UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Characterizing IgA1 and IgG1 Constant Region Diversity in South Africans

Itai T Ncube 1577419

Master in Science (MSc) in medicine in the field of Vaccinology

Supervisors:

Dr Cathrine Scheepers

Prof Penny Moore

Dr Bronwen Lambson

Dr Clare Cutland

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Itai Tendai Ncube (Student number: 1577419) am a student registered for the degree of Master in Science (MSc) in medicine in the field of Vaccinology in the academic year 2020.

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 degree 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.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: ITNcube Date: 14 May 2021

Acknowledgements

I would like to thank the participants who participated in this study. My supervisors, Dr Cathrine Scheepers, Dr Bronwen Lambson, Prof. Penny Moore and Dr Clare Cutland for their guidance and supervision, your support and motivation made this project a success.

Abstract

Antibodies consist of five different isotypes which are encoded by different genes. There is evidence that different genotypes of these isotypes can affect antibody function. However, little is known about genotypes in individuals of African descent. Therefore, this study sought to characterise the genetic diversity of *IGHA1* and *IGHG1* genes, the genes that encode IgA1 and IgG1 antibodies which are important in viral infection, in six Black Africans from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) cohort. Genomic deoxyribonucleic acid (DNA) was used to amplify, clone and Sanger sequence these genes and the sequences were compared to reference alleles from the International Immunogenetics information system (IMGT), the global reference database for immunogenetics. Eight novel *IGHA1* alleles were identified. These alleles closely matched the *IGHA1*01* reference allele, with one to four single nucleotide polymorphisms (SNPs). The majority of SNPs were synonymous; however, one novel allele (variant 1) had a R392H amino acid substitution within the CH3 region. Within *IGHG1,* this study identified three alleles that matched known IMGT alleles, and two alleles that were novel. Overall, this pilot study highlights a high level of genetic diversity in this population that may have functional relevance to immune responses in infection and vaccination, and warrants further investigation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADCC: Antibody dependent cellular cytotoxicity

ADCP: Antibody dependent cellular phagocytosis

CDC: Complement dependent cytotoxicity.

CAP: CAPRISA participant

CAPRISA: Centre for the AIDS Programme of Research in South Africa

CH: Constant heavy

CHS: Tailpiece

CNV: Copy number variation

D: Diversity region

DNA: Deoxyribonucleic acid

dH2O: Deionised water

E.coli: *Escherichia coli*

ELISA: Enzyme linked immunosorbent assay

EtOH/NaAc: Sodium Acetate Ethanol

FcR: Fc receptor

FcαRI: Fc alpha receptor I

FcγRI: Fc gamma receptor I

HIC: High income countries

HIV: Human immunodeficiency virus

IgA1: Immunoglobulin Alpha 1

IgG1: Immunoglobulin Gamma 1

IGH: Immunoglobulin heavy

IGHC: Immunoglobulin heavy constant

IGHA1: Immunoglobulin heavy constant alpha 1 gene

IGHA2: Immunoglobulin heavy constant alpha 2 gene

IGHE: Immunoglobulin heavy constant epsilon gene

IGHG1: Immunoglobulin heavy constant gamma 1 gene

IGHV: Immunoglobulin heavy variable genes

IGKC: Immunoglobulin kappa constant gene

IMGT: International immunogenetics information system

J: Joining region

kb: Kilobase

L: DNA ladder

LB: Luria-Bertani

LMIC: Low and middle income countries

N: Negative control

NGS: Next generation sequencing

P: Positive control

PBMCs: Peripheral blood mononuclear cells

PCR: Polymerase chain reaction

RAG: Recombination activating genes

RNA: Ribonucleic acid

RPM: Revolutions per minute

SNP: Single nucleotide polymorphism

s: Seconds

µL: Micro litres

UV: Ultra violet

v: Variant

V: Variable region

1. Introduction

1.1. The importance of antibodies

Antibodies or immunoglobulins are a fundamental part of the immune system and are the main drivers of the humoral response. These Y-shaped molecules are divided into a variable region referred to as the Fab and a constant region referred to as the Fc (Figure 1). The variable region gives an antibody its specificity in binding antigens, while the constant region triggers effector functions. These molecules either exist in a membrane bound form, where they are referred to as B-cell receptors, or in a soluble form, where they are known as secretory immunoglobulins (1). The function of the membrane bound B-cell receptors is to mediate an activating signal after contact with an antigen (1). Upon activation, and with the help of CD4 helper T-cells, B-cells differentiate and become plasma cells which secrete soluble antibodies (1,2).

Soluble antibodies perform various functions such as neutralization, opsonisation, complement dependent cytotoxicity (CDC) and antibody dependent cellular cytotoxicity (ADCC). Neutralization is mediated by the Fab region and results in the blocking of pathogens from infecting host cells (1). The remaining functions are mediated via the Fc region through specific Fc receptors (FcRs), where opsonisation results in the internalization and degradation of bound pathogens through phagocytosis; CDC involves the recruitment of complement proteins that can directly kill a pathogen and enhance its phagocytosis; and ADCC involving the recruitment of cytotoxic cells such as natural killer cells and directing their activity to infected cells, tumor cells, and/or pathogens (1) (Figure 1).

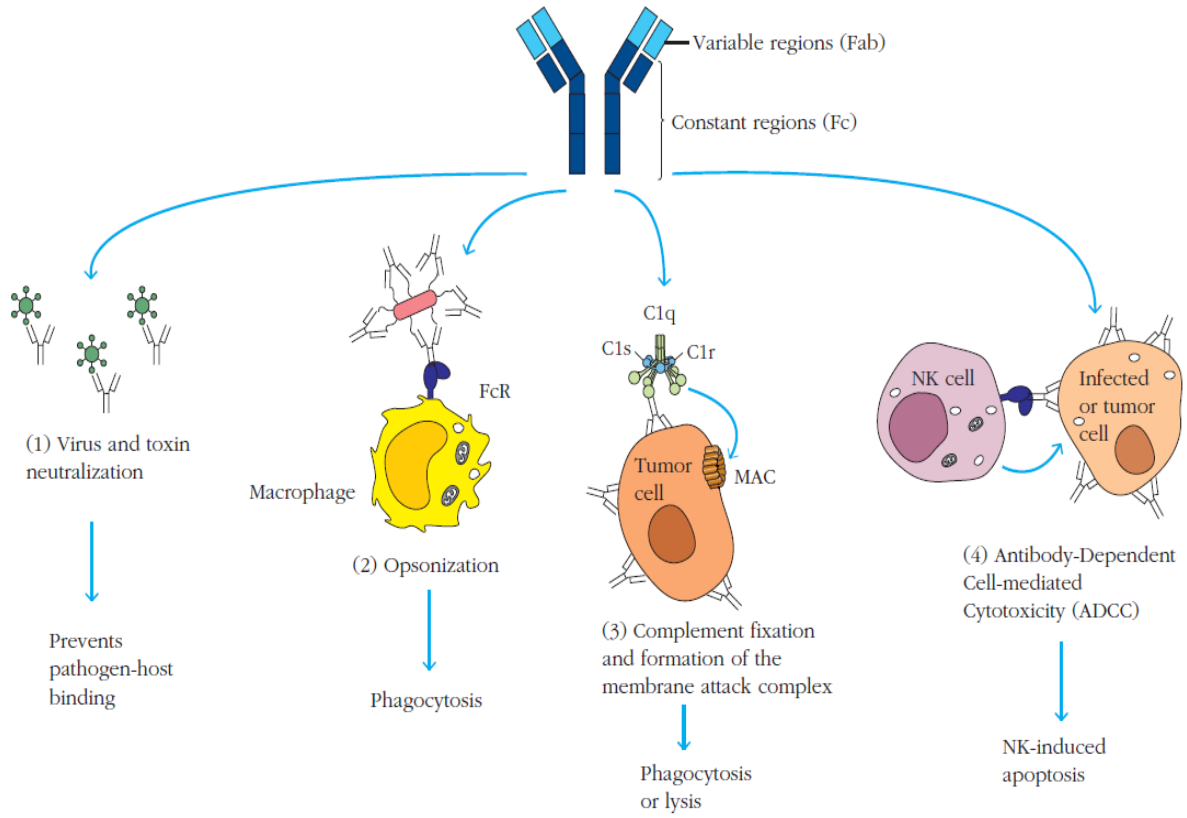

Figure 1: Schematic of the functions mediated by antibodies

The variable region binds to antigenic epitopes to block and neutralize pathogens or toxins. The constant region mediates opsonisation, complement fixation and ADCC by interacting with Fc receptors (FcR) on effector cells such as macrophages and natural killer (NK) cells (1).

In addition to being produced as a result of natural infection, antibodies are also used as therapeutics via passive infusion of polyclonal or monoclonal antibodies isolated from animals (such as horses) or humans and transferred to another human by injection or infusion (3). Monoclonal antibody therapy for cancer, autoimmune diseases and more recently for infectious diseases is a substantial pharmaceutical market (4,5). Furthermore, antibodies form the basis of many diagnostic and research assays such as the detection of antibodies associated with a disease or infection via enzyme linked immunosorbent assay (ELISA) or to identify proteins of interest such as cytokines (1).

1.2. Vaccines

Vaccines are man-made biological preparations that provide immunity to a particular infection but without risk of disease (3). Vaccines stimulate the immune system to produce antigen-specific humoral and cellular responses which generally lasts for many years, sometimes a lifetime because they induce memory. Vaccines can be classified into four basic types: live attenuated, inactivated, polysaccharide and recombinant vaccines, each of which have different characteristics which determine how they are administered (3).

The route of administration can be intranasal, oral and parenteral, and this impacts the nature of the immune response. For example, intranasal administration preferentially stimulates antibodies in mucosal compartments rather than systemic responses (6). Vaccines that target the humoral response such as the hepatitis and smallpox containing vaccine induce secretory antibodies which are able to neutralize these pathogens as well as induce effector functions as observed in natural infection. In addition to humoral responses vaccines can also induce cellular responses. In the vaccines that target cellular responses, such as certain human immune deficiency virus (HIV) vaccines, measles and mumps the effector cells are CD4+ helper T-cells and CD8+ cytotoxic T-lymphocytes. Helper T-cells secrete cytokines which can activate various phagocytic cells (such as neutrophils and monocytes) and CD8+ cytotoxic T-lymphocytes enabling them to kill microorganisms and infected cells (7,8). Phagocytic cells protect the host against intracellular bacteria and cytotoxic T-lymphocytes kill virus infected cells and tumor cells (7).

The impact of vaccines on disease prevention globally has been enormous. The Diphtheria-Tetanus-Pertussis containing vaccine is 80-90% effective amongst infants who receive the full dosing schedule, and vaccines such as the Polio and Rubella vaccines are 90% and 97% effective respectively (9). These statistics translate into 50,000 deaths prevented by the

Diphtheria-Tetanus-Pertussis containing vaccine, while the Rubella and Polio containing vaccines prevent 349,967 and 65,673 infections per annum respectively (10–12).

While vaccines are generally highly efficacious and are able to confer protection from disease and infection, some vaccines have low efficacy (8) or efficacy that varies annually, such as the influenza vaccine. Vaccine efficacy is the percentage reduction in disease in a vaccinated group compared to an unvaccinated group under optimal conditions such as randomized control trials (13). Nonetheless, even moderately efficacious vaccines can have a substantial public health benefit.

1.3. Antibody sub-classes

Antibody isotype can alter antibody effector function, which is important as it not only affects an immune response to infection but can also impact vaccine efficacy (14). There are five immunoglobulin isotypes namely IgM, IgD, IgE, IgG and IgA which are encoded for by different IGHC genes. IgA and IgG are further divided into two (IgA1 and IgA2) and four (IgG1, IgG2, IgG3 and IgG4) subclasses, respectively (1). IgM is the first antibody to be produced and is used as a marker of current and/or recent infection (1). This is then followed by IgA and IgG, which are important in systemic and mucosal immunity (1,15,16). IgE has a significant role in allergic reactions and has been shown to enhance immune responses from other antibody isotypes like IgG (1,17). The function for IgD is not fully understood (1), and unlike the other antibody isotypes, IgD remains membrane bound whereas IgM exists as a pentamer, IgA as either monomeric or dimeric, IgG and IgE as monomers (1,15).

In the mucosa, which is the main entry site for many pathogenic organisms, secretory IgA plays a key role in preventing infections by neutralizing invading pathogens (1,18). Secretory IgA has an additional segment called the secretory component attached to the Fc region (1,19). The secretory component is produced by epithelial cells of mucous membranes and its function is

to protect sites that are susceptible to protease cleavage in the hinge region of secretory IgA, allowing this molecule to exist for longer in a protease-filled environment (1).

IgA has also been shown to actively contribute to the initiation of inflammation at both mucosal and non-mucosal sites (19). IgA interacts with Fc alpha receptor I (FcαRI) through its constant region to initiate effector functions by myeloid cells (neutrophils, monocytes etc.) (19). The formation of immune complexes triggers FcαRI to synergize with various other receptors to amplify an immune response (15,19). This FcαRI mediated inflammation plays a key role in host defences and the generation of tissue specific immunity (19).

IgG, which is the most abundant antibody isotype, plays an important protective role in the systemic compartment (1). The IgG subclasses are more than 90% identical at the amino acid level, however, each subclass has a unique profile with respect to triggering effector functions, antigen binding, serum half-life and placental transport (1,20). IgG interacts with the Fc gamma receptor I (FcγRI) found on myeloid cells and the neonatal Fc receptor on the placenta however IgG subclasses have different affinities for these receptors (20–23). IgG1 is the main antibody produced against soluble protein antigens and membrane proteins and readily crosses the placenta to protect the developing foetus better than other IgG subclasses (20,24). Furthermore, the IgG1 subclass is key in maintaining a balanced humoral response, and a deficiency of IgG1 has been associated with recurrent infections (20).

IgA and IgG have been shown to have a protective role in individuals that were highly exposed to HIV through repeated sexual exposure by mediating neutralization and ADCC (8,20,25). Thus IgA and IgG isotypes are pivotal in the prevention of infections at mucosal and non-mucosal sites and are the focus of this study.

1.4. Antibody structure

Antibodies are comprised of two identical heavy and two identical light chains, the latter of which are either Kappa (k) or Lambda (λ) chains (1) (Figure 2).

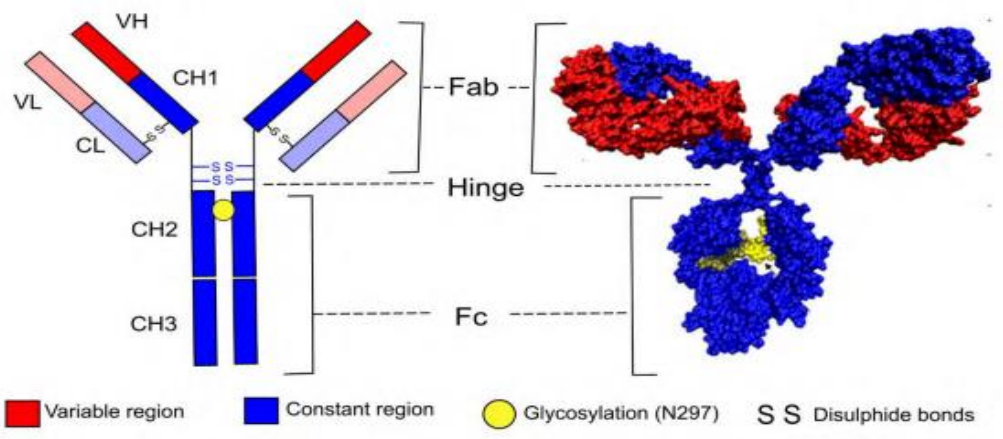

Figure 2: Schematic showing the variable and constant region of an antibody, figure taken from Richardson, PhD thesis (26)

The primary structure of an antibody (left) showing constant region heavy chains in blue, constant light chains in light blue. Variable heavy and light chains are red and pink respectively. The quaternary structure (right) showing associations between domains and each solid ball represents an amino acid. In each structure the yellow part represents a glycan.

The two main regions of an antibody are the Fab and the Fc region. The Fab region comprises of the heavy chain variable region and constant (CH) 1 domain that is bound to light chain variable region and constant region by disulphide bonds as shown in Figure 2. Located below the Fab region is the hinge, which is only found on IgA, IgD and IgG. The hinge region is rich in prolines and gives the arms of IgA, IgD and IgG segmental flexibility, allowing them to assume a wide variety of angles with respect to one another, which facilitates efficient antigen binding (1). The Fc region located below the hinge region can have two CH domains as seen in IgA, IgD and IgG, however, IgE and IgM each have three CH domains (1). Antibodies are glycosylated, which is critical to the structure, function and stability of the immunoglobulin

molecule (1,27). Furthermore, IgA antibodies possess an 18 amino acid tail piece (CHS) after the CH3 domain, which allows for polymerization of IgA molecules.

1.5. Antibody development via gene rearrangement

Antibodies are generated as a result of somatic recombination of individual V, D and J genes, which results in the variable regions of the heavy and light chains of an immunoglobulin (1,28). Somatic recombination occurs in B-cells and uses repeated and highly homologous sets of genes located at three immunoglobulin loci. These loci encompass λ, k and heavy chain genes located on chromosomes 2, 22 and 14 respectively (1,29,30). Each locus includes variable (V), diversity (D), joining (J) genes and the constant genes, however, the light chain (λ and k) loci lack diversity genes (1). During B-cell maturation in the bone marrow, VDJ genes in these loci are rearranged randomly and brought together to form a functional VDJ gene cluster paired with constant genes (1).

The process of somatic recombination happens in a stepwise process first involving the rearrangement of one heavy chain D (D_H) gene and one heavy chain J (J_H) gene. Any deoxyribonucleic acid (DNA) between these genes is deleted (7). The D_H-J_H recombination is followed by joining of one heavy chain V (V_H) gene to the D_H-J_H gene cluster, forming a rearranged V_H-D_H-J_H gene cluster and all other genes between V_H and D_H genes are deleted (7,24,25). This creates the DNA template for a primary transcript (unspliced ribonucleic acid (RNA)) containing the V_H-D_H-J_H (variable) region of the heavy chain and the constant region genes (7). Translation of this mature messenger RNA leads to synthesis of the IgM heavy chain (31,32) (Figure 3). For the antibody to be fully functional, synthesis of light chains follows a similar process to that of heavy chains (7,24). The process of VDJ recombination is mediated by enzymes such as the recombination activating genes (RAG) 1 and 2 (1).

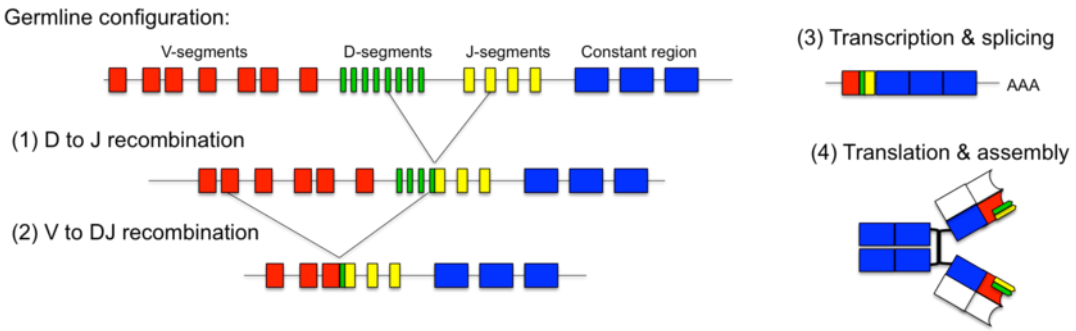

Figure 3: VDJ recombination steps in the synthesis of immunoglobulin heavy chains

VDJ recombination requires the use randomly selected VDJ genes in the synthesis of the Fab region. The primary transcript in step 3 is unspliced RNA that is processed to have a polyadenylated tail and the constant mu and delta chains which will later be translated and synthesized as membrane bound immunoglobulin (1).

Following VDJ recombination and the synthesis of membrane bound antibodies with the mu and delta heavy chains, B-cell activation via the B-cell receptor results in isotype class-switching or class-switch recombination (1). Class-switch recombination changes the production of one antibody isotype to another. Isotype class-switching generates antibodies with the same Fab, but with different constant regions (1,14,25). This is important because the new isotypes are able to recognize the same antigen but interact with different effector cells which improves the ability to clear the invading pathogens (1).

Class-switch recombination involves the pairing of gamma, epsilon or alpha heavy chain genes with the functional VDJ gene segment, through cutting and rejoining of heavy chain DNA as shown in Figure 4 (1). After cutting and re-joining of heavy chain DNA, the next constant region will lie directly downstream from the rearranged VDJ gene cluster. For class-switching to occur, there is recombination between the donor and acceptor switch regions located upstream from each Constant Region gene (1).

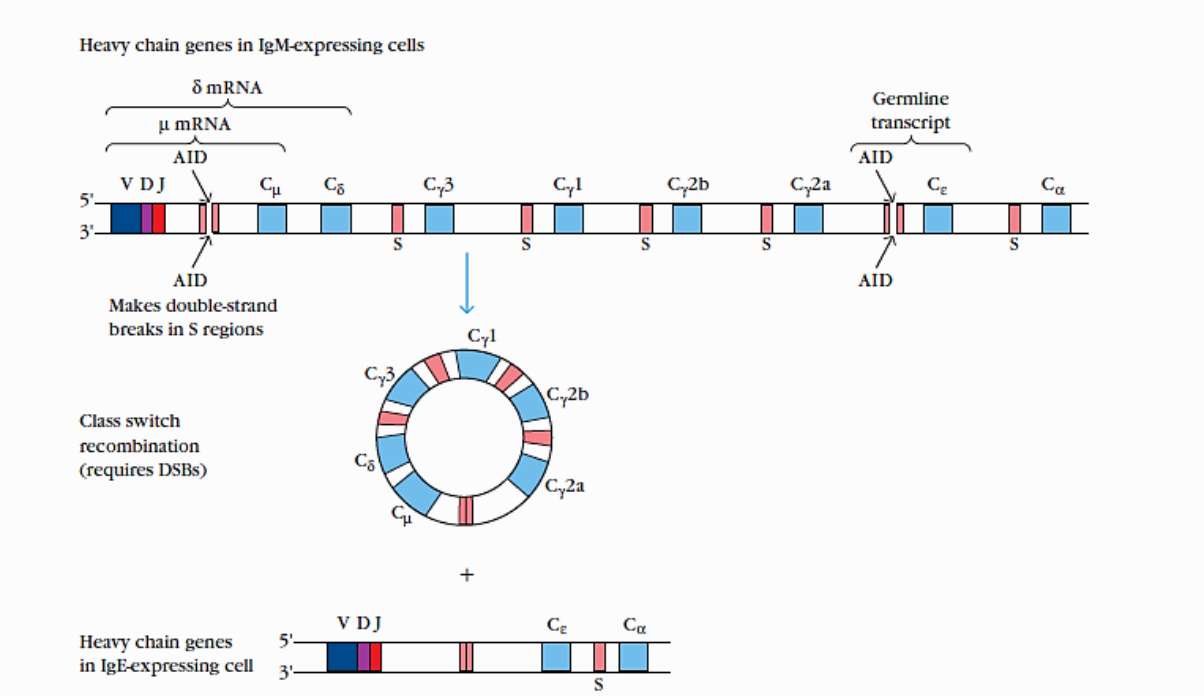

Figure 4: Class-switch recombination from a mu to an epsilon heavy chain constant region gene.

Class-switch recombination requires that activation induced cytidine deaminase (AID) enzyme makes double stranded breaks in the switch regions. This results in the loss of the intervening DNA sequences, and the double stranded breaks are resolved by normal DNA repair mechanisms (1).

1.6. IGH locus and antibody genetic diversity

The immunoglobulin heavy chain locus is found on chromosome 14, at band 14q32.33 at the telomeric extremity of the long arm in humans (29,30) (Figure 5). The IGH locus is 1250kb in length, including 170-176 functional genes and 41 pseudogenes. There are 123-129 functional IGHV genes; 27 for IGHD; 9 for IGHJ and 11 IGHC genes (those responsible for encoding the different antibody isotypes) (29).

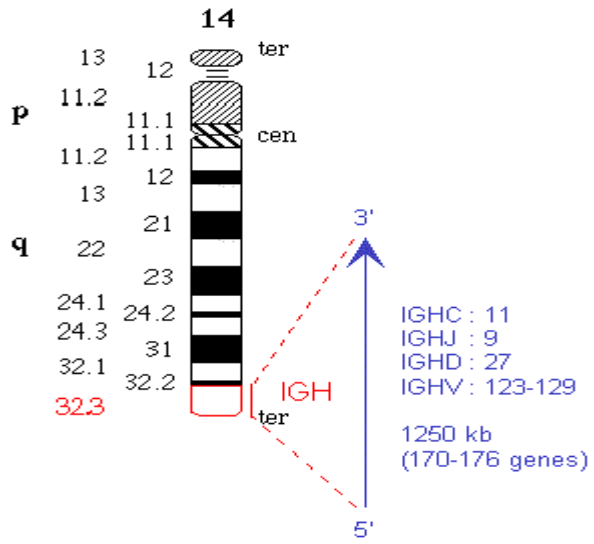

Figure 5: The IGH locus, where the IGHC genes are located on chromosome 14q32.33

The human IGH locus is in reverse orientation: being located on the chromosome long arm, the 5' end of the locus is telomeric and the 3' end is centromeric (68).

The IGH locus is susceptible to a high frequency of insertion and deletion events as it contains multiple copies of homologous genes (29,31). These insertion or deletion events can create new genetic variants that can be longer or shorter, and these genetic variants can alter the function of the translated antibody (30). Whole gene insertion or deletion events result in copy number variation (CNV) (33). Sequencing studies in healthy individuals have shown that multigene deletions in the IGH locus can cause copy number variation of five to eleven IGHV genes, though little is known about CNV within IGHC genes (34).

Genes in the IGH locus are also known to be polymorphic which leads to high levels of allelic variation (30). Within IGH genes, allelic variants can involve a haplotype of single nucleotide polymorphisms (SNPs) or involve a single SNP both are referred to by using a "*" allele. In some cases, these nucleotide sequence changes can lead to amino acid substitutions which can have an impact on Fc effector function, thus impacting responses to infection and vaccination (30). Haplotype differences within the germline at the nucleotide level are referred to as alleles, whereas haplotypic changes in the protein amino acid level are referred to as allotypes (1).

Allotypes have been determined for some heavy chain constant genes (*IGHA2*, *IGHE* and *IGHG* antibodies) and the Kappa light chain (*IGKC*) (1,36,37). Allotypes can differ between individuals and within and between ethnicities (1). This means that exposure to non-self-allotypes may elicit an anti-allotype response especially after blood transfusions. Allotypes are designated by the class, subclass and the allele number. Thus, for example for IgG1 they are called G1m1 - G1m28, and for IgA2 are referred to as A2m1 and A2m2. Increasingly, new research is highlighting the role of heavy chain allotypes in affecting the effector function of immunoglobulins. These heavy chain allotypes have been shown to be associated with susceptibility to infections and autoimmune diseases, and have been described for both IgA and IgG (22,37). In HIV studies, G1m3 carriers have been shown to have reduced levels of antibodies specific to HIV antigens such as gp140 compared to G1m1 carriers (8). Furthermore, IgA/IgG polymorphisms have been shown to affect serum concentrations of these antibodies (22,42).

These findings highlight the need for further research to increase our understanding of the IGH locus (37). A major research gap is the need to glean additional information about genotypes from other populations. A recent study in Brazil that investigated the immunoglobulin constant heavy gamma chain (*IGHG*) genes found 28 novel alleles and many single nucleotide polymorphisms (SNPs) which points to population differences compared to known alleles (43). We have previously examined the immunoglobulin heavy gamma 1 (*IGHG1*), *IGHG2* and *IGHG3* gene fragments amongst individuals from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) cohort, in collaboration with Dr Mary Carrington, from the National Cancer Institute. In that study, using gene fragments, 63 novel SNPs were identified across the constant region genes (unpublished data). However, no studies thus far have examined the genetic diversity across the entire *IGHG* and *IGHA* genes to identify allelic and allotypic diversity in black South Africans.

1.7. Aims and Objectives

This study was undertaken to provide insight into the diversity of *IGHA1* and *IGHG1* constant region genes using DNA extracted from PBMC's collected from six Black South African women who were participants in CAPRISA 002/004 studies.

The specific objectives of this study were:

1) To sequence the *IGHA1* and *IGHG1* constant region genes from six CAPRISA 002/004 participants.
2) To compare gene sequences to genomic data in the International Immunogenetics information system (IMGT) (www.imgt.org, the global standard reference database for immunogenetics) to identify new single nucleotide polymorphisms (SNPs), novel alleles and/or allotypes.

2. Methods

2.1. Study population

The samples used in this study were collected from young women between the ages of 18-40 years, from urban and rural KwaZulu Natal enrolled into the CAPRISA 002/004 studies (44,45). The CAPRISA 002 study assessed the natural history of acute HIV infection and the CAPRISA 004 cohort was established to evaluated the effectiveness of Tenofovir gel as a microbicide (44,45). This current study is an extension of a previous study assessing levels of genetic diversity in the immunoglobulin heavy variable (*IGHV*) genes within the CAPRISA cohort (46). Ethics application for the consenting of participants and sample collection was approved by the University of KwaZulu-Natal Biomedical Research Ethics Committee for the CAPRISA 002/004 studies, ethics/reference number Ref: E111/06. Ethics clearance to sequence the *IGHA1* and *IGHG1* genes in the participants involved in this research was approved by the Faculty of Health Sciences Human Research Ethics Committee University of the Witwatersrand (Ethics clearance certificate number M191075, Appendix 1).

2.2. DNA Extraction

Stored DNA from the six CAPRISA participants (CAP88, CAP177, CAP206, CAP255, CAP256 and CAP257) were used for this study. The DNA from these samples had previously been extracted by Alaine Marsden from 10 ml of blood using the Wizard® Genomic DNA purification kit (Promega; Wisconsin, USA). Extraction entailed lysis of the cells by mixing 30 ml cell lysis solution and 10 ml of blood, after mixing the solution was incubated for 10 minutes at room temperature. Following lysing of red cells, the samples were spun for 10 minutes at 2000 x g and the supernatant was discarded leaving only white cells in the samples. The white cells were vortexed for 10 - 15 seconds (s) to resuspend the cells and 10 ml of nuclei lysis solution was added to the cells in order to lyse them. To ensure adequate lysis of the cells the solution was mixed using a pipette 5-6 times. After lysing the cells, protein precipitation

solution was added to the nuclear lysate and the solution was vortexed vigorously for 10 - 20s, then it was spun for 10 minutes at 2000 x g at room temperature. After centrifugation, the supernatant was transferred to a 50ml centrifuge tube with 10ml isopropanol to precipitate the DNA. After precipitation, the solution was spun for 10 minutes at 2000 x g at room temperature, during this step DNA forms a white pellet and the supernatant was discarded. Following this step, the DNA was washed with 70% ethanol at room temperature and was rehydrated using 800µL of rehydration solution overnight at room temperature. The DNA concentration of the samples was assessed using the NanoDrop® ND-1000 (ThermoFisher Scientific; Waltham, USA), and diluted to 10ng/µL.

2.3. Amplification of the *IGHA1* and *IGHG1* genes

PCR is a molecular biology technique used to amplify fragments of DNA using primers and thermostable polymerase enzymes (47). The amplified DNA can be used for downstream applications such as DNA sequencing, as is the case for this research project. Complete *IGHA1* and *IGHG1* genes were amplified as single fragments. The extracted DNA was amplified using the proofreading DNA polymerase Q5 (NEB) (New England Biolabs® Inc.; Ipswich, USA) and *IGHA1* or *IGHG1* primers that were positioned on introns pre-CH1, post-CH3 and/or tailpiece (CHS) (Figure 6). The *IGHG1* primers were designed by Dr Mary Carrington of the National Cancer Institute, Frederick, MD, USA and have been used in her laboratory to amplify DNA samples from different donors. The *IGHA* primers were designed at the NICD Centre for HIV and STIs by Dr Bronwen Lambson and amplify both *IGHA1* and *IGHA2*.

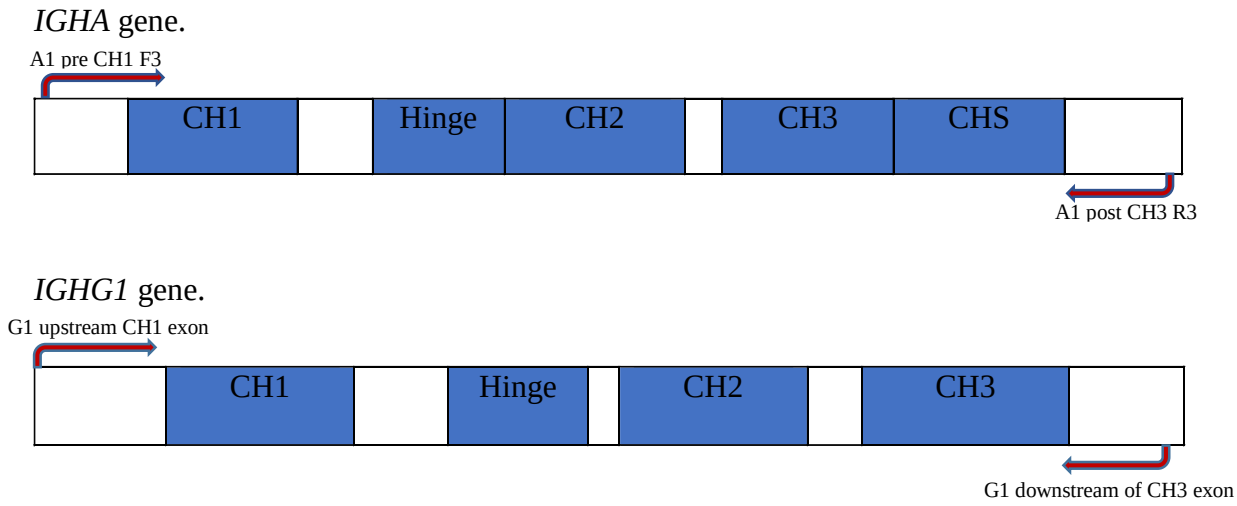

Figure 6: Schematic of primer positions within IGHA and IGHG1 genes

The genes were ±1.5kb. Blue blocks= exons and white blocks= introns. Arrows indicate where the primers bind on the genes.

Table 1: Primers used for the amplification of *IGHA* and *IGHG1* genes as single fragments

Primer name	Primer sequence
A1 pre CH1 F3	5’ GGG CCG CGT CCT CAC AGT GCA TTC 3’
A1 post CH3 R3	5’ CAG ACA GGC GCC GGA TGG AAG CG 3’
G1 upstream CH1 exon	5’ CTG AAC CTC GCG GAC AGT T 3’
G1 down stream of CH3 exon	5’ GAA AGA ACC ATC ACA CAG TCT CG 3’

The reaction mixtures for PCR amplification were set up as shown in Table 2. PCR thermocycling conditions to amplify the *IGHA1* and *IGHA2* genes were set at 98ºC for 30s and then 98ºC for 10s; 72ºC for 20s; 72ºC for 45s (35 cycles) and final extension at 72ºC for 2 minutes. For the *IGHG1* gene the annealing temperature was 56ºC, but cycling conditions were otherwise identical.

Table 2: PCR reaction mix used in the amplification of the *IGHA* and *IGHG1* genes

Name	Volume (1x reaction)
dH2O	11µL
Forward primer (25µM)	0.12µL
Reverse primer (25µM)	0.12µL
*Q5 Hot start master mix	12.5µL
Sample (DNA 10ng/µL)	1µL
Total	24.74µL

* The Q5 hot start master mix contains dNTPs, Mg^{++} and a buffer requiring only the addition of primers and DNA template.

2.4. Agarose gel electrophoresis of PCR products

The principle of agarose gel electrophoresis is that charged molecules such as DNA which is negatively charged can be separated by size under a constant current in an aqueous solution (48). DNA can migrate through an agarose gel from the negative node to the positive node and the size of the molecule determines how fast it will migrate towards the positive node (23). The size of the PCR products was confirmed using 1% agarose gel electrophoresis. The gel was prepared by mixing 100 ml of Tris-Acetate-EDTA-TAE buffer and 1g agarose powder. This solution was heated to dissolve the agarose powder and 2.5µL of 10mg/ml ethidium bromide was added to the solution after heating. Electrophoresis was performed for 30 minutes at 80 volts. Products were visualised under ultraviolet (UV) light because DNA binds to the ethidium bromide, forming a DNA-ethidium bromide complex which gives off a fluorescent light in the presence of UV light indicating the location of the DNA fragments. The DNA fragments were compared to a marker Quick load® purple 1 kb DNA ladder (New England Biolabs®; Ipswich, USA) containing DNA fragments of predetermined sizes. The gel was imaged using the Chemidoc™ imaging system (Bio-Rad; Hercules, USA).

After confirming that the amplified DNA fragments bands were of the correct size (approximately 1.5 kb), the DNA was purified using the Qiagen, QIAquick® PCR Purification kit (Qiagen; Hilden, Germany). This kit uses a silica membrane to bind DNA in a high salt concentration while a washing step ensures that residual proteins and polysaccharides are

removed from the DNA. After washing the DNA was eluted using the kit supplied elution buffer.

2.5. Cloning and transformation of *IGHA* and *IGHG1* plasmids

DNA cloning is a process that is used to make multiple, identical copies of a gene. For this study, blunt end cloning was utilized and the gene of interest was inserted into a plasmid and DNA ligase ligated the gene to the plasmid, producing plasmid/recombinant DNA (49, 50). This recombinant DNA was then transferred into bacteria in a process called transformation, which allows for making large quantities of the plasmid DNA. Cloning is used to separate alleles in a heterozygous sample as each plasmid accepts only a single amplicon, allowing for the identification of individual haplotypes. After transformation, bacterial colonies containing plasmids with the gene of interest can then be grown and sequenced and haplotypes identified (49,50).

Blunt end cloning was used to create large numbers of clones of the *IGHA1* and the *IGHG1* genes. Due to sequence similarity, the primers used to amplify the *IGHA1* gene also amplified the *IGHA2* gene. Therefore clones belonging to the *IGHA2* gene were observed during analysis of results and these alleles reported on. Blunt end ligation occurs because of random collisions between a blunt ended PCR product and a blunt end cloning vector which leads to the DNA insert and vector interacting. During this interaction the DNA ligase, ligates the DNA insert to the vector by creating a phosphodiester bond between the nucleotides of the DNA insert and vector, joining them and forming a permanent bond, which then creates a circular plasmid that can be propagated in bacteria (51). The Q5 polymerase produced blunt ended PCR products. These PCR products were ligated into the pJET1.2/ Blunt cloning vector CloneJET (Thermo Scientific; Waltham, USA) using the reaction mix as shown in Table 3. The samples were then mixed by flicking the sample tubes 2-3 times and incubated at room temperature for 5 minutes to allow for ligation. After ligation the samples were transformed into chemically competent

JM109 *Escherichia coli* (*E. coli)* (Zymo Research; Irvine, USA). In the event that few or no colonies grew overnight using the JM109 cells, the more competent XL 10 Gold (*E. coli)* ultra-competent cells (Stratagene; San Diego, USA) were used as an alternative according to manufacturers' protocol. The samples were then plated on Luria-Bertani (LB) (Invitrogen ™; Carlsbad, USA) agar with 100 µg/ml carbenicillin and were incubated at 37ºC overnight.

Table 3: Ligation reaction mix

Name	**Volume (1x reaction)**
2X reaction buffer	10µL
Purified PCR product (5-10 ng/µL)	3µL
pJET1.2/Blunt Cloning Vector	1µL
Water	5µL
T4 DNA ligase	1µL
Total	20µL

2.6. Screening of bacterial colonies

Following cloning, colony PCR was used to screen for colonies with the target insert of 1.5kb. To do this, 4-6 single colonies were picked using a sterile pipette tip per plate and transferred into individual wells of a 96 well plate. Each well contained 100µL of LB (Invitrogen ™; Carlsbad, USA) broth supplemented with 100 µg/ml carbenicillin. These 96 well plates were incubated for 2 to 3 hours in an incubator at 37ºC without shaking, after which 1µL of sample per well, was added to a PCR reaction mix, containing the DreamTaq green PCR Master Mix (2X) (Thermo Scientific; Waltham, USA). As follows:

Table 4: Colony PCR reaction mix

Name	Volume (1x reaction)
DreamTaq Green PCR master mix	10µL
0.16 pJET Forward primer (vector) (25 µM)	0.4µL
0.16 pJET Reverse primer (vector) (25 µM)	0.4µL
dH2O (nuclease free)	8µL
Sample	1µL
Total	19.8µL

Colony PCR was run using the following thermocycling conditions: 95ºC for 1 min and then 95ºC for 30s; 55ºC for 30s; 72ºC for 1 min (35 cycles) and the final extension was run at 72ºC for 10 min. Following the amplification, the results were visualised using gel electrophoresis (1% agarose gel electrophoresis was performed at 80 volts for 30 minutes) to identify a well/colony with the DNA insert of 1.5kb in size. The samples were compared to a marker containing DNA fragments of predetermined sizes as described above in Section 3.4.

2.7. Plasmid preparation

The colonies that had an insert of the correct size were then miniprepped in 15 ml of LB broth (pre-warmed to 37ºC) and 100 µg/ml carbenicillin and incubated in a shaking incubator at 150-200 revolutions per minute (RPM) at 37ºC overnight. The following day plasmid DNA was extracted from the bacteria using the Qiagen QIAprep® spin kit (Qiagen; Hilden, Germany). This was achieved by centrifugation of the samples at 13000 RPM for three minutes to pellet the bacterial culture, the supernatant was discarded and the pellet was resuspended in 250µL buffer P1 (resuspension buffer). The resuspended samples were transferred to a micro centrifuge tube, after which 250µL of buffer P2 (lysis buffer) was added to the samples which were mixed by inverting four times until the solution became clear. After this step, 350µL of buffer N3 (neutralization buffer) was added to the samples, which were mixed and spun for 10 minutes at 13000 RPM and after spinning 800µL of the supernatant was transferred to a QIAprep 2.0 spin column in a micro centrifuge tube. The QIAprep 2.0 spin column binds to

extracted plasmid DNA while cellular debris are washed and discarded. To wash the sample 750µL of buffer PE was added to the samples in the QIAprep 2.0 spin columns which were then spun for 30-60s at 13000 RPM, the flow through was discarded and the samples spun again for a minute to remove residual wash buffer. After washing the QIAprep 2.0 spin columns were placed in clean micro centrifuge tubes to elute DNA. DNA was eluted by adding 50µL of buffer EB to the samples, this stood for 1 minute and was spun for 1 minute at 13000 RPM. The concentration of the extracted DNA was measured using the NanoDrop ® ND-1000 spectrophotometer (ThermoFisher Scientific; Waltham, USA).

2.8. Sanger sequencing of *IGHA* and *IGHG1* genes

Minipreps were Sanger sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems; Forster city, USA). Sanger sequencing uses dideoxynucleotides bound to a fluorescent dye (four different fluorophores are used for the four nucleotides) to generate DNA fragments of different lengths. During sequencing incorporation of a labelled dideoxynucleotide terminates the nucleotide chain synthesis (52). This generates fragments of different lengths, which a sequence analyser assembles into an electropherogram. The sequencer (3500XL Genetic analyser (Applied Biosystems; Forster City, USA)) uses an electric current to separate individual strands of DNA based on size and identifies the type of dideoxynucleotide at the end of each fragment. This pattern of fluorescence is then used to generate an electropherogram that can be read as a DNA sequence. For sequencing, two vector and four internal primers were used, shown in Table 5. The internal primers were different for each gene, whereas the same vector primers were used as the inserts were cloned into the same vector. Sanger sequencing has reliable reads of approximately 700 nucleotides whereas the length of *IGHA* and *IGHG1* are 1.5kb, therefore internal primers allowed for generating overlaps within the 700 nucleotides to obtain full coverage of the genes.

Table 5: Vector and internal primers used to sequence the *IGHA* and *IGHG1* genes.

Primer name	Primer sequence
A1 F1 (*IGHA1*)	5'GGG AAC GTG GTC ATC GCC 3'
A1 F2 (*IGHA1*)	5' GAC CGG CCT GAG AGA TGC C 3'
A1 F3 (*IGHA1*)	5' GCA CTG CTG CCT ACC CCG 3'
A1 F4 (*IGHA1*)	5' GGA CGT GCT GGT TCG CTG GC 3'
A1 Post CH2 R1 (*IGHA1*)	5' ACC TAA GAG CAG GTC CTC GAG G3'
VHCH F01 (*IGHG1*)	5' CAC CTC TGG GGG CAC AGC3'
CH2 F (*IGHG1*)	5'GGA CCG TCA GTC TTC CTC TTC CC 3'
HC rev (*IGHG1*)	5'GCA ACA CCA AGG TGG ACA AG 3'
G-CH3rev3 (*IGHG1*)	5' CTTCTCATGCTCCGTGATGCATGAGG 3'
Vector Forward primer (*IGHA1*/ *IGHG1*)	5' GGG CCG CGT CCT CAC AGT GCA TTC 3'
Vector Reverse primer (*IGHA1*/ *IGHG1*)	5' CAG ACA GGC GCC GGA TGGAAG CG 3'

The vector primers attached on the vector and they allowed for sequencing in the forward and reverse directions. The internal primers were mostly forward primers, however for *IGHG1* gene two reverse internal primers were introduced to ensure full coverage of the gene (Figure 7).

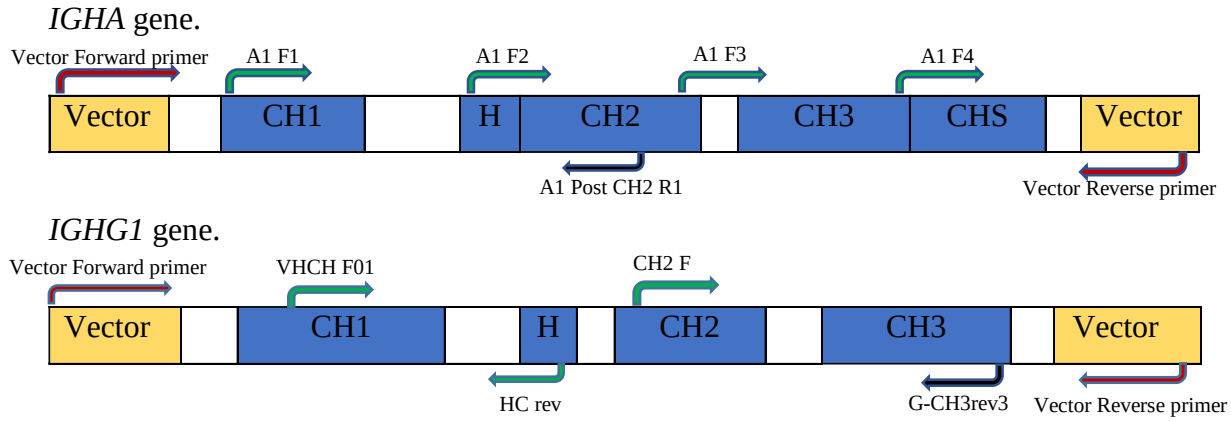

Figure 7: Location of sequencing primers

Red arrows represent vector primers. Green arrows represent *IGHA/IGHG1* (F) internal primers; Gold blocks = pJet vector; White blocks = introns; Blue blocks = exons and Black arrow = *IGHA/G1* (R) internal primer. Note: genes are not to scale.

The sample reaction mix were set up as shown in Table 6.

Table 6: Sequencing reaction mix

Name	**Volume (1x reaction)**
BigDye ™ terminator ver. 3.1 Ready reaction mix	1µL
BigDye ™ terminator ver. 3.1 5x sequencing buffer	1.5µL
Primer 2pmol/µL (F/R)	3µL
dH2O (nuclease free)	3.5µL
Sample (PCR~50ng or plasmid 300ng per well)	1µL
Total	10µL

Sanger sequencing was run using the following thermocycling conditions, 96ºC for 10s; 50ºC for 5s; 60ºC for 4 minutes for 25 cycles. Sequencing reactions were cleaned using the Sodium Acetate Ethanol (EtOH/NaAc) precipitation method. For the clean-up, 50 µL EtOH/NaAc (96.15% EtOH and 3.85% NaAc) was aliquoted to each well to precipitate the DNA, the samples were centrifuged for 30 minutes at 2000 x g on a table top centrifuge. After draining the wells by inverting the plates, the plate was centrifuged inverted for 1 minute at 150 x g. 100 µL of cold ethanol was added to the wells with DNA pellets to wash any remaining salts from the DNA, the plate was then further centrifuged for 5 minutes at 2000 x g. The plates were inverted for drainage and further centrifuged for 1 minute at 150 x g. This was followed by drying of the plate in a thermocycler at 65°C for 3 minutes, with the thermocyclers lid open and the reaction caps were removed. The samples were run on an ABI 3500xL Genetic Analyser (Applied Biosystems; Forster city, USA). At least 6 clones were sequenced per participant.

2.9. Data analysis

Sequencing results were imported into Sequencher® 5.4.6 DNA sequence analysis software (Gene codes Corporation; Ann Arbor, USA). Each chromatogram from a single clone was trimmed to remove nucleotides of poor quality. The chromatograms were aligned to generate contigs. These contigs were aligned to IMGT reference alleles on Sequencher® 5.4.6. The

alignments were done for each exon of the constant region for each gene, this included the tailpiece (CHS) for the *IGHA* genes. These alignments were exported into BioEdit Sequence Alignment editor (Ibis Biosciences, Inc.; Carlbad, USA) for further analysis. This step allowed for comparing the sequencing results to the IMGT reference alleles and easy identification of nucleotide substitutions and new genetic variants. The *IGHG1* results from this study were also compared to data generated from the same participants using the next generation sequencing (NGS) on PacBio Sequel I, generated as part of another study. Furthermore, these results (Sanger and NGS) were compared with the data generated previously using small fragments of the *IGHG1* constant region genes. As is typical within immunogenetics research alleles refer to either a haplotype or the entire gene sequence compared to a reference gene sequence, which may include either a single of multiple SNPs. Allotypes refer to a haplotype of amino acid changes that alter antibody function.

3. Results

This research project investigated genetic diversity in the *IGHA1* and *IGHG1* genes, which encode for the constant regions of IgA1 and IgG1, in six black South African women. As there are substantial sequence similarities between *IGHA1* and *IGHA2* genes, both genes were amplified and reported on in this study. The sequence identity between the *IGHA1* and *IGHA2* subclasses is high (~90%), and what differentiates the two subclasses is the 13 amino acid extension in the hinge region of *IGHA1* (36). Other changes such as the N-linked glycosylation motifs located in the CH1 and CH2 domains at positions 166, 211 and 337 are important differences (36). These features were used to distinguish *IGHA1* and *IGHA2* gene sequences.

3.1. PCR amplification of the *IGHA1*, *IGHA2* and *IGHG1* genes

All genes *(IGHA1*, *IGHA2* and *IGHG1)* were amplified with the same PCR protocol with the exception of the annealing temperatures which were 72°C for *IGHA1* and *IGHA2* and 56°C for *IGHG1*. After successful amplification of the target genes, DNA fragments of 1.5kb in size were observed following agarose gel electrophoresis (Figure 8).

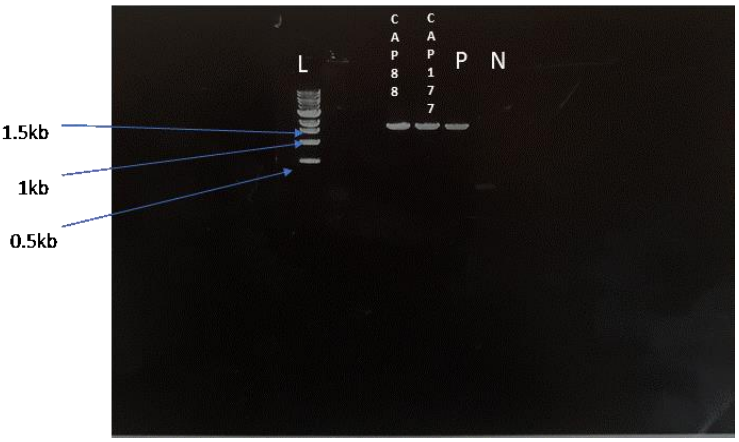

Figure 8: PCR products after agarose gel electrophoresis

An example of the PCR products from two participants visualised using agarose gel electrophoresis. The two participant fragments (CAP88 and CAP177) illustrate the correct DNA size. L= DNA ladder, CAP177/CAP88 = CAPRISA samples, P=positive control, N= negative control. Arrows show the different fragment sizes.

3.2. Colony PCR after cloning and bacterial transformation

Colony PCR was performed to assess if the transformed *E. coli* bacteria contained plasmids with the correct size DNA (*IGHA1*, *IGHA2* and *IGHG1*) insert. Colony PCR utilizes forward and reverse primers that bind to the plasmid on either side of the DNA insert in order to amplify

the DNA insert. Following amplification, the size of the insert was visualized using an agarose gel (Figure 9). Only the colonies with the correct size insert were miniprepped. Colonies that did not have the correct size insert were discarded and new colonies were picked and amplified. From the colonies with the correct DNA size, six and four plasmids were Sanger sequenced for each participant (n=6) for *IGHA* and *IGHG1* inserts, respectively.

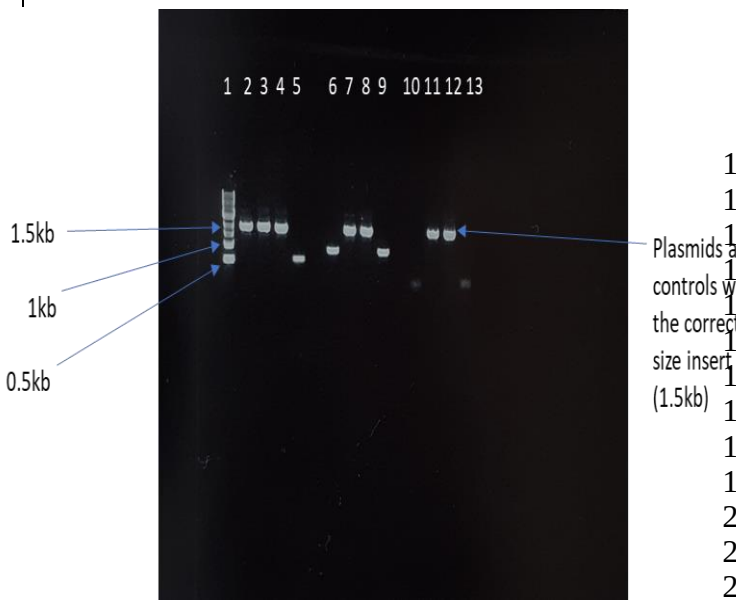

Figure 9: Amplicons generated using colony PCR, following agarose gel electrophoresis

Colony PCR result showing the DNA fragments (*IGHA1*, *IGHA2* and/or *IGHG1*) of the correct size amplified from five participants, CAP88, CAP177, CAP206, CAP255 and CAP256. Each participant had 2 fragments. Lane 1= DNA ladder of different sizes compared to plasmids screened by colony PCR to identify DNA inserts of 1.5kb. Lane 2 and 3 =CAP88 (*IGHA1*), lane 4 and 5= CAP177 (*IGHA1*), lane 6 and 7= CAP206 (*IGHA2*), lane 8 and 9=CAP255 (*IGHA2*), lane 10 and 11=CAP256 (*IGHG1*). CAP = CAPRISA samples/plasmids, 12 = positive control and 13 = negative control.

3.3. Comparison of the *IGHA1* and *IGHA2* sequences to IMGT alleles

All plasmids sequenced in this study were from the six participants. Primers used to amplify the *IGHA1* gene also amplified the *IGHA2* gene as explained in section 2.5, hence the *IGHA2* plasmid results are also reported on. Of the 36 plasmids sequenced for the *IGHA* clones, 24 (67%) matched *IGHA1*, 6 (17%) plasmids matched *IGHA2* and another 6 (16%) did not have a complete coverage because of poor sequence quality in the CH2 exon (Figure 10).

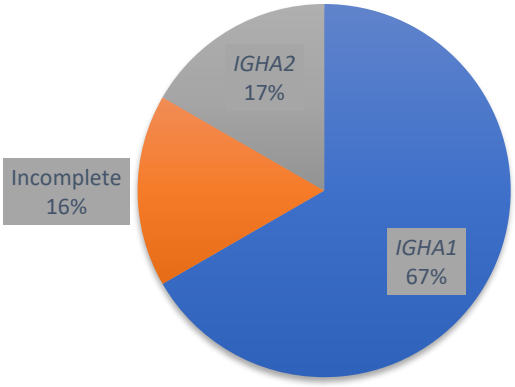

Figure 10: Percentage of plasmids containing *IGHA1* and *IGHA2* DNA inserts from the total *IGHA* amplicons generated

Blue represents *IGHA1* plasmids; Grey: *IGHA2* plasmids and Orange: samples with incomplete data.

None of the alleles sequenced for the *IGHA1* were exact matches to known IMGT alleles and thus likely represent novel alleles. Whereas one *IGHA2* allele was an exact match to *IGHA2*01* IMGT allele. Overall, there were 13 unique alleles detected, eight novel *IGHA1* alleles (identified from 24 clones sequenced) and five novel *IGHA2* alleles (from six clones sequenced).

The closest matching IMGT allele for all novel *IGHA1* alleles was *IGHA1*01* and thus novel alleles were referred to as *IGHA1*01* variant 1-8. The variant alleles were observed across participants, however, one (*IGHA1*01* variant 1) was observed in three participants (CAP88, CAP255 and CAP256). Other variants, such as variant 2 and 4 were observed in CAP177, variant 3 in CAP256, variant 5 and 7 in CAP206, variant 6 in CAP255 and variant 8 was observed in CAP257. The distribution of SNPs in the novel alleles identified in the *IGHA1* gene is highlighted in Figure 11. SNP positions are according to IMGT numbering (53), whereas the amino acid position numbering are based on the Bur numbering system due to the differences in the hinge and/or CH exon lengths between *IGHA1* and *IGHA2* (54). All variants had at least one SNP in intron 1, 5/8 had SNPs in the CH1 exon, 4/8 had SNPs in intron 2 and

3/8 had SNPs in the CH3 exon. Only one variant had a SNP in the CH2 exon, and no SNPs were detected in the CHS or hinge regions.

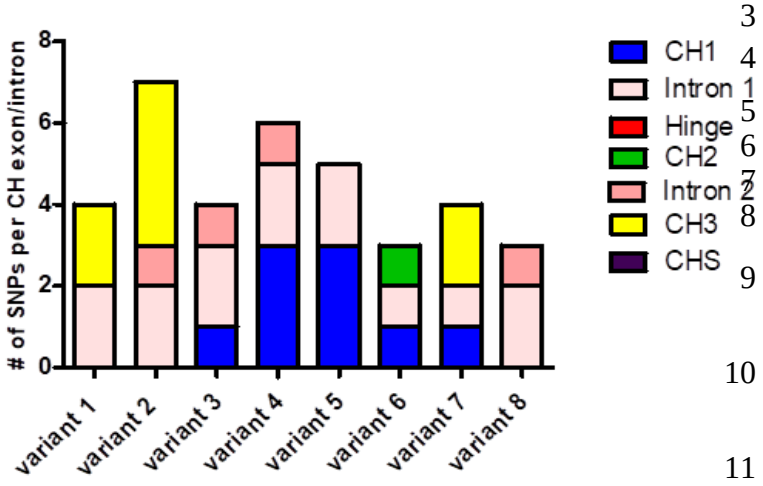

Figure 11 : Distribution of SNPs within IGHA1 genes

Shown are the number of SNPs observed in the CH exons (colours = blue, green and yellow) and introns (colours = pink).

The SNPs that were of particular interest from a functional perspective are those that resulted in amino acid changes within CH domains as they could have an impact on antibody function. Of the 17 SNPs on the CH exons, only 4 resulted in amino acid changes: R392H observed in the CH3 of variant 1, S175R observed in the CH1 of variant 3 and F172L and E295D observed in CH1 and CH2, respectively, of variant 5 (Figure 12).

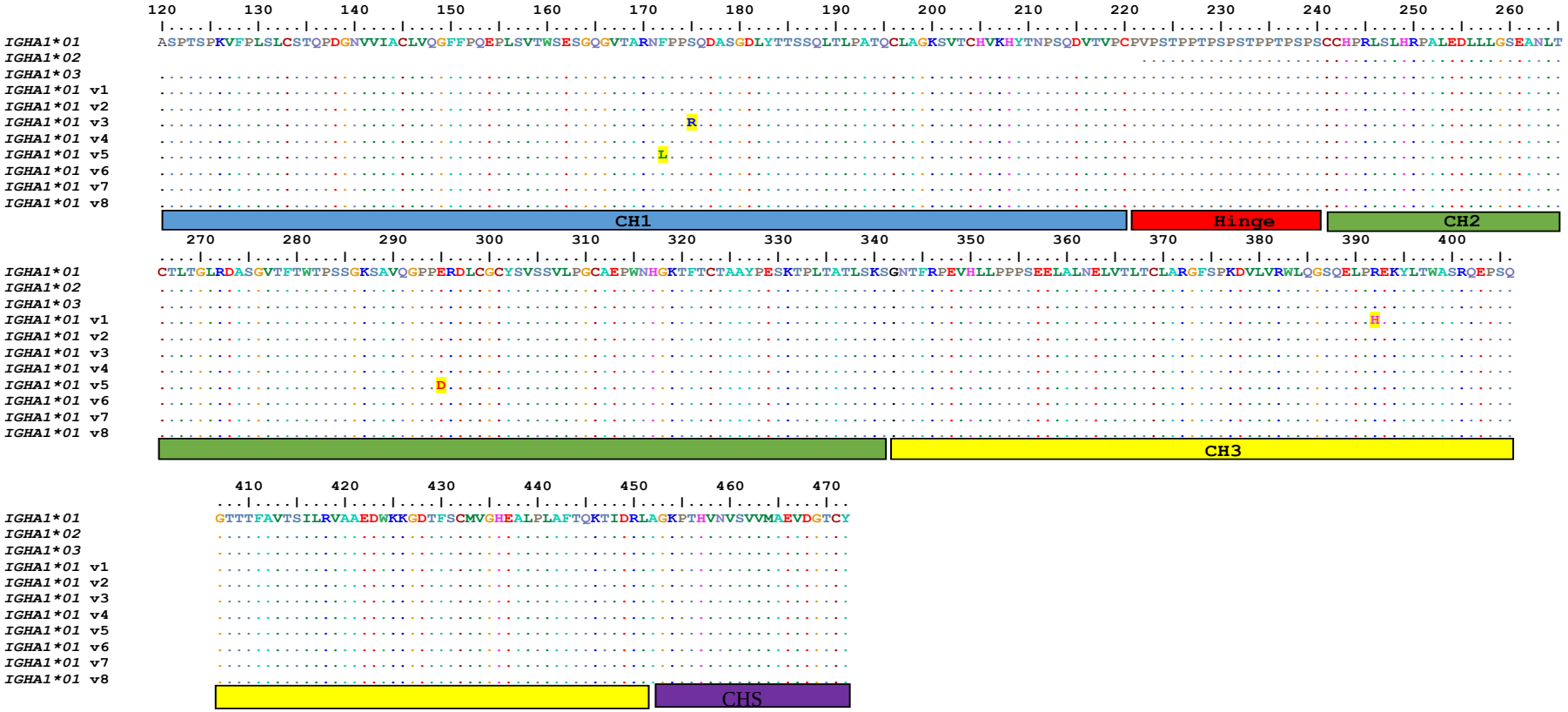

Figure 12: Amino acid alignment of *IGHA1* alleles

Coloured bars indicate the different antibody domains (blue = CH1, red = hinge, green = CH2, yellow = CH3, purple = CHS, the intron is not indicated); highlighted letters indicate amino acid differences between IMGT alleles and/or participant alleles. Highlighted amino acids: L = leucine; R = arginine; D = aspartic acid; H = histidine. Dots indicate that the amino acid on that position was the same between IMGT reference alleles as well as sample alleles.

Of the five novel *IGHA2* alleles, three were more closely related to the *IGHA2*02* reference allele and two were most similar to the *IGHA2*01* IMGT allele. Most alleles had SNPs in the CH1, CH2, CH3 domains and intron 2. Variant 1 was observed in CAP257 and variant 2 in CAP256 had SNPs in the CH2 and CH1 domain only. Two variants (3 and 5) observed in CAP177 and CAP257 had at least six SNPs in the CH1 exon, and at least two SNPs in intron 2 and in the CH3 exon. One variant (observed in CAP88) had SNPs in both the CH2 and CH3 exons only (Figure 13). Overall, none of the novel alleles had SNPs in intron 1, the hinge or the CHS.

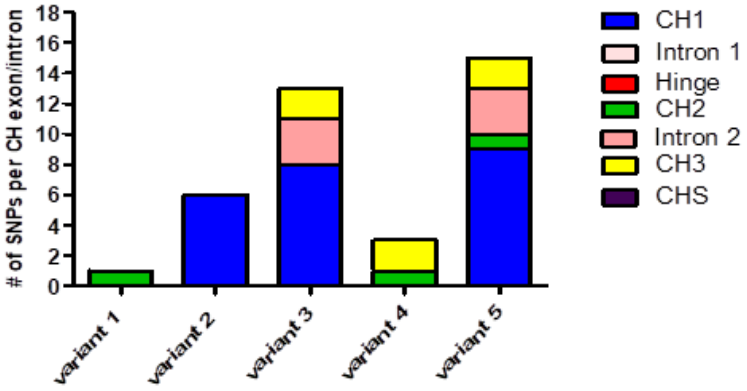

Figure 13: Distribution of SNPs within *IGHA2* genes

Shown are the number of SNPs observed in the CH exons (colours = blue, green and yellow) and introns (colours = pink).

Three variants (2, 3 and 5) shared amino acid substitutions D133C, P136Q, Q137P and V143I, with variant 5 having additional SNPs in the CH1 and CH2 domains (Figure 14). Interestingly, the majority of these amino acid changes match the beginning of *IGHA1* rather than *IGHA2* (Figure 15). These sequences therefore could be a result of PCR mediated recombination between the *IGHA1* and *IGHA2* alleles in these participants, which often occurs when incompletely extended DNA fragments anneal to closely related sequences generating recombination between starting templates (55).

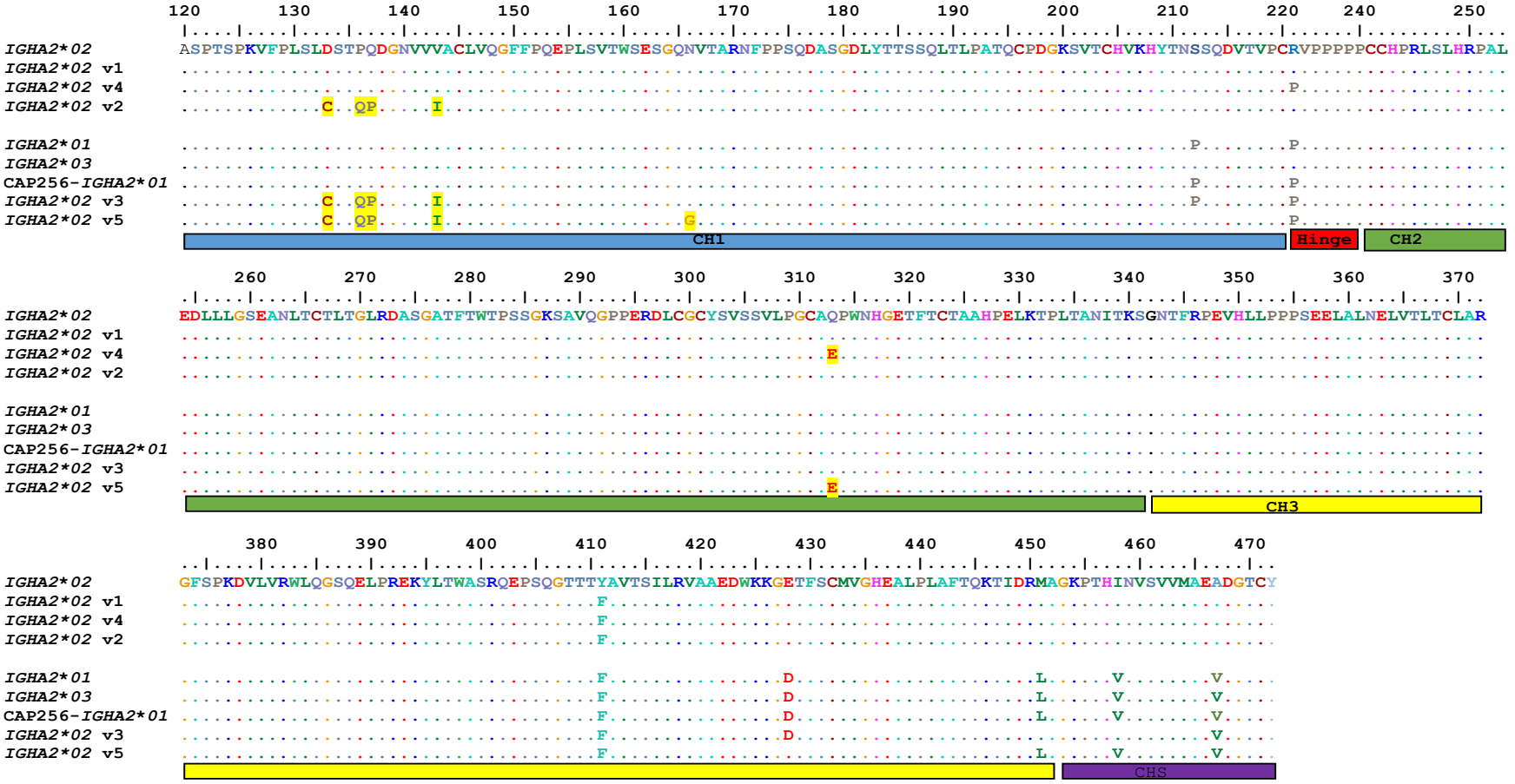

Figure 14: Amino acid alignment of *IGHA2* alleles

Coloured bars indicate the different antibody domains; highlighted letters indicate amino acid differences between IMGT alleles and/or participant alleles. Amino acids that are highlighted: C = cysteine; Q = glutamine; G = glycine; P = proline; I = isoleucine; E = glutamic acid; V = valine. Amino acids in bold: F = phenylalanine; L = leucine; S = serine; R = arginine; D = aspartic acid, Y = tyrosine, M = methionine. Missing numbers are because of the 13 amino acid deletion in the hinge region of the *IGHA2* gene which impacts the numbering of *IGHA2*. Dots indicate that the amino acids on that position are the same between IMGT reference alleles as well as sample alleles.

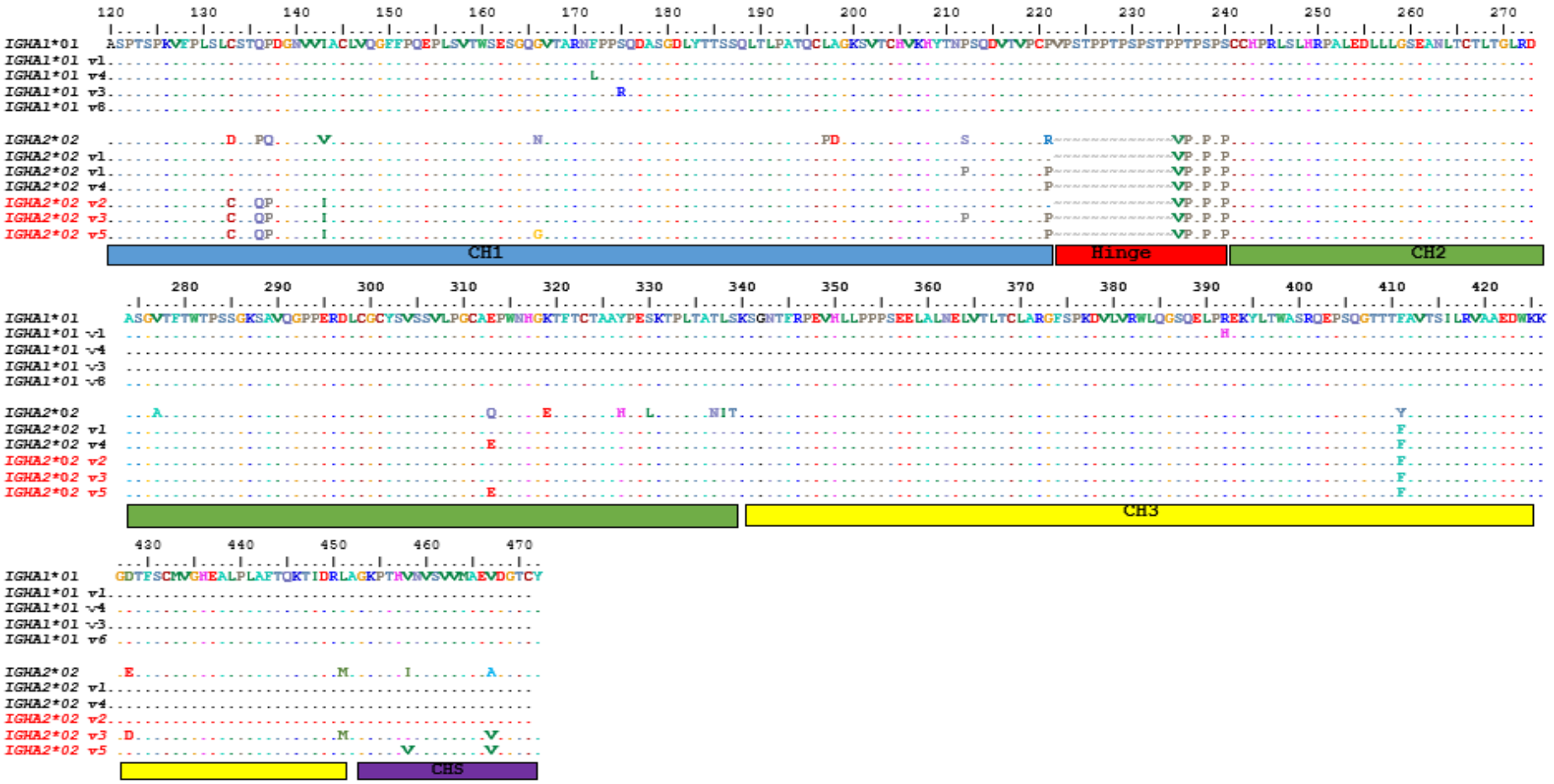

Figure 15: Amino acid alignment of *IGHA1* and *IGHA2* alleles

In this alignment, the 1st five sequences represent the *IGHA1* gene and the last six the *IGHA2* gene. The potential PCR mediated recombination is in the CH1 domain for the alleles shown in red, where the "C, Q, P, I" motif at positions 133, 136, 137 and 143 match those seen in the *IGHA1* alleles rather than the D, P, Q, V motif seen in *IGHA2* alleles. Letters represent amino acids: A=alanine; N=asparagine; C=cysteine; Q=glutamine; G=glycine; P=proline; I=isoleucine; V=valine; F=phenylalanine; L=leucine; S=serine; R=arginine; E=glutamic acid; D=aspartic acid; H=histidine, Y=tyrosine, M=methionine. Dots indicate that the amino acids on that position are the same between IMGT reference alleles as well as sample alleles.

3.4. *IGHA2* Allotypes

Allotypes have been described for the *IGHA2* gene, but not for the *IGHA1* gene. The two allotypes that have been described for *IGHA2* are A2m1 and the A2m2 (37,59). A third allotype (A2n) has been described, this allotype has the same CH1 domain as A2m2, matches both A2m1 and A2m2 in the CH2 domain, interestingly it has a CH3 domain similar to the *IGHA1* gene and A2m1 allotype however, its prevalence in the global population is still being investigated (36,61). The A2m1 allotype has prolines at positions 212 and 221 in the CH1 domain whereas A2m2 has a serine at 212 and an arginine at 221 (37). Of the five variant alleles sequenced for the *IGHA2* gene, two (variant 1 and 2) matched with allotypic determinants of the A2m2 allotype and variant 3 matched with the A2m1 allotype (Table 7). There is a possibility that variants 4 and 5 could be novel allotypes as they had a serine and a proline at positions 212 and 221 respectively.

Table 7: Known *IGHA2* allotypes compared to observed alleles

	Domain: CH1	
Allotype/Sample allele	Amino acid on Position 212	Amino acid on Position 221
A2m1	**Proline**	**Proline**
A2m2	**Serine**	**Arginine**
A2n	**Serine**	**Arginine**
*IGHA2*02* variant 1	Serine	Arginine
*IGHA2*02* variant 2	Serine	Arginine
*IGHA2*02* variant 3	Proline	Proline
*IGHA2*02* variant 4	Serine	Proline
*IGHA2*02* variant 5	Serine	Proline

Amino acids in bold are used to identify and differentiate the known allotypes from each other. Yellow shading represents A2m1 allotypes, orange shading represents A2m2 allotypes and unshaded variants represent a novel allotype.

3.5. Comparison of *IGHG1* sequences to IMGT alleles

Within the *IGHG1* gene, four plasmids were sequenced per participant (n=6) to get a total of 24 sequences. Most of the plasmids sequenced for *IGHG1* suggested that the participants were heterozygous for the gene however, due to the low number of plasmids analysed this could not be determined accurately. From the 24 sequences, ten were exact matches to the *IGHG1*01*

IMGT allele, nine matched the *IGHG1*05* IMGT allele and one matched the *IGHG1*13* IMGT allele. The *IGHG1*01* and the *IGHG1*05* IMGT alleles were the most common alleles identified in these participants (Figure 16). In addition, two novel alleles were identified and are referred to as *IGHG1*01* variant 1 and *IGHG1*01* variant 2 observed in CAP88 and CAP256 respectively (Figure 17). None of the novel alleles were shared between participants and these two alleles each had a silent SNP in the CH2 exon or CH3 exon.

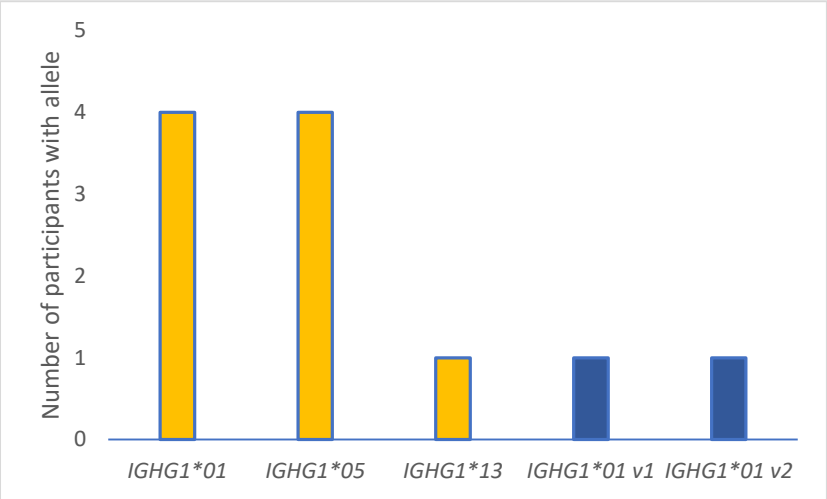

Figure 16: Number of Participants with novel alleles and alleles matching IMGT alleles

Orange bars = IMGT alleles and blue bars = novel alleles.

The two previously described IMGT alleles (*IGHG1*01* and *IGHG1*05*) were observed in four individuals, CAP88, CAP177, CAP206 and CAP256, *IGHG1*13* was observed in CAP257. Additionally, none of the sequenced alleles had SNPs in the introns. The two novel *IGHG1* alleles detected in this study were also identified in the same participants, using SNP detection and Sanger sequencing in collaboration with Dr Mary Carrington from the National Cancer Institute, USA and with PacBio Sequencing in our laboratory as part of another project at the National Institute for Communicable Disease, South Africa.

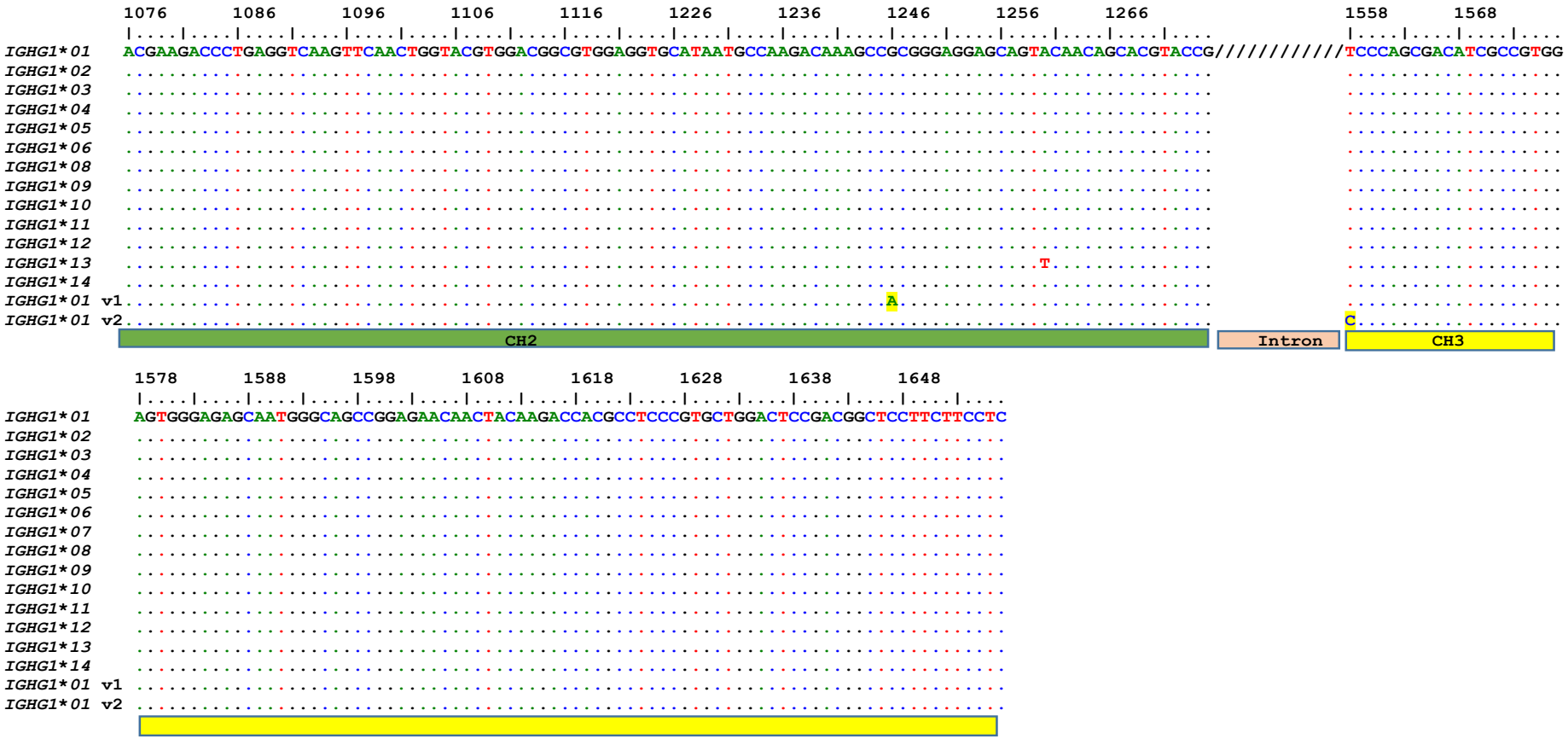

Figure 17: An alignment showing novel SNPs identified in the IGHG1 gene exons

Alignment showing parts of the *IGHG1* gene in which novel SNPs were detected. Coloured bars indicate the different exons; highlighted letters indicate nucleotide differences between IMGT alleles and/or participant alleles. Gene numbering is according to IMGT. Letters represent nucleotides: A = adenine, C = cytosine, G = guanine and T = thymine. Dots indicate that the nucleotides were the same on that position for the IMGT reference alleles and/or the samples.

3.6. *IGHG1* Allotypes

The six known allotypes for IgG1 are the G1m17, G1m3, G1m1, G1m2, G1m27 and the G1m28 (57). Some of these allotypes have been shown to confer altered functional properties such as Fc functions and serum half-life in this antibody isotype (58). For example the G1m17 allotype has been associated with augmented antibody responses against the Hepatitis C virus (58). In this study, all *IGHG1* alleles share allotypic determinants that correlate with the G1m17 allotype. This was confirmed by the presence of a lysine at position 214 in the CH1 domain. Other allotypes and their corresponding amino acids are shown in Table 8.

Table 8: Known allotypes observed in the *IGHG1* translated protein alignments compared to sample alleles

	Domain: CH1		**Domain: CH3**					
Allotype/Sample allele	**214**		**356**	**358**	**422**	**431**	**435**	**436**
	K	R	D	L	I	G	R	Y
G1m17	+	–	–	–	–	–	–	–
G1m3	–	+	–	–	–	–	–	–
G1m1	–	–	+	+	–	–	–	–
G1m2	–	–	–	–	–	+	–	–
G1m27	–	–	–	–	+	–	–	–
G1m28	–	–	–	–	–	–	+	+
*IGHG1*01* variant 1	+	–	–	–	–	–	–	–
*IGHG1*01* variant 2	+	–	–	–	–	–	–	–

Numbers in bold indicate amino acid positions. Letters represent amino acids: K = lysine; G = glycine; I = isoleucine; L = leucine; R = arginine; D = aspartic acid; Y = tyrosine. Orange shading highlights G1m17 allotype.

3.7. Summary of alleles identified in this study

Table 9 is a summary of the alleles observed across the genes analysed in this study, including known IMGT reference alleles (highlighted in yellow) and novel alleles. Overall, there were eight novel alleles identified for the *IGHA1* gene. One of these, variant 1, was found in 3 participants. In the *IGHA2* gene, five novel alleles were identified, these alleles were not shared between participants. In the *IGHG1* gene, there were two novel alleles identified.

Table 9: Summary of alleles observed between the genes

Allele	Closest IMGT allele	Amino Acid change	Number of participants with allele
*IGHA1*01* v1	*IGHA1*01*	R392H	3
*IGHA1*01* v2	*IGHA1*01*	None	1
*IGHA1*01* v3	*IGHA1*01*	S174 R	1
*IGHA1 *01* v4	*IGHA1*01*	None	1
*IGHA1*01* v5	*IGHA1*01*	F171L, E296D	1
*IGHA1*01* v6	*IGHA1*01*	None	1
*IGHA1*01* v7	*IGHA1*01*	None	1
*IGHA1 *01* v8	*IGHA1*01*	None	1
*IGHA2*02* v1	*IGHA2*02*	Y411F	1
*IGHA2*02* v2	*IGHA2*02*	D133C, P136Q, Q137P, V143I, Y411F	1
*IGHA2*02* v3	*IGHA2*02*	D133C, P136Q, Q137P, V143I, Y411F, E428D, A467V	1
*IGHA2*02* v4	*IGHA2*02*	Q313E, Y411F	1
*IGHA2*02 v5	*IGHA2*02*	D133C, P136Q, Q137P, V143I N166G, Q313E, Y411F, M451L, I458V, A467V	1
*IGHA2*01*		None	1
*IGHG1*01* v1	*IGHG1*01*	None	1
*IGHG1*01* v2	*IGHG1*01*	None	1
*IGHG1*01*		None	4
*IGHG1*05*		None	4
*IGHG1*13*		None	1

Letters represent amino acids: A = alanine; N = asparagine; C = cysteine; Q = glutamine; G = glycine; P = proline; I = isoleucine; V = valine; F = phenylalanine; L = leucine; S = serine; R = arginine; E = glutamic acid; D = aspartic acid; H = histidine; Y = tyrosine. Alleles that were 100% matches to IMGT alleles are highlighted in yellow.

4. Discussion

This study investigated the *IGHA1*, *IGHA2* and *IGHG1* immunoglobulin constant region genes in six Black South African women. These genes encode IgA1, IgA2 and IgG1 antibodies which are an important aspect of the humoral immune response in both the systemic and mucosal compartments, protecting against infections of different types in these locations (1,8,20,24,59). Most of the information available concerning these genes on databases such as IMGT was obtained from individuals of European descent or Asian ancestry (60). This research investigated these genes in individuals of African ancestry and identified a total of 15 novel alleles between the three genes. From those, eight had nonsynonymous SNPs which altered amino acids observed in the *IGHA1* and *IGHA2* genes. Furthermore, two alternative approaches from other studies were used to confirm the SNPs identified in the *IGHG1* gene. These novel alleles will be reported to IMGT for inclusion into the database.

A surprising finding in this study was the identification of alleles that appeared to have half of the *IGHA1* CH1 domain and half of the *IGHA2* CH1 domain. This may be due to a recombination event during the PCR reaction. Such recombination events usually occur in PCR reactions that have a high number of PCR cycles, or those that use a high concentration of template and a low primer to target amplicon ratio (55). PCR recombination can be reduced by decreasing the number of PCR cycles and increasing the extension time, using a different enzyme with higher fidelity and template specific primers as well as through replicates (67). Although there will be some consistent bias in terms of where the recombination will be observed in relation to previous experiments, the rate of recombination will not be significant, however, working with complex genes increases the possibility of recombination by ~30% (67). Though it is likely that these variants (*IGHA2*02* v2, v3 and v5) occurred as a result of PCR crossover, it is also possible that these represent novel alleles which arose evolutionary

by recombination, unequal crossover or looping-out excision during duplications and deletions of these genes, as has been previously observed within the IGH locus (29,38,61). An example of this involves the CHS of *IGHA2*01* and *IGHA2*03* which matches the CHS of *IGHA1*01* rather than *IGHA2*02* (61). *IGHA1* and *IGHA2* genes are ~90% similar (36), other unique features (such as N-linked glycosylation motifs and the 13 amino acid extension in *IGHA1*) as explained in chapter 3 (section 3.3) distinguish the two genes from each other and these can be used as guides when designing new primers. Therefore, future studies involving the use of primers specific for unique gene specific regions within the pre-CH1, CH2 and CHS domains of *IGHA1* and *IGHA2* genes could be used to determine whether these alleles came about through PCR recombination or represent novel alleles.

At the amino acid level, the changes observed in the alleles across the heavy chain domains, suggested that the CH2 domain and CHS are conserved compared to other domains across the genes. The CH2 domain is responsible for triggering effector functions by the different antibody isotypes (19,20,22,62). Therefore, conserving the CH2 domain may be an important means of maintaining the function and stability of an antibody molecule (27,63). This also applies to the CHS which serves to allow for polymerisation of the IgA molecules, thus creating multiple binding sites for effective clearing of antigens and pathogens (1,15,18). This study did not identify novel nucleotide or amino acid substitutions in the tailpiece, in line with this need for high levels of conservation. However, we note that this study involved a very small sample size and would therefore require a larger study to confirm this. Of note though was a single amino acid change (R392H) located within the CH3 domain of the *IGHA1*01* v1 allele which was shared in 50% of the participants (3/6) suggesting that there may be population specific differences and therefore warrants further investigation.

Additionally, a novel IgA2 allotype was identified, this allotype encodes a serine and proline at positions 212 and 221 respectively, which was observed in two participants. Both known

IgA2 allotypes (A2m1 and A2m2) which were also observed in our population have been associated with different disease outcomes. The A2m1 allotype, which has prolines at both positions has been associated with a higher risk of IgA nephropathy, whereas having serine and arginine at the same positions (A2m2) is associated with lower risk for the disease (56,64). Thus, understanding the effects of a serine, proline allotype would be of interest. This could be done by assessing the relationship between particular genotypes, and the susceptibility of individuals to particular diseases or conditions. However such studies would rely on genotyping and phenotyping in large numbers of donors to establish such effects.

All *IGHG1* alleles matched the G1m17 allotype (37,68), which is unsurprising as this allotype is extensively found in many populations (65). This allotype in association with a kappa light chain allotype Km1 has been associated with an increased acquisition of HIV (66). Additionally, G1m and Km allotypes may play a significant role in IgG subclass distribution, as G1m carriers seem to have more IgG1 antibodies (58). Furthermore, studies have highlighted the importance of G1m allotypes in enhancing ADCC and ADCP, which are important in the response to infection, vaccination or cancer therapy (8,20,58,65). The overall influence of different allotypes on vaccine efficacy however remains unclear (58). Furthermore, non-synonymous SNPs, and/or deletions in these loci have been associated with recurrent infections from bacteria, viruses and fungi as well as autoimmune diseases (such as celiac disease and systemic lupus erythematosus) which can have severe health implications (1,30). The advancement of DNA sequencing technologies will now make it easy to identify these associations and lead to better treatment and management of patients.

Our study has highlighted the high diversity of antibody genes, the impact of which is important to understand when designing globally efficacious vaccines, autoimmune and cancer therapies. This study had limitations such as a small sample size and use of primers that could amplify

both the *IGHA1* and the *IGHA2* genes. Future studies should include substantially more donors in order to determine how common these novel alleles are, to identify further novel alleles and to accurately assess the diversity of the *IGHC* genes. It will also be important to have gene specific primers and avoid using primers that can amplify two genes, as this will increase accuracy in gene analysis, and reduce issues such as PCR mediated recombination. Furthermore, future studies should assess the impact of these novel mutations, by comparing the function of antibodies with the same variable (VDJ) regions but different constant regions involving the known and novel alleles.

5. Conclusion

In this small pilot study, involving 6 Black South African donors, a total of 15 novel *IGHC* alleles were identified within the *IGHA1*, *IGHA2* and *IGHG1* genes as well as a novel IgA2 allotype. These data highlight the high level of genetic diversity within this locus and the need to study more globally representative populations. Importantly, this genetic diversity could have functional impact on antibody immune responses. Furthermore, the understanding of *IGHA1* and *IGHA2* genes has previously led to the design of antibodies that have an increased serum half-life (36). Thus, having a greater understanding of global antibody genetic diversity could lead to the development of more effective monoclonal antibody therapeutics with robust effector functionality. Overall, these results have implications for Fc effector functions, vaccine development and monoclonal antibody therapies and thus further investigations are required to have a complete understanding of the diversity of the *IGHC* genes.

6. References

1. Owen JA, Jenni Punt SAS & Stanford PPJ. (2013). Kuby Immunology. 7th edtion. New York: W. H. Freeman.

2. Pabst O. New concepts in the generation and functions of IgA. *Nature Reviews Immunology*. 2012;12(12):821–32. https://doi/10.1038/nri3322

3. CDC, Ncird. Immunology and Vaccine-Preventable Diseases – Pink Book – Pneumococcal Disease. *Morbity and mortality weekly report*. 2011;59:279–96.

4. Gaughan CL. The present state of the art in expression, production and characterization of monoclonal antibodies. *Journal of Molecular Diversity*. 2016;20(1):255–70. https://doi/10.1007/s11030-015-9625-z

5. Pavlou AK & Belsey MJ. The therapeutic antibodies market to 2008. *European Journal of Pharmaceutics and Biopharmaceutics*. 2005;59(3):389–96. https://doi/10.1016/j.ejpb.2004.11.007

6. Haneberg B, Dalseg R, Wedege E, Høiby EA, Haugen IL & Oftung F. Intranasal Administration of a Meningococcal Outer Membrane Vesicle Vaccine Induces Persistent Local Mucosal Antibodies and Serum Antibodies with Strong Bactericidal Activity in Humans. *Journal of Infection and Immunity*. 1998;66(4):1334–41. https://doi/10.1128/IAI.66.4.1334-1341.1998

7. Wilson NA, Reed J, Napoe GS, Piaskowski S, Szymanski A & Furlott J. Vaccine-Induced Cellular Immune Responses Reduce Plasma Viral Concentrations after Repeated Low-Dose Challenge with Pathogenic Simian Immunodeficiency Virus SIVmac239. *Journal of Virology*. 2006;80(12):5875–85. https://doi/10.1128/JVI.00171-06

8. Lopez E, Shattock RJ, Kent SJ & Chung AW. The Multifaceted Nature of Immunoglobulin A and Its Complex Role in HIV. *Journal of AIDS Research and Human Retroviruses*. 2018;34(9):727–38. https://doi/10.1089/aid.2018.0099

9. Organization WH. Immunization coverage. Fact sheet N 378. 2015 [cited 2019 Dec 19]. p. 1. Available from: http://www.who.int/mediacentre/factsheets/fs378/en/

10. World Health Organization. Weekly Epidemiology Record. *Releve epidemiologique hebdomadaire*. 2017;92(31):417–35.

11. WHO Publication. Rubella vaccines: WHO position paper—Recommendations. *Vaccine*. 2011;29(48):8767–8. https://doi/10.1016/j.vaccine.2011.08.061

12. World Health Organization. Polio vaccines: WHO position paper, March 2016–recommendations. *Vaccine*. 2017;35(9):1197–9. https://doi/10.1016/j.vaccine.2016.11.017

13. Design of vaccine efficacy trials to be used during public health emergencies -Points of considerations and key principles 2. [cited 2020 Nov 15]. Available from: https://www.who.int/docs/default-source/blue-print/working-group-for-vaccine-evaluation-(4th-consultation)/ap1-guidelines-online-consultation.pdf

14. French MA, Tjiam MC, Abudulai LN & Fernandez S. Antiviral functions of human immunodeficiency virus type 1 (HIV-1)-specific IgG antibodies: Effects of antiretroviral therapy and implications for therapeutic HIV-1 vaccine design. *Frontiers in Immunology*. 2017;8:1–13. https://doi/10.3389/fimmu.2017.00780

15. Woof JM & Kerr MA. The function of immunoglobulin A in immunity. *Journal of Pathology*. 2006;208(2):270–82. https://doi//10.1002/path.1877

16. Xu H, Wang X & Veazey RS. Mucosal immunology of HIV infection. *Immunological*

Reviews. 2013;254(1):10–33. https://doi/10.1111/imr.12072

17. Kelly BT & Grayson MH. Immunoglobulin E, what is it good for? A*nnals of Allergy, Asthma & Immunology*. 2016;116(3):183–7. https://doi/10.1016/j.anai.2015.10.026

18. de Sousa-Pereira P & Woof JM. IgA: Structure, Function, and Developability. *Antibodies*. 2019;8(4):57. https://doi/10.3390/antib8040057

19. Hansen IS, Baeten DLP & den Dunnen J. The inflammatory function of human IgA. *Journal of Cellular and Molecular Life Sciences*. 2019;76(6):1041–55. https://doi/10.1007/s00018-018-2976-8

20. Vidarsson G, Dekkers G & Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology*. 2014;5:1–17. https://doi/10.3389/fimmu.2014.00520/abstract

21. Bournazos S & Ravetch J V. Fcγ Receptor Function and the Design of Vaccination Strategies. *Journal of Immunity*. 2017;47(2):224–33. https://doi/ 10.1016/j.immuni.2017.07.009

22. de Taeye SW, Rispens T & Vidarsson G. The Ligands for Human IgG and Their Effector Functions. *Antibodies*. 2019;8(2):30. https://doi/10.3390/antib8020030

23. Hovenden M, Hubbard MA, AuCoin DP, Thorkildson P, Reed DE & Welch WH. IgG Subclass and Heavy Chain Domains Contribute to Binding and Protection by mAbs to the Poly γ-D-glutamic Acid Capsular Antigen of Bacillus anthracis. Roy CR, editor. *PLoS Pathogen*. 2013;9(4):e1003306. https://doi/10.1371/journal.ppat.1003306

24. Vidarsson G, van der Pol W-L, van den Elsen JMH, Vilé H, Jansen M & Duijs J. Activity of Human IgG and IgA Subclasses in Immune Defense Against Neisseria meningitidis Serogroup B. *Journal of Immunology*. 2001;166(10):6250–6.

https://doi/10.4049/jimmunol.166.10.6250

25. Lu LL, Suscovich TJ, Fortune SM & Alter G. Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology*. 2018;18(1):46–61. https://doi/10.1038/nri.2017.106

26. Richardson, SI 2018, 'Defining The Role Of Fc Effector Function in Natural HIV-1 Infection',PhD thesis, University of the Witwatersrand, Johannesburg.

27. Schroeder HW & Cavacini L. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*. 2010;125(2):S41–52. https://doi/10.1016/j.jaci.2009.09.046

28. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M & Schein J. Complete Haplotype Sequence of the Human Immunoglobulin Heavy-Chain Variable, Diversity, and Joining Genes and Characterization of Allelic and Copy-Number Variation. *The American Journal of Human Genetics*. 2013;92(4):530–46. https://doi/10.1016/j.ajhg.2013.03.004

29. Bottaro A, Cariota U, DeMarchi M & Carbonara AO. Pulsed-field electrophoresis screening for immunoglobulin heavy-chain constant-region (IGHC) multigene deletions and duplications. *The American Journal of Human Genetics*. 1991;48(4):745–56.

30. Watson CT & Breden F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Journal of Genes and Immunity*. 2012;13(5):363–73. https://doi/10.1038/gene.2012.12

31. Ruiz M, Pallarès N, Contet V, Barbié V & Lefranc MP. The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. *Journal of Experimental and*

Clinical Immunogenetics. 1999;16(3):173–84.

32. Lefranc M-P. Nomenclature of the Human Immunoglobulin Heavy (IGH) Genes. *Journal of Experimental and Clinical Immunogenetics*. 2001;18(2):100–16. https://doi/10.1159/000049189

33. Beckmann JS, Estivill X & Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews Genetics*. 2007;8(8):639–46. https://doi/10.1038/nrg2149

34. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M & Schein J. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *The American Journal of Human Genetics*. 2013;92(4):530–46. https://doi/10.1016/j.ajhg.2013.03.004

35. Allotype - an overview | ScienceDirect Topics. [cited 2020 Nov 18]. Available from: https://www.sciencedirect.com/topics/medicine-and-dentistry/allotype

36. Lombana TN, Rajan S, Zorn JA, Mandikian D, Chen EC & Estevez A. Production, characterization, and in vivo half-life extension of polymeric IgA molecules in mice. *MAbs*. 2019;11(6):1122–38. https://doi/10.1080/19420862.2019.1622940

37. Warrender AK & Kelton W. Beyond Allotypes: The Influence of Allelic Diversity in Antibody Constant Domains. *Frontiers in Immunology*. 2020;11:2016. https://doi/10.3389/fimmu.2020.02016/full

38. Jefferis R & Lefranc M-P. Human immunoglobulin allotypes. *MAbs*. 2009;1(4):332–8. https://doi/abs/10.4161/mabs.1.4.9122

39. Owen JA, Punt J, Stranford SA, Jones PP, Kuby J. Kuby immunology LK -

https://wits.on.worldcat.org/oclc/848210406. 7th ed. TA - TT -. New York: W.H. Freeman; 2013.

40. Pandey JP, Namboodiri AM, Luo Y, Wu Y, Elston RC & Thomas DL. Genetic Markers of IgG Influence the Outcome of Infection with Hepatitis C Virus. *The Journal of Infectious Diseases*. 2008;198(9):1334–6. https://doi/10.1086/592282

41. Giha HA, Nasr A, Iriemenam NC, Arnot D, Troye-Blomberg M & Theander TG. Antigen-specific influence of GM/KM allotypes on IgG isotypes and association of GM allotypes with susceptibility to Plasmodium falciparum malaria. *Malar Journal*. 2009;8(1):306. https://doi/10.1186/1475-2875-8-306

42. Durandy A & Kracker S. Immunoglobulin class-switch recombination deficiencies. *Arthritis Research and Therapy*. 2012;14(4):218. https://doi/10.1186/ar3904

43. Calonga-Solís V, Malheiros D, Beltrame MH, Vargas L de B, Dourado RM & Issler HC. Unveiling the Diversity of Immunoglobulin Heavy Constant Gamma (IGHG) Gene Segments in Brazilian Populations Reveals 28 Novel Alleles and Evidence of Gene Conversion and Natural Selection. *Frontiers in Immunology*. 2019;10. https://doi/10.3389/fimmu.2019.01161/full

44. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L & Gray CM. Establishing a Cohort at High Risk of HIV Infection in South Africa: Challenges and Experiences of the CAPRISA 002 Acute Infection Study. Ugarte-Gil CA, editor. *PLoS One*. 2008;3(4):e1954. https://doi/10.1371/journal.pone.0001954

45. Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C & Mansoor LE. Effectiveness and Safety of Tenofovir Gel, an Antiretroviral Microbicide, for the Prevention of HIV Infection in Women. *Science*. 2010;329(5996):1168–74. https://doi/10.1126/science.1193748

46. Scheepers C, Shrestha RK, Lambson BE, Jackson KJL, Wright IA & Naicker D. Ability To Develop Broadly Neutralizing HIV-1 Antibodies Is Not Restricted by the Germline Ig Gene Repertoire. *The Journal of Immunology*. 2015;194(9):4371–8. https://doi/10.4049/jimmunol.1500118

47. Polymerase Chain Reaction - an overview | ScienceDirect Topics. [cited 2020 Feb 29]. Available from: https://www.sciencedirect.com/topics/neuroscience/polymerase-chain-reaction

48. Agarose gel electrophoresis. [cited 2020 Mar 2]. Available from: https://geneticeducation.co.in/agarose-gel-electrophoresis/

49. Motohashi K. A novel series of high-efficiency vectors for TA cloning and blunt-end cloning of PCR products. *Scientifc Reports*. 2019;9(1):6417. https://doi/10.1038/s41598-019-42868-6

50. Romanos MA, Scorer CA & Clare JJ. Foreign gene expression in yeast: a review. *Yeast*. 1992;8(6):423–88. https://doi/10.1002/yea.320080602

51. Addgene: Protocol - How to Ligate Plasmid DNA. [cited 2020 Mar 2]. Available from: https://www.addgene.org/protocols/dna-ligation/

52. DNA sequencing | Learn Science at Scitable. [cited 2020 Feb 29]. Available from: https://www.nature.com/scitable/definition/dna-sequencing-205/

53. Lefranc M-P & Lefranc G. The immunoglobulin factsbook. London: Academic Press; 2001.

54. Putnam F, Sheng Y, Liu V & Low T. Primary structure of a human IgA1 the alpha 1 heavy chain . Primary Structure of a Human IgAl Immunoglobulin. *The Journal of Biological Chemistry*. 1979;254(8):2865–74.

55. Lahr DJG & Katz LA. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques*. 2009;47(4):857–66. https://doi/10.2144/000113219

56. Crowley-Nowick PA, Julian BA, Wyatt RJ, Galla JH, Wall BM & Warnock DG. IgA nephropathy in Blacks: Studies of IgA2 allotypes and clinical course. *Kidney International*. 1991;39(6):1218–24. https://doi/10.1038/ki.1991.154

57. IMGT Repertoire (IG and TR). [cited 2020 Oct 7]. Available from: http://www.imgt.org/IMGTrepertoire/Proteins/allotypes/human/IGH/IGHC/G1m_allotypes.html

58. Kratochvil S, McKay PF, Chung AW, Kent SJ, Gilmour J & Shattock RJ. Immunoglobulin G1 Allotype Influences Antibody Subclass Distribution in Response to HIV gp140 Vaccination. *Frontiers in Immunology*. 2017;8(1883). https://doi/10.3389/fimmu.2017.01883/full

59. Davis SK, Selva KJ, Kent SJ & Chung AW. Serum IgA Fc effector functions in infectious disease and cancer. *Immunology and Cell Biology*. 2020;98(4):276–86. https://doi/10.1111/imcb.12306

60. Popejoy AB & Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161–4. https://doi/10.1038/538161a

61. Chintalacharuvu KR, Raines M & Morrison SL. Divergence of human alpha-chain constant region gene sequences. A novel recombinant alpha 2 gene. *Journal of Immunology*. 1994;152(11):5299–304.

62. Breedveld A & van Egmond M. IgA and FcαRI: Pathological Roles and Therapeutic Opportunities. *Frontiers in Immunology*. 2019;10.

https://doi/10.3389/fimmu.2019.00553/full

63. Prabakaran P, Vu BK, Gan J, Feng Y, Dimitrov DS & Ji X. Structure of an isolated unglycosylated antibody CH2 domain. *Acta Crystallographica Section D: Biological Crystallography*. 2008;64(10):1062–7. https://doi/10.1107/S0907444908025274

64. Galla JH. IgA nephropathy. *Kidney International*. 1995;47(2):377–87. https://doi/10.1038/ki.1995.50

65. Lefranc M-P & Lefranc G. Human Gm, Km, and Am Allotypes and Their Molecular Characterization: A Remarkable Demonstration of Polymorphism. In: *Methods in Molecular Biology*. 2012;882:635–80. https://doi/10.1007/978-1-61779-842-9_34

66. Pandey JP, Namboodiri AM, Bu S, De Dieu Tapsoba J, Sato A & Dai JY. Immunoglobulin genes and the acquisition of HIV infection in a randomized trial of recombinant adenovirus HIV vaccine. *Virology*. 2013;441(1):70–4. https://doi/10.1016/j.virol.2013.03.007

67. Jun-Liang Zhou, Jianping Xu, An-Guo Jiao, Li Yang, Jie Chen, Philippe Callac, Yu Liu and Shou-Xian Wang. Patterns of PCR Amplification Artifacts of the Fungal Barcode Marker in a Hybrid Mushroom. Front. Microbiol. 2019;2686. https://doi.org/10.3389/fmicb.2019.02686

68 IMGT Repertoire (IG and TR). [cited 2020 Oct 7]. Available from: http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=chromosomes&species=human&group=IGH

7. Appendix 1

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Post Graduate Affairs)

TO: Mr IT Ncube
School of Pathology
Antibody Immunity Research Unit
National Institute of Infectious Diseases

E-mail: itaincube2@gmail.com

CC: Supervisor: Dr C Scheepers <catherinem@nicd.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2019/10/31

REF: *R14/49*

PROTOCOL NO: **M191075** *(This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)*

PROJECT TITLE: *Characterization of IgA1 and IgG1 constant region diversity in South Africa*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mr IT Ncube

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M191075

NAME: (Principal Investigator)	Mr IT Ncube
DEPARTMENT:	School of Pathology Antibody Immunity Research Unit National Institute of Infectious Diseases
PROJECT TITLE:	Characterization of IgA1 and IgG1 constant region diversity in South Africa
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Laboratory Study
SUPERVISOR:	Dr C Scheepers
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2019/10/31

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **October** and will therefore reports and re-certification will be due early in the month of **October** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date

PLEASE QUOTE THE CLEARANCE CERTIFICATE NUMBER IN ALL ENQUIRIES