 200	6%				
CAP322*	2	92 200	39%	CAP295	4	19 100	11%				
				CAP305*	4	67 200	0%				
Median	**3**	**75 850**	**53%**	**Median**	**4**	**47 300**	**0%**				

Supplementary Table 3.2: List of all glycan components spotted onto the glycan arrays

Glycan Type	Glycan Name	Description	Linker	Core	Sugar Ratio
Blood Group A	2'F-A type 2-Sp - 05	GalNAca1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp	Sp	BSA	5
Blood Group A	2'F-A type 2-Sp - 13	GalNAca1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp	Sp	BSA	13
Blood Group B	2'F-B type 2-Sp - 03	Gala1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp	Sp	BSA	3
Blood Group B	2'F-B type 2-Sp - 07	Gala1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp	Sp	BSA	7
Blood Group B	2'F-B type 2-Sp - 15	Gala1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp	Sp	BSA	15
Blood Group H	2'FucLac (BG-H5)	Fuca1-2Galb1-4Glc-BSA	3 atom	BSA	7
carb-Sia	3'SLacNAc	Sialyla2-3Galb1-4GlcNAc - BSA (3'SLacNAc)	NA	BSA	19
carb-GlcNAc	3'GN type1-Sp - 04	GlcNAcb1-3Galb1-3GlcNAcb-Sp	Sp	BSA	4
carb-GlcNAc	3'GN type1-Sp - 16	GlcNAcb1-3Galb1-3GlcNAcb-Sp	Sp	BSA	16
carb-GlcNAc	3'GN-Di-LacNAc-Sp - 06	GlcNAcb1-3(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	6
carb-GlcNAc	3'GN-Di-LacNAc-Sp - 14	GlcNAcb1-3(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	14
carb-Sia	3'KDNLacNAc-Sp - 05	KDNa2-3Galb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	3'KDNLacNAc-Sp - 12	KDNa2-3Galb1-4GlcNAcb-Sp	Sp	BSA	12
carb-Sia	3'-KDNLeC-Sp - 04	KDNa2-3Galb1-3GlcNAcb-Sp	Sp	BSA	4
carb-Sia	3'-KDNLeC-Sp - 12	KDNa2-3Galb1-3GlcNAcb-Sp	Sp	BSA	12
carb-Sia	3'S(Gc)LacNAc-Sp - 06	Neu5Gca2-3Galb1-4GlcNAcb-Sp	Sp	BSA	6
carb-Sia	3'S(Gc)LacNAc-Sp - 10	Neu5Gca2-3Galb1-4GlcNAcb-Sp	Sp	BSA	10
carb-Sia	3'S(Gc)LeC-Sp - 05	Neu5Gca2-3Galb1-3GlcNAcb-Sp	Sp	BSA	5
carb-Sia	3'S(Gc)LeC-Sp - 12	Neu5Gca2-3Galb1-3GlcNAcb-Sp	Sp	BSA	12
carb-Sia	3'S-Di-LacNAc-Sp - 06	Neu5Aca2-3(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	6
carb-Sia	3'S-Di-LacNAc-Sp - 13	Neu5Aca2-3(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	13
carb-Sia	3'Sia-3-FL	Siaa2-3Galb1-4(Fuca1-3)Glc - BSA	3 atom	BSA	7
carb-Sia	3'SLacNAc-Sp - 05	Neu5Aca2-3Galb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	3'SLacNAc-Sp - 10	Neu5Aca2-3Galb1-4GlcNAcb-Sp	Sp	BSA	10
carb-Sia	3'SLDN-Sp - 05	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	3'SLDN-Sp - 11	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp	Sp	BSA	11
carb-Sia	3'SLeC-Sp - 05	Neu5Aca2-3Galb1-3GlcNAcb-Sp	Sp	BSA	5
carb-Sia	3'SLeC-Sp - 12	Neu5Aca2-3Galb1-3GlcNAcb-Sp	Sp	BSA	12
Lewis	3'-sulpho-LeA	3-SO3-Galb1-3[Fuca1-4)GlcNAc-	3 atom	BSA	15
Lewis	3'-sulpho-LeX	3-SO3-Galb1-4[Fuca1-3)GlcNAc-	3 atom	BSA	15
carb-Sia	6'SLac	Sialyla2-6Galb1-4Glc-APD-HSA	APD	HSA	

		(6'Slac)			
carb-Sia	6'S(Gc)LacNAc-Sp - 05				
carb-Sia	6'S(Gc)LacNAc-Sp - ## (High)	Neu5Gca2-6Galb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	6'S-Di-LacNAc-Sp - 05	Neu5Aca2-6[Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	5
carb-Sia	6'S-Di-LacNAc-Sp - 13	Neu5Aca2-6[Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	13
carb-Sia	6'SLacNAc-Sp - 05	Neu5Aca2-6Galb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	6'SLacNAc-Sp - 11	Neu5Aca2-6Galb1-4GlcNAcb-Sp	Sp	BSA	11
carb-Sia	6'SLDN-Sp - 05	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	6'SLDN-Sp - 13	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp	Sp	BSA	13
Lewis	6'-sulpho-LeA	6-SO3-Galb1-3[Fuca1-4)GlcNAc-	3 atom	BSA	16
Lewis	6'-sulpho-LeX	6-SO3-Galb1-4[Fuca1-3)GlcNAc-	3 atom	BSA	8
carb-Sia	9OAc3'SLacNAc-Sp - 04	Neu5Ac(9Ac)a2-3Galb1-4GlcNAcb-Sp	Sp	BSA	4
carb-Sia	9OAc3'SLacNAc-Sp - 10	Neu5Ac(9Ac)a2-3Galb1-4GlcNAcb-Sp	Sp	BSA	10
carb-Sia	9OAc3'SLeC-Sp - 05	Neu5Ac(9Ac)a2-3Galb1-3GlcNAcb-Sp	Sp	BSA	5
carb-Sia	9OAc3'SLeC-Sp - 12	Neu5Ac(9Ac)a2-3Galb1-3GlcNAcb-Sp	Sp	BSA	12
Blood Group A	A tetra type 1-Sp - 05	GalNAca1-3[Fuca1-2]Galb1-3GlcNAcb-Sp	Sp	BSA	5
Blood Group A	A tetra type 1-Sp - 15	GalNAca1-3[Fuca1-2]Galb1-3GlcNAcb-Sp	Sp	BSA	15
Blood Group A	A tetra type 2-Sp - 05	GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp	Sp	BSA	5
Blood Group A	A tetra type 2-Sp - 07	GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp	Sp	BSA	7
Blood Group A	A tetra type 2-Sp - 17	GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp	Sp	BSA	17
peptide-Tn	Ac-A-Tn(Thr)-S-G - 05	Ac-Ala-(GalNAca)Thr-Ser-Gly-Hex (muc4)	hex	BSA	5
peptide-Tn	Ac-A-Tn(Thr)-S-G - 08	Ac-Ala-(GalNAca)Thr-Ser-Gly-Hex (muc4)	hex	BSA	8
peptide-Tn	Ac-A-Tn(Thr)-S-G - 23	Ac-Ala-(GalNAca)Thr-Ser-Gly-Hex (muc4)	hex	BSA	23
peptide-Tn	Ac-G-V-Tn(Thr)-S-A-G - 04	Ac-Gly-Val-(GalNAca)Thr-Ser-Ala-Gly-Hex (muc1)	hex	BSA	4
peptide-Tn	Ac-G-V-Tn(Thr)-S-A-G - 21	Ac-Gly-Val-(GalNAca)Thr-Ser-Ala-Gly-Hex (muc1)	hex	BSA	21
peptide-Tn	Ac-P-Tn(Thr)-T-G - 05	Ac-Pro-(GalNAca)Thr-Thr-Gly-Hex (muc2)	hex	BSA	5
peptide-Tn	Ac-P-Tn(Thr)-T-G - 08	Ac-Pro-(GalNAca)Thr-Thr-Gly-Hex (muc2)	hex	BSA	8
peptide-Tn	Ac-P-Tn(Thr)-T-G - 22	Ac-Pro-(GalNAca)Thr-Thr-Gly-Hex (muc2)	hex	BSA	22
peptide-GlcNAca	Ac-S-Ser(GlcNAca)-S-G - 07	AcSer-(GlcNAca)Ser-Ser-Gly-Hex-BSA	hex	BSA	7
peptide-GlcNAca	Ac-S-Ser(GlcNAca)-S-G - 24	AcSer-(GlcNAca)Ser-Ser-Gly-Hex-BSA	hex	BSA	24
peptide-GlcNAcb	Ac-S-Ser(GlcNAcB)-S-G - 07	AcSer-(GlcNAcb)Ser-Ser-Gly-Hex-BSA	hex	BSA	7

peptide-GlcNAcb	Ac-S-Ser(GlcNAcB)-S-G - 24	AcSer-(GlcNAcb)Ser-Ser-Gly-Hex-BSA	hex	BSA	24
peptide	Ac-S-S-S-G	SerSerSerGly-BSA	hex	BSA	24
peptide-TF	Ac-S-TF(Ser)-S-G - 04	AcSer-(Galb1-3GalNAca)Ser-Ser-Gly-Hex-BSA (S-TF-S)	hex	BSA	4
peptide-TF	Ac-S-TF(Ser)-S-G - 16	AcSer-(Galb1-3GalNAca)Ser-Ser-Gly-Hex-BSA (S-TF-S)	hex	BSA	16
peptide-TF	Ac-S-TF(Ser)-S-G - 28	AcSer-(Galb1-3GalNAca)Ser-Ser-Gly-Hex-BSA (S-TF-S)	hex	BSA	28
peptide	Ac-S-Thr-S-G - 18	Ac-Ser-Thr-Ser-Gly-Hex	hex	BSA	18
peptide-Tn	Ac-S-Tn(Ser)-S-G - 04	AcSer-(GalNAca)Ser-Ser-Gly-Hex-BSA (STnS)	hex	BSA	4
peptide-Tn	Ac-S-Tn(Ser)-S-G - 22	AcSer-(GalNAca)Ser-Ser-Gly-Hex-BSA (STnS)	hex	BSA	22
peptide-Tn	Ac-S-Tn(Ser)-S-G - 33	AcSer-(GalNAca)Ser-Ser-Gly-Hex-BSA (STnS)	hex	BSA	33
peptide-Tn	Ac-S-Tn(Thr)-A-G - 04	Ac-Ser-(GalNAca)Thr-Ala-Gly-Hex (muc1)	hex	BSA	4
peptide-Tn	Ac-S-Tn(Thr)-A-G - 08	Ac-Ser-(GalNAca)Thr-Ala-Gly-Hex (muc1)	hex	BSA	8
peptide-Tn	Ac-S-Tn(Thr)-A-G - 22	Ac-Ser-(GalNAca)Thr-Ala-Gly-Hex (muc1)	hex	BSA	22
peptide-Tn	Ac-S-Tn(Thr)-G-G - 03	Ac-Ser-(GalNAca)Thr-Gly-Gly-Hex (muc4)	hex	BSA	3
peptide-Tn	Ac-S-Tn(Thr)-G-G - 07	Ac-Ser-(GalNAca)Thr-Gly-Gly-Hex (muc4)	hex	BSA	7
peptide-Tn	Ac-S-Tn(Thr)-G-G - 19	Ac-Ser-(GalNAca)Thr-Gly-Gly-Hex (muc4)	hex	BSA	19
peptide-Tn	Ac-S-Tn(Thr)-S-G - 04	AcSer-(GalNAca)Thr-Ser-Gly-Hex-BSA (STnS)	hex	BSA	4
peptide-Tn	Ac-S-Tn(Thr)-S-G - 24	AcSer-(GalNAca)Thr-Ser-Gly-Hex-BSA (STnS)	hex	BSA	24
peptide-Tn	Ac-S-Tn(Thr)-S-G HSA-04	AcSer-(GalNAca)Thr-Ser-Gly-Hex-HSA (STnS)	hex	HSA	4
peptide-Tn	Ac-S-Tn(Thr)-S-G HSA-23	AcSer-(GalNAca)Thr-Ser-Gly-Hex-HSA (STnS)	hex	HSA	23
peptide-Tn	Ac-S-Tn(Thr)-Tn(Thr)-G - 05	Ac-Ser-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	5
peptide-Tn	Ac-S-Tn(Thr)-Tn(Thr)-G - 09	Ac-Ser-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	9
peptide-Tn	Ac-S-Tn(Thr)-Tn(Thr)-G - 22	Ac-Ser-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	22
peptide-Tn	Ac-S-Tn(Thr)-V-G - 04	Ac-Ser-(GalNAca)Thr-Val-Gly-Hex	hex	BSA	4
peptide-Tn	Ac-S-Tn(Thr)-V-G - 22	Ac-Ser-(GalNAca)Thr-Val-Gly-Hex	hex	BSA	22
peptide-TF	Ac-TF(Ser)-G - 04	Ac(Galb1-3GalNAca)Ser-Gly-Hex-BSA	hex	BSA	4
peptide-TF	Ac-TF(Ser)-G - 24	Ac(Galb1-3GalNAca)Ser-Gly-Hex-BSA	hex	BSA	24
peptide-Tn	Ac-Tn(Ser)Tn(Ser)Tn(Ser)-G - 03	Ac(GalNAca)Ser-(GalNAca)Ser-(GalNAca)Ser-Gly-Hex-BSA (Tn3)	hex	BSA	3
peptide-Tn	Ac-Tn(Ser)-Tn(Ser)-Tn(Ser)-G - 16	Ac(GalNAca)Ser-(GalNAca)Ser-(GalNAca)Ser-Gly-Hex-BSA (Tn3)	hex	BSA	16
peptide-Tn	Ac-Tn(Ser)Tn(Ser)Tn(Ser)-G - 27	Ac(GalNAca)Ser-(GalNAca)Ser-(GalNAca)Ser-Gly-Hex-BSA (Tn3)	hex	BSA	27

peptide-Tn	Ac-Tn(Thr)-G - 21	Ac(GalNAca)Thr-Gly-Hex-BSA	hex	BSA	21
peptide-Tn	Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 05	Ac-(GalNAca)Thr-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	5
peptide-Tn	Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 08	Ac-(GalNAca)Thr-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	8
peptide-Tn	Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 20	Ac-(GalNAca)Thr-(GalNAca)Thr-(GalNAca)Thr-Gly-Hex (muc2)	hex	BSA	20
peptide-Tn	Ac-T-Tn(Thr)-P-G - 04	Ac-Thr-(GalNAca)Thr-Pro-Gly-Hex (muc2,6,7)	hex	BSA	4
peptide-Tn	Ac-T-Tn(Thr)-P-G - 08	Ac-Thr-(GalNAca)Thr-Pro-Gly-Hex (muc2,6,7)	hex	BSA	8
peptide-Tn	Ac-T-Tn(Thr)-P-G - 21	Ac-Thr-(GalNAca)Thr-Pro-Gly-Hex (muc2,6,7)	hex	BSA	21
peptide-Tn	Ac-V-Tn(Thr)-S-G - 04	Ac-Val-(GalNAca)Thr-Ser-Gly-Hex (muc1)	hex	BSA	4
peptide-Tn	Ac-V-Tn(Thr)-S-G - 08	Ac-Val-(GalNAca)Thr-Ser-Gly-Hex (muc1)	hex	BSA	8
peptide-Tn	Ac-V-Tn(Thr)-S-G - 19	Ac-Val-(GalNAca)Thr-Ser-Gly-Hex (muc1)	hex	BSA	19
Blood Group A	Adi - 04	GalNAca1-3Galb-BSA (Adi)	MEAG	BSA	4
Blood Group A	Adi - 17	GalNAca1-3Galb-BSA (Adi)	MEAG	BSA	17
glycoprotein	AGE60	Advanced glycation endproducts day 60 (AGE60)	na	BSA	na
Blood Group A	A-LeB hexa	GalNAca1-3(Fuca1-2)Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-	(Glc)	BSA	6
glycoprotein	Alpha-1-acid glycoprotein	alpha1 Acid Glycoprotein	na		na
glycoprotein	Alpha-fetoprotein	alpha fetoprotein (AFP)	na		na
non-human-aGal	alphaGal	Gala1-3Galb1-4GlcNAc-BSA (alphaGal)	14 atom spacer	BSA	8
non-human	alpha-Gal tetra - 04	Gala1-3Galb1-4GlcNAcb1-3Galb1-	(Glc)	BSA	4
non-human	alpha-Gal tetra - 17	Gala1-3Galb1-4GlcNAcb1-3Galb1-	(Glc)	BSA	17
non-human-aGal	alphaGal-6-deoxy	Gala1-3Galb1-4(6deoxy-GlcNAc)-HSA (alphaGal)	3 atom spacer	HSA	11
non-human	Ara5	Araa1-5Araa1-5Araa1-5Araa1-5Araa1-BSA (Ara5)	(Ara)	BSA	20
Blood Group B	B tetra type 1-Sp - 04	Gala1-3[Fuca1-2]Galb1-3GlcNAcb-Sp	Sp	BSA	4
Blood Group B	B tetra type 1-Sp - 16	Gala1-3[Fuca1-2]Galb1-3GlcNAcb-Sp	Sp	BSA	16
Blood Group B	B tetra type 2-Sp - 05	Gala1-3[Fuca1-2]Galb1-3GlcNAcb-Sp	Sp	BSA	5
Blood Group B	B tetra type 2-Sp - 07	Gala1-3[Fuca1-2]Galb1-4GlcNAcb-Sp	Sp	BSA	7
Blood Group B	B tetra type 2-Sp - 20	Gala1-3[Fuca1-2]Galb1-4GlcNAcb-Sp	Sp	BSA	20
non-human	B6 di - 06	Gala1-3Galb‰ÛÒ BSA (Bdi)	(Glc)	BSA	6
non-human	B6 di - 16	Gala1-3Galb‰ÛÒ BSA (Bdi)	(Glc)	BSA	16
non-human	Bdi	Gala1-3Gal‰ÛÒ BSA (Bdi)	14 atom spacer	BSA	23
Blood Group A	BG-A	GalNAca1-3(Fuca1-2)Galb•Û_ - BSA [BG-A]	6 atom spacer	BSA	19
Blood Group A	BG-A1	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3Galb1-4(Glc)-APD-HSA (BG-A1)	APD	HSA	5

Blood Group B	BG-B (Dextra)	Gala1-3(Fuca1-2)Galb-BSA [BG-B] from Dextra	6 atom spacer	BSA	13
Blood Group H	BG-H1	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb‰ÛÒAPD-HSA [BG-H1]	APD-	HSA	20
Blood Group H	BG-H2	Fuca1-2Galb1-4GlcNAcb-HSA (BG-H2)	unknown	HSA	16
control	BSA	Bovine serum albumin	na	BSA	na
control	BSA - C5 (Alkyne) - 10	DF-168B-175-1	C5	BSA	
control	BSA - C5 (Alkyne) - 23	DF-168C-16-B5	C5	BSA	
control	BSA-#2		na		na
control	Triazole linker from Xuefei	BSA-linker-triazole from Xuefei	adipic acid	BSA	43
glycoprotein	BSM	Bovine submaxillary mucin (STn, STF, S-GlcNAcb1-3, ~20% of Sia is acetylated at 7,8, or 9)	na	BSM	na
glycoprotein	BSM (asialo)	Asialo-Bovine submaxillary mucin (aBSM, Tn, TF, GlcNAcb1-3GalNAc)	na	BSM	na
glycoprotein	BSM (deacetylated)	Deacetylated-Bovine submaxillary mucin	na	BSM	na
glycoprotein	BSM (ox)	periodate oxidized bovine submaxillary mucin	na		na
glycoprotein	CEA	carcinoembryonic antigen (CEA)	na		na
non-human	Cellobiose	Glcb1-4Glcb-BSA (Cellobiose)	(Glc)	BSA	13
non-human	Cellotriose	Glcb1-4Glcb1-4Glcb-BSA (Cello3)	(Glc)	BSA	13
non-human	Chito 3	GlcNAcb1-4GlcNAcb1-4GlcNAcb-BSA (Chito 3)	(GlcNAc)	BSA	8
non-human	Chito 3 - 20	GlcNAcb1-4GlcNAcb1-4GlcNAcb-BSA (Chito 3)	(GlcNAc)	BSA	20
carb-Sia	CT/Sda-Sp - 05	Neu5Aca2-3[GalNAcb1-4]Galb1-4GlcNAcb-Sp	Sp	BSA	5
carb-Sia	CT/Sda-Sp - 13	Neu5Aca2-3[GalNAcb1-4]Galb1-4GlcNAcb-Sp	Sp	BSA	13
control	Cy3	Cy3-BSA (20mg/mL + BSA, 125mg/mL total)	na	BSA	
control	Cy5	Cy5-BSA (30mg/mL+ BSA, 125mg/mL total)	na	BSA	
Lewis	DFLNH(c)	Galb1-4GlcNAcb1-6[Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3]Galb1-BSA	(Glc)	BSA	8
Lewis	DFLNnH	Galb1-4(Fuca1-3)GlcNAcb1-6[Galb1-4(Fuca1-3)GlcNAcb1-3]Galb1-BSA	(Glc)	BSA	10
Lewis	DFpLNH I	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb1-3Galb-BSA	(Glc)	BSA	9
carb-type 2	Di-LacNAc-Sp - 06	(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	6
carb-type 2	Di-LacNAc-Sp - 16	(Galb1-4GlcNAcb1-3)2b-Sp	Sp	BSA	16
Lewis	Di-LeC-Sp - 06	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp	Sp	BSA	6
Lewis	Di-LeC-Sp - 16	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp	Sp	BSA	16
carb-Sia	DSLNT	Siaa2-3Galb1-3(Siaa2-6)GlcNAcb1-3Galb1-BSA (DSLNT)	(Glc)	BSA	5.6
glycoprotein	FABP	Fatty Acid Binding Protein (FABP)	na		na

glycoprotein	fetuin	fetuin (Sia2-3LacNAc, Sia2-6LacNAc, SiaLeC, STF)	na		na
glycoprotein	fetuin (asialo)	asialofetuin (Galb1-4GlcNAc, Galb1-3GlcNAc, Galb1-3GalNAc			
glycoprotein	Fetuin (ox)	periodate oxidized fetuin	na		na
non-human	Forssman Di - 04	GalNAca1-3GalNAcb1-BSA	MEAG	BSA	4
non-human	Forssman Di - 21	GalNAca1-3GalNAcb1-BSA	MEAG	BSA	21
non-human	Forssman Di - 31	GalNAca1-3GalNAcb1-BSA	MEAG	BSA	31
non-human	Forssman Tetra-BSA - 05	GalNAca1-3GalNAcb1-3Gala1-4Galb-BSA	(Glc)	BSA	5
non-human	Forssman Tetra-BSA - 13	GalNAca1-3GalNAcb1-3Gala1-4Galb-BSA	(Glc)	BSA	13
Lewis	Fuc, Sia-LNnH-APD-HSA	Galb1-4[Fuca1-3]GlcNAcb1-6[Neu5Aca2-3Galb1-4GlcNAcb1-3]Galb1-APD-HSA	APD	HSA	12 to 15
carb-Fuc	Fuc-a - 04	Fuc-a - BSA	MEAG	BSA	4
carb-Fuc	Fuc-a - 22	Fuc-a - BSA	MEAG	BSA	22
non-human	Fuc-b - 04	Fuc-b - BSA	MEAG	BSA	4
non-human	Fuc-b - 22	Fuc-b - BSA	MEAG	BSA	22
glycolipid	Fuc-GM1a - 08	Fuca1-2Galb1-3GalNAcb1-4(Siaa2-3)Galb1-4	(Glc)	BSA	8
non-human	G2M4	Manb1-4(Gala1-6)Manb1-4(Gala1-6)Manb1-4Manb1-BSA (G2M4)	(man)	BSA	7
glycolipid	GA1 - 06	Galb1-3GalNAcb1-4Galb1-BSA (GA1tri or asialo-GM1)	(Glc)		6
glycolipid	GA1 - 20	Galb1-3GalNAcb1-4Galb1-BSA (GA1tri or asialo-GM1)	(Glc)	BSA	20
glycolipid	GA1di	Galb1-3GalNAcb ‰ÛÒ HSA (GA1di)	3 atom spacer)	HSA	11
glycolipid	GA2di - 04	GalNAcb1-4Galb - BSA (GA2di or asialo-GM2)	MEAG	BSA	4
glycolipid	GA2di - 16	GalNAcb1-4Galb - BSA (GA2di or asialo-GM2)	MEAG	BSA	16
glycolipid	GA2di - 37	GalNAcb1-4Galb - BSA (GA2di or asialo-GM2	MEAG	BSA	37
glycolipid	GA2di (accurate)	GalNAcb1-4Galb - BSA (GA2di or asialo-GM2)	unknown	BSA	28
non-human-aGal	Gal3	Gala1-3Galb1-4Gala-BSA (Gal3)	(1-3Gal)	BSA	7
carb-Gal	Gal-a	Gal-a - BSA	MEAG	BSA	24
non-human	Gala1-2Gal	Gala1-2Gal	3 atom	BSA	13
carb-Gal	Gala1-4Galb	Gala1-4Galb-CETE-BSA	CETE	BSA	11
non-human-aGal	Gala3-type1	Gala1-3Galb1-3GlcNAc-BSA	3 atom	BSA	9
carb-Gal	Gal-b	Gal-b - BSA	MEAG	BSA	21
non-human	Galb1-4Gal	Galb1-4Gal-BSA	3 atom spacer	BSA	
non-human	Galb1-6Man-a	Galb1-6Man-a - BSA	MEAG	BSA	13
non-human-aGal	Galilli	Gala1-3Galb1-4Glc-BSA	3 atom	BSA	21
carb-GalNAc	GalNAc-a - 04	GalNAc-a - BSA	MEAG		4
carb-GalNAc	GalNAc-a - 22	GalNAc-a - BSA	MEAG	BSA	22
carb-GalNAc	GalNAca1-6Galb - 04	GalNAca1-6Galb-BSA	MEAG		4

carb-GalNAc	GalNAca1-6Galb - 22	GalNAca1-6Galb-BSA	MEAG	BSA	22
carb-GalNAc	GalNAc-b	GalNAc-b - BSA	MEAG	BSA	21
glycolipid	Gb4	GalNAcb1-3Gala1-4Galb1-BSA (Gb4 or P antigen)	(Glc)	BSA	9
glycolipid	Gb4 tetra (P1 tetra)-Sp - 06	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp	Sp	BSA	6
glycolipid	Gb4 tetra (P1 tetra)-Sp - 15	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp	Sp	BSA	15
glycolipid	Gb5/SSEA3 - 04	Galb1-3GalNAcb1-3Gala1-4Galb1-	(Glc)	BSA	4
glycolipid	Gb5/SSEA3 - 12	Galb1-3GalNAcb1-3Gala1-4Galb1-	(Glc)	BSA	12
glycolipid	GD1a-Sp - 05	Neu5Aca2-3[Neu5Aca2-3Galb1-3GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	5
glycolipid	GD1a-Sp - 10	Neu5Aca2-3[Neu5Aca2-3Galb1-3GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	10
glycolipid	GD1b	Siaa2-8Siaa2-3(Galb1-3GalNAcb1-4)Galb1-4-BSA	(Glc)	BSA	5
glycolipid	GD2-Sp - 04	Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	4
glycolipid	GD2-Sp - 10	Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	10
glycolipid	GD3-Sp - 04	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	4
glycolipid	GD3-Sp - 08	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	8
carb-Glc	Glc-a	Glc-a - BSA	MEAG	BSA	22
carb-Glc	Glc-a - 05	Glc-a - BSA	MEAG	BSA	5
carb-Glc	Glca1-6Glca1-4Glca1-4Glcb	Glca1-6Glca1-4Glca1-4Glcb-CETE-BSA	CETE	BSA	15
carb-Glc	Glc-b	Glc-b - BSA	MEAG	BSA	23
carb-GlcNAc	GlcNAca1-4Galb - 03	GlcNAca1-4Galb-BSA	MEAG	BSA	3
carb-GlcNAc	GlcNAca1-4Galb - 20	GlcNAca1-4Galb-BSA	MEAG	BSA	20
carb-GlcNAc	GlcNAc-b	GlcNAc-b - BSA	MEAG	BSA	21
N-linked	GlcNAc-Man3	Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb-BSA	(GlcNAc)	BSA	2
N-linked	GlcNAc-Man5	Mana1-6(Mana1-3)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb-BSA	(GlcNAc)	BSA	3
Blood Group A	Globo A	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-BSA	(Glc)	BSA	9
Blood Group H	Globo H	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-BSA	(Glc)	BSA	10
glycoprotein	glycophorin (asialo)	asialo-glycophorin A (aGn)	na	Gn	na
glycoprotein	Glycophorin A	Glycophorin A (Gn)	na	Gn	na
glycolipid	GM1a	Galb1-3GalNAcb1-4(Siaa2-3)Galb1-4(Glc)HSA	(Glc)	HSA	29
glycolipid	GM2-Sp - 04	Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	4
glycolipid	GM2-Sp - 07	Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	7
glycolipid	GM2-Sp - 14	Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	14
glycolipid	GM3	Sialyla2-3Galb1-4Glc-APD-HSA	APD	HSA	15-Dec

glycolipid	GM3(Gc)-Sp - 05	Neu5Gca2-3Galb1-4Glcb-Sp	Sp	BSA	5
glycolipid	GM3(Gc)-Sp - 14	Neu5Gca2-3Galb1-4Glcb-Sp	Sp	BSA	14
glycolipid	GM3-Sp - 04	Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	4
glycolipid	GM3-Sp - 11	Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	11
carb-GlcNAc	GNLacNAc-Sp - 06	GlcNAcb1-3Galb1-4GlcNAcb-Sp	Sp	BSA	6
carb-GlcNAc	GNLacNAc-Sp - 16	GlcNAcb1-3Galb1-4GlcNAcb-Sp	Sp	BSA	16
glycolipid	GQ2-Sp - 03	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	3
glycolipid	GQ2-Sp - 06	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	6
glycolipid	GT2-Sp - 03	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	3
glycolipid	GT2-Sp - 08	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3[GalNAcb1-4]Galb1-4Glcb-Sp	Sp	BSA	8
glycolipid	GT3-Sp - 03	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	3
glycolipid	GT3-Sp - 07	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp	Sp	BSA	7
GAG	Hep-5000	heparin polysaccharide (MW ~5000)	na	BSA	1
GAG	Hep-N-acetylated	fully N-acetylated heparin polysaccharide	na	BSA	1
control	HSA	Human serum albumin (isolated from serum)	na	HSA	na
control	HSA (recomb)	human serum albumin (recombinant)	na	HSA	na
glycoprotein	hsp90	Heat Shock Protein 90 (hsp90)	na		na
GAG	Hya8	(GlcNAcb1-4GlcAb1-3)4b1-	(GlcNAc)	BSA	3
GAG	Hya9	(GlcAb1-3GlcNAcb1-4)4b1-3GlcAb1-	(GlcNAc)	BSA	3
N-linked	Hybrid-M5N4B	GlcNAcb1-2Mana1-3[Mana1-3(Mana1-6)Mana1-6](GlcNAcb1-4)Manb1-4GlcNAcb1-	(GlcNAc)	BSA	3
carb-type 1	iLNO	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6 (Galb1-3GlcNAcb1-3)Galb1-	(Glc)	BSA	6
carb-Glc	Isomaltose	Glca1-6Glcb-BSA	MEAG	BSA	13
glycoprotein	KLH	Keyhole limpet hemocyanin	na	KLH	na
glycoprotein	KLH (oxidized)	periodate oxidized Keyhole limpet hemocyanin	na	KLH	na
carb-type 2	LacNAc	Galb1-4GlcNAc ‰ÛÒ BSA (LacNAc)	14 atom spacer	BSA	22
carb-type 2	LacNAc (trimeric)	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-APE-HSA (TriLacNAc)	APE	HSA	8
N-linked	LacNAc-Man5	Mana1-6(Mana1-3)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb-BSA	(GlcNAc)	BSA	2
carb-type 2	LacNAc-Sp - 06	Galb1-4GlcNAcb-Sp	Sp	BSA	6
carb-type 2	LacNAc-Sp - 15	Galb1-4GlcNAcb-Sp	Sp	BSA	15
carb-Gal	Lactose	Galb1-4Glcb ‰ÛÒ BSA (Lac)	unknown	BSA	33
carb-Gal	Lactose-C5 - 05	Galb1-4Glcb ‰ÛÒ BSA (Lac)	C5	BSA	5

carb-Gal	Lactose-C5 - 14	Galb1-4Glcb ‰ÛÒ BSA (Lac)	C5	BSA	14
carb-GalNAc	LDN-Sp - 05	GalNAcb1-4GlcNAcb-Sp	Sp	BSA	5
carb-GalNAc	LDN-Sp - 14	GalNAcb1-4GlcNAcb-Sp	Sp	BSA	14
Lewis	LeA	Galb1-3[Fuca1-4)GlcNAcb1-3Galb1-4Glcb- BSA (Lea)	3 atom spacer	BSA	18
Lewis	LeA-LeX	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-APD-HSA	APD	HSA	21
Lewis	LeB	Fuca1-2Galb1-3[Fuca1-4)GlcNAcb1-3Galb1-4Glcb-BSA (Leb)	3 atom spacer	BSA	9
Lewis	LeC-Sp - 06	Galb1-3GlcNAcb-Sp	Sp	BSA	6
Lewis	LeC-Sp - 07	Galb1-3GlcNAcb-Sp	Sp	BSA	7
Lewis	LeC-Sp - 15	Galb1-3GlcNAcb-Sp	Sp	BSA	15
Lewis	LeX (dimeric)	Galb1-4[Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-APE-BSA	APE	BSA	7
Lewis	LeX (monomeric)	Galb1-4[Fuca1-3)GlcNAc-APD-HSA (Lex)	APD	HSA	x
Lewis	LeY	Fuca1-2Galb1-4[Fuca1-3)GlcNAc ‰ÛÒHSA (Ley)	unknown	HSA	8
carb-type 1+2	LNH - 13	Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-BSA	(Glc)	BSA	13
carb-type 2	LNnH - 11	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-BSA	(Glc)	BSA	11
carb-type 2	LNnT - 04	Galb1-4GlcNAcb1-3Galb1-BSA (LNnT)	(Glc)	BSA	4
carb-type 2	LNnT - 14	Galb1-4GlcNAcb1-3Galb1-BSA (LNnT)	(Glc)	BSA	14
carb-type 1	LNT - 05	Galb1-3GlcNAcb1-3Galb-BSA (LNT)	(Glc)	BSA	5
carb-type 1	LNT - 21	Galb1-3GlcNAcb1-3Galb-BSA (LNT)	(Glc)	BSA	21
carb-GlcNAc	LNT-2-Sp - 06	GlcNAcb1-3Galb1-4Glcb-Sp	Sp	BSA	6
carb-GlcNAc	LNT-2-Sp - 15	GlcNAcb1-3Galb1-4Glcb-Sp	Sp	BSA	15
carb-type 1	LNT-Sp - 06	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp	Sp	BSA	6
carb-type 1	LNT-Sp - 15	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp	Sp	BSA	15
carb-Sia	LSTa	Siaa2-3Galb1-3GlcNAcb1-3Galb1-BSA (LSTa)	(Glc)	BSA	10
carb-Sia	LSTb	Galb1-3(Siaa2-6)GlcNAcb1-3Galb1-BSA (LSTb)	(Glc)	BSA	11
carb-Sia	LSTc	Siaa2-6Galb1-3GlcNAcb1-3Galb1-BSA (LSTc)	(Glc)	BSA	7
carb-Glc	Maltopentaose	Glca1-4Glca1-4Glca1-4Glca1-4Glca-BSA (Malto5)	(Glc)	BSA	11
carb-Glc	Maltose	Glca1-4Glcb-BSA (Maltose)	MEAG	BSA	23
N-linked	Man1 - 04	Manβ1-4GlcNAcβ1-4GlcNAcβ1-	4HB	BSA	4
N-linked	Man1 - 12	Manβ1-4GlcNAcβ1-4GlcNAcβ1-	4HB	BSA	12
N-linked	Man3	Manα1-6(Manα1-3)Manβ1-4GlcNAc -BSA (Man3)	(GlcNAc)	BSA	5
N-linked	Man5	Manα1-6(Manα1-3)Manα1-6(Manα1-3)Manβ1-4GlcNAc-BSA (Man5)	(GlcNAc)	BSA	5
N-linked	Man6 - I	Manα1-6(Manα1-3)Manα1-6(Manα1-2Manα1-3)Manβ1-BSA	(GlcNAc)	BSA	4
N-linked	Man6 - II	Manα1-2Manα1-3Manα1-6(Manα1-2Manα1-3)Manβ1-BSA	(GlcNAc)	BSA	5

		(Man6 - II)			
N-linked	Man7D1	Manα1-6(Manα1-3)Manα1-6(Manα1-2Manα1-2Manα1-3)Manβ1-4GlcNAc-BSA	(GlcNAc)	BSA	10
N-linked	Man7D3	Manα1-2Manα1-6(Manα1-3)Manα1-6(Manα1-2Manα1-3)Manβ1-4GlcNAc-BSA	(GlcNAc)	BSA	8
N-linked	Man8D1D3	Manα1-2Manα1-6(Manα1-3)Manα1-6(Manα1-2Manα1-2Manα1-3)Manβ1-4GlcNAc-BSA	(GlcNAc)	BSA	9
N-linked	Man9	Manα1-2Manα1-6(Manα1-2Manα1-3)Manα1-6(Manα1-2Manα1-2Manα1-3)Manβ1-4GlcNAc-BSA	(GlcNAc)	BSA	5
carb-Man	Man-a	Man-α - BSA	MEAG	BSA	20
carb-Man	Man-a - 05	Man-α - BSA	MEAG	BSA	5
carb-Man	Mana1-6Man-a	Manα1-6Man-α - BSA	MEAG	BSA	15
carb-Man	Mana1-6Man-a - 04	Manα1-6Man-α- BSA	MEAG	BSA	4
non-human	Manb4 - 04	Manβ1-4Manβ1-4Manβ1-4Manβ1-BSA (Manb4)	(Man)	BSA	14
non-human	Manb4				
carb-Man	ManT	Manα1-6[Manα1-3]Manβ-BSA [ManT]	14 atom spacer	BSA	26
carb-type 1	MFiLNO(1-3)	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6 (Galb1-3GlcNAcb1-3)Galb1-	(Glc)	BSA	9
Lewis	MFLNH I	Galb1-4GlcNAcb1-6 (Fuca1-2Galb1-3GlcNAcb1-3)Galb1-	(Glc)	BSA	11
Lewis	MFLNH III	Galb1-4(Fuca1-3)GlcNAcb1-6 (Galb1-3GlcNAcb1-3)Galb1-	(Glc)	BSA	14
Lewis	MSMFLNH I	Siaa2-6Galb1-4GlcNAcb1-6 (Fuca1-2Galb1-3GlcNAcb1-3)Galb1-	(Glc)	BSA	11
Lewis	MSMFLnNH	Galb1-4(Fuca1-3)GlcNAcb1-6 (Siaa1-3Galb1-4GlcNAcb1-3)Galb1-	(Glc)	BSA	9
N-linked	NA2	Galb1-4GlcNAcb1-2Mana1-6[Galb1-4GlcNAcb1-2Mana1-3]Manb1-4GlcNAc -BSA (NA2)	(GlcNAc)	BSA	8
N-linked	NA3	Galb1-4GlcNAcb1-2Mana1-6[Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3]Manb1-4GlcNAc -BSA (NA3)	(GlcNAc)	BSA	5
N-linked	NA4	Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-6)Mana1-6[Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3]Manb1-4GlcNAc -BSA (NA4)	(GlcNAc)	BSA	5
N-linked	NGA2	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAc -BSA (NGA2)	(GlcNAc)	BSA	7
N-linked	NGA2B	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)(GlcNAcb1-4)Manb1-4GlcNAc -BSA (NGA2B)	(GlcNAc)	BSA	5
N-linked	NGA3	GlcNAcb1-2Mana1-6[GlcNAcb1-2(GlcNAcb1-4)Mana1-3]Manb1-4GlcNAc -BSA (NGA3)	(GlcNAc)	BSA	1

N-linked	NGA3B	GlcNAcb1-2Mana1-6[GlcNAcb1-2(GlcNAcb1-4)Mana1-3](GlcNAcb1-4)Manb1-4GlcNAc -BSA (NA3)	(GlcNAc)	BSA	6
N-linked	NGA4	GlcNAcb1-2(GlcNAcb1-6)Mana1-6[GlcNAcb1-2(GlcNAcb1-4)Mana1-3]Manb1-4GlcNAc -BSA	(GlcNAc)	BSA	6
N-linked	NGA4(B)2	GlcNAcb1-2(GlcNAcb1-4)(GlcNAcb1-6)Mana1-6[GlcNAcb1-2Mana1-3](GlcNAcb1-4)Manb1-4GlcNAc -BSA [NGA4(B)2]	(GlcNAc)	BSA	4
N-linked	NGA5B	GlcNAcb1-2(GlcNAcb1-4)(GlcNAcb1-6)Mana1-6[GlcNAcb1-2(GlcNAcb1-4)Mana1-3](GlcNAcb1-4)Manb1-4GlcNAc -BSA (NGA5B)	(GlcNAc)	BSA	2
glycoprotein	OSM	Ovine submaxillary mucin (94% STn, 4% TF, 2% Fuca1-2Galb1-3GalNAc)	na	OSM	na
glycoprotein	OSM (asialo)	asialo-Ovine submaxillary mucin (aOSM)	na	OSM	na
glycoprotein	OSM (ox)	periodate oxidized ovine submaxillary mucin	na		na
glycoprotein	ovalbumin	ovalbumin (56% Man5+Man6)	na		na
glycoprotein	Ovalbumin (ox)	periodate oxidized ovalbumin	na		na
glycolipid	P1	Gala1-4Galb1-4GlcNAc-BSA (P1)		BSA	9
control	PEG-linker	OH-(CH2)2-NH-Gly-CO-PEG7-NH-(CO)Hept-SH-Mal-Cychex-CO-BSA	PEG-linker	BSA	6
glycolipid	Pk or Gb3	Gala1-4Galb1-4Glc-HSA [Pk or Gb3 or CD77]		HSA	13
carb-type 1	pLNH - 07	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-BSA (pLNH)	(Glc)	BSA	7
carb-type 1	pLNH - 21	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-BSA (pLNH)	(Glc)	BSA	21
glycoprotein	PSA	Prostate Specific Antigen (PSA)			
non-human	Rha-a	Rha-a ‰ÛÒ BSA	MEAG	BSA	18
non-human	Rha-b	Rha-b - BSA	MEAG	BSA	21
peptide-Tn	R-Tn(Ser)-Tn-hydroxyethylamide	BSA-linker-Tn(Ser)-hydroxyethylamide	adipic acid	BSA	36
Blood Group H	Sia-LNF V	Fuca1-2Galb1-3(Neu5Aca2-6)GlcNAcb1-3Galb1-APD-HSA	APD	HSA	12 to 15
carb-Sia	Sia-LNnT	Siaa2-3Galb1-4GlcNAcb1-3Galb1-APD-HSA	APD	HSA	9
Lewis	Sialyl LeA	Siaa2-3Galb1-3[Fuca1-4)GlcNAcb1-3Galb1-APD-HSA (SLeA)	APD	HSA	12
Lewis	Sialyl LeX	Sialyla2-3Galb1-4[Fuca1-3)GlcNAc ‰ÛÒ BSA	14 atom spacer	BSA	9
glycolipid	SSEA-4-Sp - 05	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp	Sp	BSA	5
glycolipid	SSEA-4-Sp - 12	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp	Sp	BSA	12

Blood Group H	TFiLNO(1-2,1-2,1-3)	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb-BSA	(Glc)	BSA	4
Blood Group H	TFiLNO(1-2,1-2,1-4)	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6[Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3]Galb-BSA	(Glc)	BSA	6
glycoprotein	Tgl	Thyroglobulin (Tgl)	na		na
non-human	X3Glc3	Xyla1-6Glcb1-4(Xyla1-6)Glcb1-4(Xyla1-6)Glcb1-BSA (X3Glc3)	(Glc)	BSA	15
non-human	Xylb4	Xylb1-4Xylb1-4Xylb1-4Xylb1-BSA (Xylb4)	(xyl)	BSA	22
cancer-associated glycopeptide mucins	DTVPLPTAHG-TF(Thr)-SASSTG				
cancer-associated glycopeptide mucins	DTVPLPTAHG-TF(Thr)-TF(Ser)-ASSTG				
cancer-associated glycopeptide mucins	DTVPLPTAHGTSASSTG				
cancer-associated glycopeptide mucins	DTVPLPTAHGT-TF(Ser)-ASSTG				
cancer-associated glycopeptide mucins	DTVPLP-TF(Thr)-AHGTSASSTG				
glycoprotein	gp120	gp120 glycoprotein			

Supplementary Table 3.3: List of individuals used for the gp120 competition assay

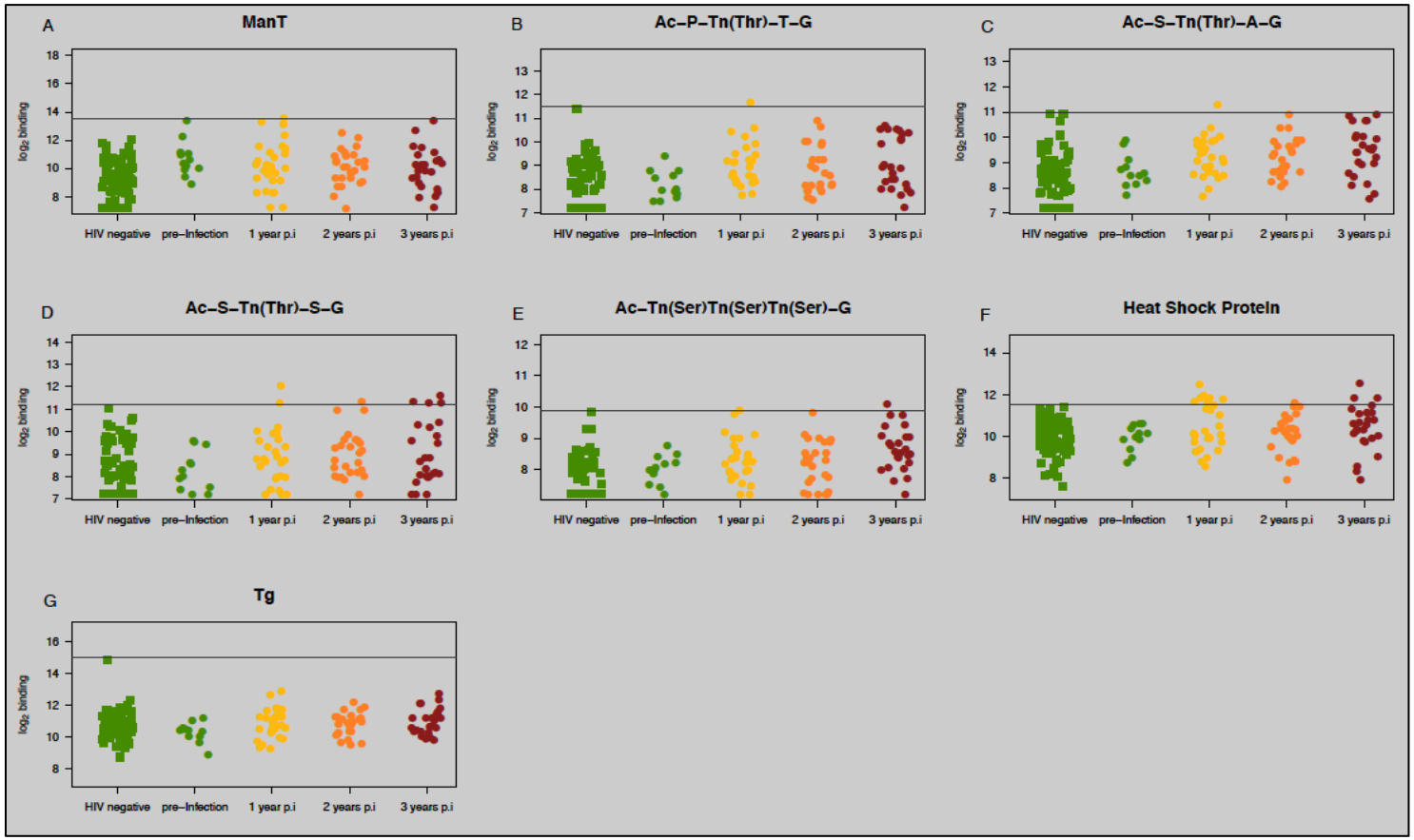
Patient ID (PTID)	Neutralization
CAP200	NON-BCN
CAP206	BCN
CAP221	NON-BCN
CAP244	INTERMEDIATE
CAP248	BCN
CAP255	BCN
CAP314	BCN
CAP322	BCN

Supplementary Table 3.4: p-values and fold-changes for glycans affected by the gp120 competition assay

Type	Name	P-value	Fold-change
N-linked	GlcNAc-Man3	0.005	1.6
	LacNAc-Man5	0.033	1.5
	Man7	0.044	2.9
	Man8	0.007	10.9
	Man9	0.017	2.5
Mannose	Man-a - 20	0.033	1.5
	Mana1-6Man-a - 15	0.030	1.4
	ManT	0.009	1.3
	G2M4 - 07	0.018	1.2
	Galb1-6Man-a - 13	0.006	1.3
Tn-peptide	Ac-A-Tn(Thr)-S-G - 05 & 08	0.005, 0.0177	1.7, 1.8
	Ac-P-Tn(Thr)-T-G - 05	0.004	1.7
	Ac-S-Tn(Ser)-S-G - 04 & 22	0.010, 0.015	1.8, 1.23
	Ac-S-Tn(Thr)-A-G - 04	0.004	2.1
	Ac-S-Tn(Thr)-G-G - 03 & 07	0.007, 0.008	1.6, 1.5
	Ac-S-Tn(Thr)-S-G - 04	0.023	1.6
	Ac-Tn(Ser)Tn(Ser)Tn(Ser)-G - 03	0.014	1.5
	Ac-Tn(Thr)-Tn(Thr)-Tn(Thr)-G - 08	0.022	1.2
	Ac-T-Tn(Thr)-P-G - 04 & 08	0.008, 0.05	1.8, 1.6
	Ac-V-Tn(Thr)-S-G - 04 & 08	0.014, 0.04	2.0, 1.7
Glycolipid	GM2-Sp - 14	0.024	2.1
	Gb4 - 09 & tetra	0.001, 0.009	1.4, 1.3
Glycoprotein	CEA	0.026	2.7
	fetuin (asialo)	0.034	1.2
	gp120	2.0E-06	21.4
	hsp90	0.015	1.3
Heparin	Hep-5000	0.016	2.9
Blood Group	BG-A1	0.007	1.7
	BG-H1	0.005	2.3
	BG-H2	0.029	2.5

	Globo H	0.007	1.8
Lewis Antigen	Sialyl LeX	0.05	1.7
Fucosylated	Fuc-a	0.029	1.1
	Fuc-b	0.012	1.3
GalNAc carbohydrate	GalNAc-a - 04 & 22	0.006, 0.02	1.4, 1.2
	GalNAca1-6Galb - 04	0.003	1.3
	GlcNAca1-4Galb - 03 & 20	0.004, 0.03	4.4, 1.4
	GalNAc-b - 21	0.011	1.2
	GlcNAc-b - 21	0.004	1.3
Glc carbohydrate	Glc-b - 23	0.003	1.2
GlcNAc carbohydrate	3'GN-Di-LacNAc-Sp - 06	0.05	1.2
Type 2 carbohydrate	LacNAc - 22 & trimeric	0.019	1.5
	LNnT - 04	0.031	2.0

Shown are the p-values and fold-change differences between the PBS and gp120-incubated sera for the glycans that had p-values ≤ 0.05. Highlighted in red are the glycans that were previously selected in Method 1, Method 2 and/or bound by bNAbs. Highlighted in green are the glycans that showed an elevation in binding in the gp120-incubated sample compared to the PBS-incubated sample. All other glycans showed a decrease in binding in the gp120-incubated sample.



Supplementary Figure 3.1: Longitudinal binding of glycans with elevated binding during HIV-1 infection.

A) ManT longitudinal binding. Shown are the log_2 binding values for all sera, with HIV-negative and pre-infection in green, one year post-infection in yellow, two years post-infection in orange and three years post-infection in red. **B-E)** Tn-peptide longitudinal binding. **F)** Thyroglobulin binding. **G)** Fucosyl-b longitudinal binding.