KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI, GHANA.

COLLEGE OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY AND BIOTECHNOLOGY

INVESTIGATING THE PHARMACOGENETIC BASIS OF NEVIRAPINE-INDUCED CUTANEOUS REACTIONS AMONG SELECTED HIV PATIENTS IN GHANA.

By

ALI BABA ISSAH

JULY, 2017

INVESTIGATING THE PHARMACOGENETIC BASIS OF NEVIRAPINE-INDUCED CUTANEOUS REACTIONS AMONG SELECTED HIV PATIENTS IN GHANA.

By

ALI BABA ISSAH

A Thesis submitted to the Department of Biochemistry and Biotechnology,

College of Science,

in partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE

(BIOTECHNOLOGY).

JULY, 2017

DECLARATION

I, ALI BABA ISSAH, hereby declare that the contents of this work are the results of my own original research and that, to the best of my knowledge, it contains no materials previously published by another person, nor materials accepted by another university for the award of any degree; except where due acknowledgement has been made in the text.

…...……. ……..

(Signature) (Date)

ALI BABA ISSAH (PG2480814)

Student Name & ID

…...……. …… …...……. ……

(Signature) (Date) (Signature) (Date)

Dr. KWABENA OWUSU DANQUAH **Prof. FAREED K.N. ARTHUR**

Supervisor Supervisor

…...……. ……..

(Signature) (Date)

Dr. PETER TWUMASI

Head of Department

ACKNOWLEDGEMENT

I thank the Almighty God for his protection and provision for my life, without which I would not be alive to undertake this project. Lord, you have been good to me and I am eternally grateful.

I also sincerely wish to express my gratitude toward Lady Reverend Dr. Joy Bruce, for making my stay here in Kumasi worthwhile. I was lost and away from home, but your constant *voice* in my life served as a wonderful guide that always kept me at peace and in check. I honestly do not know how my stay here in Kumasi would have been if I did not have the privilege of knowing you as my Pastor. God richly bless you, Mummy.

I do wish to acknowledge the immense contributions of my supervisors, Dr. Kwabena O. Danquah, and Prof. F.K.N. Arthur, whose guidance and supervision ensured the success of this work. I am very grateful for your support.

Finally, to all my friends and loved ones who have supported me in diverse ways, I say a very big thank you. To First Love Church, Ayeduase City Campus, words cannot express the love we share. You guys are the very blessing of a family that I needed at this time of my life.

DEDICATION

To my lovely mother, Mary Duah. I do not know any other mother who would do what you have already done for me. I love you, Mom and I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADR	Adverse Drug Reaction
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
cART	Combination Antiretroviral Therapy
CD4/8	Cluster of Differentiation 4/8
CDC	Centers for Disease Control
CHRPE	Committee on Human Research, Publication and Ethics
CIRF	Cancer and Infection Research Facility
CPIC	Clinical Pharmacogenetics Implementation Consortium
CYP	Cytochrome P450
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide
EDTA	Ethylenediaminetetraacetic Acid
FDC	Fixed-Dose Combination
GRID	Gay-Related Immune Deficiency
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Syndrome
HLA	Human Leucocyte Antigen
HSR	Hypersensitivity Reaction
HTLV	Human T Cell Lymphotropic Virus
IMGT	ImMunoGeneTics project
IPD	Immuno Polymorphism Database
KNUST	Kwame Nkrumah University of Science and Technology
LAV	Lymphadenopathy-Associated Virus
MHC	Major Histocompatibility Complex
MLR	Mixed Lymphocyte Reaction

NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NNRTI	Non-Nucleoside/Nucleotide Reverse Transcriptase Inhibitor
NRTI	Nucleoside/Nucleotide Reverse Transcriptase Inhibitor
NVP	Nevirapine
PCR	Polymerase Chain Reaction
PI	Protease Inhibitor
RBC	Red Blood Cell
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
SBT	Sanger Sequence-Based Typing
SDS	Sodium Dodecyl Sulphate
SIV	Simian Immunodeficiency Virus
SJS	Stevens-Johnson Syndrome
SNP	Single Nucleotide Polymorphism
SSOP	Sequence-Specific Oligonucleotide Probe
SSP	Sequence-Specific Primer
STI	Sexually Transmitted Infection
TEN	Toxic Epidermal Necrolysis
TNF	Tumour Necrosis Factor
UGT	Uridine 5′-Diphospho-Glucuronosyltransferase
UNAIDS	United Nations Programme on HIV and AIDS
US FDA	United States Food and Drugs Authority
WBC	White Blood Cell
WHO	World Health Organization

ABSTRACT

Nevirapine (NVP) is a potent nonnucleoside reverse transcriptase inhibitor, widely prescribed for the treatment of HIV-1 infections, as part of HAART treatment in developing countries. The use of NVP, however, is linked to some mild to severe adverse drug reactions (ADRs) of which various skin reactions and liver toxicities are notable. Major NVP-related hypersensitivity reactions have been associated with various mutations in cytochrome P450 enzymes and HLA alleles but these associations are yet to make their translations into clinical usefulness. As more pharmacogenetic associations are established, it may soon become necessary that patients be offered genetic testing prior to NVP treatment to reduce their risk of developing these related ADRs — as in the case of Abacavir hypersensitivity. HLA class I and II molecules, by means of their extraordinary high polymorphic nature, have and continue to pose great challenges to high resolution HLA typing using current available methods. In the present study, HLA-A and -B genes from a Ghanaian HIV population were typed by Sanger sequence-based typing (SBT) using locus-specific PCR. Following BLAST analysis in the IPD-IMGT/HLA database, 10% of HLA-A and 28.6% HLA-B samples typed were successfully assigned genotypes at the allele-level, whereas 83.3% HLA-A and 64.3% HLA-B genotypes were successfully assigned to the serotype level. The majority of alleles typed were ambiguous at the allele-level resolution. These results reflect the current state of HLA typing by SBT using locus-specific PCR and further corroborates the need to: a) augment locus-specific PCR with sequence-specific primers (SPPs) and/or sequence-specific oligonucleotide probes (SSOPs), b) incorporate sequences from outside the peptide binding regions for HLA typing, and c) opt for higher DNA sequencing resolving power as provided by next generation sequencing. No significant association was found between alleles observed and NVP-induced rash in this study but allele frequency distribution between the rash and tolerant groups may suggest different alleles are involved in NVP-rash among Ghanaian population — although this requires further testing.

CHAPTER 1

1.0 INTRODUCTION

1.1 BACKGROUND

The Human Immunodeficiency Virus Acquired Immune Deficiency Syndrome (HIV/AIDS) pandemic is undoubtedly one of the world's worst attacks by any singular organism and the fight against it has fallen short by too far a margin. Since its discovery in 1983 (Barré-Sinoussi *et al.*, 1983; Gallo *et al.*, 1983), HIV has continued to spread and still affects some 36.7 million people worldwide, having killed over 35 million people by the end of 2015 (UNAIDS Fact Sheet, 2016). The impact of HIV/AIDS, however, is not equally felt across the world, as different regions are affected differently with the worst hit areas being sub-Saharan Africa, accounting for approximately 70% of the entire people living with the virus (WHO, 2017). In Ghana, the prevalence of HIV hovered around 2.0% as at 2015 with over 270, 000 persons living with HIV and almost 11, 000 AIDS deaths that year alone (Ghana AIDS Commission, 2016). While this is not the highest in the sub-region, it is nonetheless unacceptably high and raises doubts on the nation's ability to eradicate the virus by the end of 2030, as part of its millennium development goals targets.

Meanwhile, treatment of HIV has evolved significantly from single antiretroviral therapy (ART), beginning from 1987 to the present day highly active antiretroviral therapy (HAART), which involves the use of a combination of antiretrovirals (ARVs). HAART, also referred to as combination antiretroviral therapy (cART), became necessary when the virus was found to develop resistance to antiretrovirals when used as monotherapy over long periods (O'brien, 2000; Pennings, 2013). Although treatment regimen varies slightly across different countries and

depends on many mixed factors, as a first-line therapy, HAART treatment typically involves the use of two nucleoside/nucleotide reverse transcriptase inhibitors (NRTI) and one non-nucleoside reverse transcriptase inhibitor (NNRTI) or protease inhibitor (PI) for ART-naïve patients (Battegay, 2014; Gulick, 2007; Pennings, 2013) — and this has been the guideline for HAART treatment in Ghana since its inception in 2003 (National HIV/AIDS/ STI Control Programme, 2008). Other classes of ARVs are usually resorted to only after the patient begins to show signs of resistance to one or more of the first-line drugs (Tang & Shafer, 2012). With the advent of HAART, HIV-1 infection has been transformed from what was once a fatal diagnosis into a chronically managed disease in patients who have access to medication and achieve durable virologic suppression (Palella *et al.*, 1998; Scourfield *et al.*, 2011).

Nevirapine (NVP) is one of the most commonly prescribed antiretrovirals as part of HAART treatment for HIV-1 infections particularly among developing countries (Phillips *et al.*, 2007). Although largely effective, like other non-nucleoside reverse transcriptase inhibitors (NNRTIs), the use of nevirapine occasionally results in adverse drug reactions (ADRs) of which various cutaneous hypersensitivity reactions and liver toxicities are notable (Shenton *et al.*, 2007; Wit *et al.*, 2008). Some of the very serious NVP hypersensitivity ADRs include Stevens - Johnson syndrome and/or Toxic Epidermal Necrolysis (SJS/TEN), which are extreme forms of exanthematous multiform characterized by widespread skin detachment (Bastuji-Garin *et al.*, 2000; Roujeau *et al.*, 2009). SJS/TEN are generally rare occurrences but have rather high mortality rates and often result in a lifelong sequelae (Aihara, 2011).

Various attempts have been made over the years to address the issue of ADRs in general and those induced by HAART treatment have not been left out of the fray. In recent years, a number of pharmacogenetics studies have been conducted in a bid to unravel many so-called idiosyncratic

ADRs in whose category SJS/TEN and other NVP hypersensitivity reactions fall (Rufini *et al.*, 2015). Many associations have been made between, notably, some specific HLA alleles and cytochrome P450 single nucleotide polymorphisms (SNPs), and a host of ADRs including SJS/TEN that result from the use of drugs such as allopurinol (Hung *et al.*, 2005; Somkrua *et al.*, 2011; Tassaneeyakul *et al.*, 2009), carbamazepine (Chung *et al.*, 2004; Hung *et al.*, 2006), and NVP as well (Carr *et al.*, 2013; Ciccacci *et al.*, 2013, 2015; Yuan *et al.*, 2011). Given the widespread use of nevirapine in Ghana, it is necessary to attempt to individualise ART to reduce the risk of NVP-induced rash among patients accessing NVP-based combination therapy — and this is the long term goal of this study.

1.2 PROBLEM STATEMENT

The real problem with NVP hypersensitivity is the severe form in which it can assume especially when it manifests as Stevens-Johnson syndrome and/or Toxic Epidermal Necrolysis (SJS/TEN). The manifestations of SJS/TEN range from mild exanthematous skin rashes to a large amount of bullae and extensive mucocutaneous sloughing (Roujeau, 1994; Wong *et al.*, 1999). In spite of the low incidence, the mortality rate is 5–10% for SJS, about 30% for SJS/TEN overlap, and nearly 50% for TEN (Dao *et al.*, 2015). Unfortunately, a consensus treatment guideline remains unavailable and treatment protocols vary widely across different healthcare providers with varying effectiveness (Aihara, 2011). Although, it can be triggered by infections such as *Mycoplasma pneumonia*, SJS/TEN remain principally drug-induced adverse reactions (Roujeau *et al.*, 2009), thereby leaving risk identification and subsequent drug avoidance in susceptible patients as the most promising approach to addressing the problem.

1.3 HYPOTHESIS AND CONCEPTUAL FRAMEWORK

Quite a number of studies have revealed NVP-related adverse drug reactions to be genetically mediated with both immune and non-immune genes implicated. Some animal model studies reported that NVP-rash is clearly immune-mediated, and partial depletion of CD4+ T cells, but not CD8+ T cells, was protective (Popovic *et al.*, 2010; Shenton *et al.*, 2007). Many pharmacogenetic associations have been made between specific HLA alleles and nevirapine hypersensitivity reactions in recent years. In 2009, Chantarangsu *et al.* and Likanonsakul *et al.* reported an association between nevirapine-induced skin rashes and *HLA-B*35:05* and *HLA-C*04* respectively, in Thai patients. Yuan *et al.* (2011) subsequently reported similar associations between NVP-induced cutaneous adverse events and *HLA-B*35* in Asians (particularly among Thais), and *HLA-C*04* among Whites, Asians and markedly, in Blacks. They also found a significant association between *HLA-DRB1*01* and NVP-hepatotoxicity among Whites, while the same allele had been previously reported to be associated with Nevirapine/efavirenz-induced hypersensitivity reactions in a French cohort population (Vitezica *et al.*, 2008). Furthermore, in 2012, Gao *et al.* also reported associations between NVP-hypersensitivity and *HLA-C*04* in a Han Chinese population while a more specific variant, *HLA-C*04:01*, has been found to increase the risk of NVP-SJS/TEN in Malawian populations (Carr *et al.*, 2013, 2017).

At the same time, *HLA-DRB1*01:02* together with *HLA-B*58:01* have also been reported to be associated with NVP-hepatotoxicity events in a South African population (Phillips *et al.*, 2013).

Additionally, some CYP P450 single nucleotide polymorphisms (SNPs) have also been found to increase the risk of NVP-related toxicities in different populations, with the most notable SNP being CYP 2B6 *516G>T* and *983T>C* (Ciccacci *et al.*, 2013; Yuan *et al.*, 2011) — and there are

yet many other NVP pharmacogenetic associations that have been made with different non-CYP gene mutations/polymorphisms (Ciccacci *et al.*, 2010, 2015; Liptrott *et al.*, 2012).

While these findings appear inconsistent, it supports the notion that different polymorphisms may be responsible for similar NVP-related adverse drug events in different populations and that the associated HLA alleles as well as other gene SNPs are at best necessary but not sufficient for the development of these NVP-ADRs. It is therefore necessary for more NVP pharmacogenetic studies especially in populations that hitherto have not been studied, if any of these associations are to have any significant impact on the healthcare of affected HIV patients — hence the decision to undertake this research in a Ghanaian population.

1.4 AIMS AND OBJECTIVES

The aim of the study was to genotype HLA-A and -B genes in the selected population and to determine possible associations between the HLA alleles observed and NVP-induced rash events.

1.4.1 Specific Objectives

Specific objectives were:

- To determine the prevalence of NVP-induced rash among HIV patients accessing NVP-based combination therapy at the Juaben Municipal Hospital.
- To assess NVP-rash and tolerant groups of patients for possible associations between their socio-demographic or clinical characteristics and NVP-induced rash.
- To determine the HLA-A and -B alleles and genotypes carried by each participant in the selected population by Sanger sequence-based typing (SBT).

- To assess the level of resolution to which the chosen method is able to type the selected HLA loci.
- To analyse the observed HLA alleles for possible association with NVP-induced rash.

1.5 JUSTIFICATION

Knowing the HLA type of an individual has implications for the success of whole organ and hematopoietic stem cell transplantations (Erlich *et al.*, 2001). Specific HLA alleles can indicate a person's elevated risks for or exclusion from certain autoimmune disorders (such as ankylosing spondylitis (Lin & Gong, 2017) and celiac disease (Fasano *et al.*, 2015)), and the development of certain drug adverse reactions. Indeed, knowing the HLA type of an HIV patient has important implications for individualizing antiretroviral therapy for that patient. This idea became undeniably apparent when a strong association between Abacavir hypersensitivity and *HLA-B*57:01* (in mainly Caucasian populations), led to the recommendation of genetic testing for all Caucasian patients by the US FDA prior to treatment to help reduce their risk for Abacavir hypersensitivity (Mallal *et al.*, 2008). A few other pharmacogenetic associations have also resulted in similar recommendations (Center for Drug Evaluation and Research, 2016) with carbamazepine-induced SJS/TEN association with *HLA-A*15:02* in Han Chinese (Chung *et al.*, 2004; Hung *et al.*, 2006) being another example. It is now standard practice for drug companies to include a 'pharmacogenetic warning' section in their drug product inserts to serve as useful information for both physicians and consumers.

Many pharmacogenetic associations are yet to make their translations into clinical usefulness (Pavlos & Phillips, 2012). Nevertheless, given the Abacavir and carbamazepine premise described

above, an association between Nevirapine hypersensitivity and HLA alleles (or any other gene SNP) could potentially have similar impacts, depending on the consistency of findings and the strength of association.

This work is therefore an attempt to further individualize Nevirapine antiretroviral therapy in Ghana using genetic (specifically HLA) data. It will also provide an idea of the suitability for adoption of Sanger sequence-based HLA typing for routine screening of HIV patients prior to HAART treatment in Ghana.

CHAPTER 2

2.0 LITERATURE REVIEW

2.1 HUMAN IMMUNODEFICIENCY VIRUS (HIV)

2.1.1 History and Origin of HIV

Acquired Immune Deficiency Syndrome (AIDS) was first observed in 1981 in the United States when a group of young homosexuals began to present with rare opportunistic infections and malignancies that were previously found only in immunocompromised patients (CDC, 1981; Sharp & Hahn, 2011). As the number of cases grew, the US Centers for Disease Control (CDC) was alerted to set up a task force to monitor the outbreak (CDC, 1982; Durack, 1981). The disease was initially given various names, according to the original discoverers, such as lymphadenopathy, Kaposi's Sarcoma and Opportunistic Infections, GRID (gay-related immune deficiency), etc., until in 1982 when the CDC, discovering that the disease was not restricted to only gay people, officially began to use the name AIDS (CDC, 1982; Greene, 2007).

Around this time, the causative organism was still unknown. However, in 1980, Robert Gallo and colleagues had previously reported the discovery of a retrovirus that infected human T lymphocytes, causing adult T-cell leukaemia, which they named human T cell lymphotropic virus (HTLV) (Poiesz *et al.*, 1980). In 1983, they gave hint of a possible causal agent of AIDS when they reported the discovery of yet another retrovirus they called human T cell lymphotropic virus III (HTLV-III) isolated from an AIDS patient, claiming it was related to their previously discovered HTLV-I (Gallo *et al.*, 1983). Interestingly, in the same year (and in the same issue of the journal *Science*), another team of scientists from the Pasteur Institute, led by Luc Montagnier reported the discovery of a retrovirus isolate from a ‘pre-AIDS’ patient suffering from acute

lymphadenopathy and demonstrating by electron microscopy that the isolate's core proteins differed from that of HTLV-I, named it lymphadenopathy-associated virus (LAV) (Barré-Sinoussi *et al.*, 1983). Nevertheless, it still remained unclear whether any of these isolates were the actual cause of AIDS until subsequent works from Gallo and his colleagues eventually helped establish a causal link in 1984 (Gallo *et al.*, 1984; Popovic *et al.*, 1984; Schüpbach *et al.*, 1984). Later in 1986, these two viruses were discovered to be the same and were subsequently renamed Human Immunodeficiency Virus (HIV) (Coffin *et al.*, 1986).

A number of theories surrounding the origin of HIV exist till date, some of which lend themselves to various conspiracies. However, the most prominent of these theories with enough scientific backing holds that the virus originated from simian immunodeficiency viruses (SIVs) native to African primates through a series of multiple zoonotic processes (Hahn *et al.*, 2000; Sharp & Hahn, 2011). While the exact periods of zoonosis are unclear, phylogenetic evidence suggests that these SIVs which have their natural hosts in African primates have crossed the species barrier at least on seven different occasions, creating slightly different pathogens in the process each time, ultimately leading to HIV (Hahn *et al.*, 2000). The two most important types of HIV, HIV-1 and HIV-2 are believed to have originated from the chimpanzee (SIVcnz) and sooty mangabey (SIVsmm), respectively (Gao *et al.*, 1999; Sharp *et al.*, 2001).

2.1.2 Clinical Importance of HIV

HIV is primarily a blood-borne pathogen principally transmitted through sexual intercourse and to a lesser degree through other body fluids (Gallo & Montagnier, 2003). The ultimate devastation by HIV upon infection is the depletion of CD4+ T cells, thereby compromising the host's cell-

mediated immunity and exposing the patient to various opportunistic infections (Greene, 2007). Untreated HIV infection remains a life-threatening condition with a long incubation period of *averagely* 10 years to progression into AIDS (Hahn *et al.*, 2000; Macher & Goosby, 2016).

HIV/AIDS has since its discovery spread to become the world's biggest pandemic affecting millions of people worldwide. The menace caused by this virus is by far the worst any singular virus or organism has probably done to the human race having infected over 78 million people and killed over 35 million people, since the start of the epidemic (WHO, 2017). By the end of 2015, 36.7 million people were living with HIV and there were 2.1 million new cases of infection that year alone (UNAIDS Fact Sheet, 2016). While the majority of HIV burden is borne by the people of sub-Saharan Africa, with a prevalence of about 1 in every 25 adults affected (WHO, 2017), new trends of infection continue to emerge in other parts of the world, making HIV a persisting global threat to human lives (Hahn *et al.*, 2000). Although the number of people accessing antiretroviral therapy has improved in recent years, from 7.5 million in 2010 to about 18.2 million as of June 2016, there have been no declines in new HIV infections among adults since 2010 —with an estimated 1.9 million new adult infection each year (UNAIDS Fact Sheet, 2016). With the life expectancy of people living with HIV increasing due to improved treatment, the overall number of people living with the virus is expected to rise.

The economic impact of HIV has been enormous costing US$ 19 billion of investments in the AIDS response in low- and middle-income countries alone by the end of 2015 with an estimated US$ 26.2 billion to be required in the year 2020 and US$ 23.9 billion required in 2030 (UNAIDS Fact Sheet, 2016). As the fight against HIV continues and until such a time when a cure or vaccine is discovered, the world will have no choice than to embrace current treatment methods as the best way to tackle this devastating menace by means of therapy.

2.1.3 Classification and Structure of HIV

Two clinically important types of AIDS causing viruses, HIV-1 and HIV-2 have been identified, distinguished by their genome organisation and phylogenetic relationship with other SIVs (Hahn *et al.*, 2000). HIV-1 which is the more virulent type (and the first to be discovered and described by Gallo *et al.* and Barré-Sinoussi *et al.* in 1983), accounts for the vast majority of all global HIV infections, being more widespread compared to the type 2, which is predominantly confined to West Africa (Sharp & Hahn, 2011). HIV-1 is further classified into four groups: M (the major group), N, O (the outlier group), and P; with more than 95 percent of HIV-1 cases involving HIV-1 group M (Daw *et al.*, 2017). Within group M lie 10 separate subtypes: A, B, C, D, F, G, H, J, and K. Also included are CRFs (circulating recombinant forms) which are essentially hybrid subtypes formed from a mixture of two subtypes (Hemelaar *et al.*, 2011; Hemelaar, 2012).

HIV is a member of the lentivirus family of retroviruses (Hahn *et al.*, 2000)— a group of viruses whose genetic material are stored in ribonucleic acids (RNAs) as opposed to deoxyribonucleic acids (DNAs) as do majority of living organisms. HIV's genetic material is stored in two single stranded RNA molecules which are closely knit to a host of viral proteins essential to the survival of the virion particle. This matrix of viral genome and proteins are surrounded by a capsid made up of thousands of copies of the viral protein p7. The capsid is in turn enclosed in a lipid bilayer envelope (taken from a human host cell membrane when new viral particles bud from their infected host cells during replication) studded with a glycoprotein complex comprising of *gp120* and *gp41 glycoproteins* which are important in the attraction and fusion of the virus to host cells.

The RNA viral genome, upon entry into a suitable host cell, is converted into double stranded DNA (dsDNA) by a *reverse transcriptase* enzyme found in the viral capsid and encoded in the viral genome as well. This dsDNA is then transported into the nucleus where it is inserted into the

host genome, aided by another viral encoded enzyme — *integrase*. A *protease* enzyme also plays an important role in the replication of the virus by cleaving proviral proteins into functional subunits for viral capsid formation during or shortly after viral budding from an infected cell. Figure 2.1 below depicts the typical structure of HIV virion.

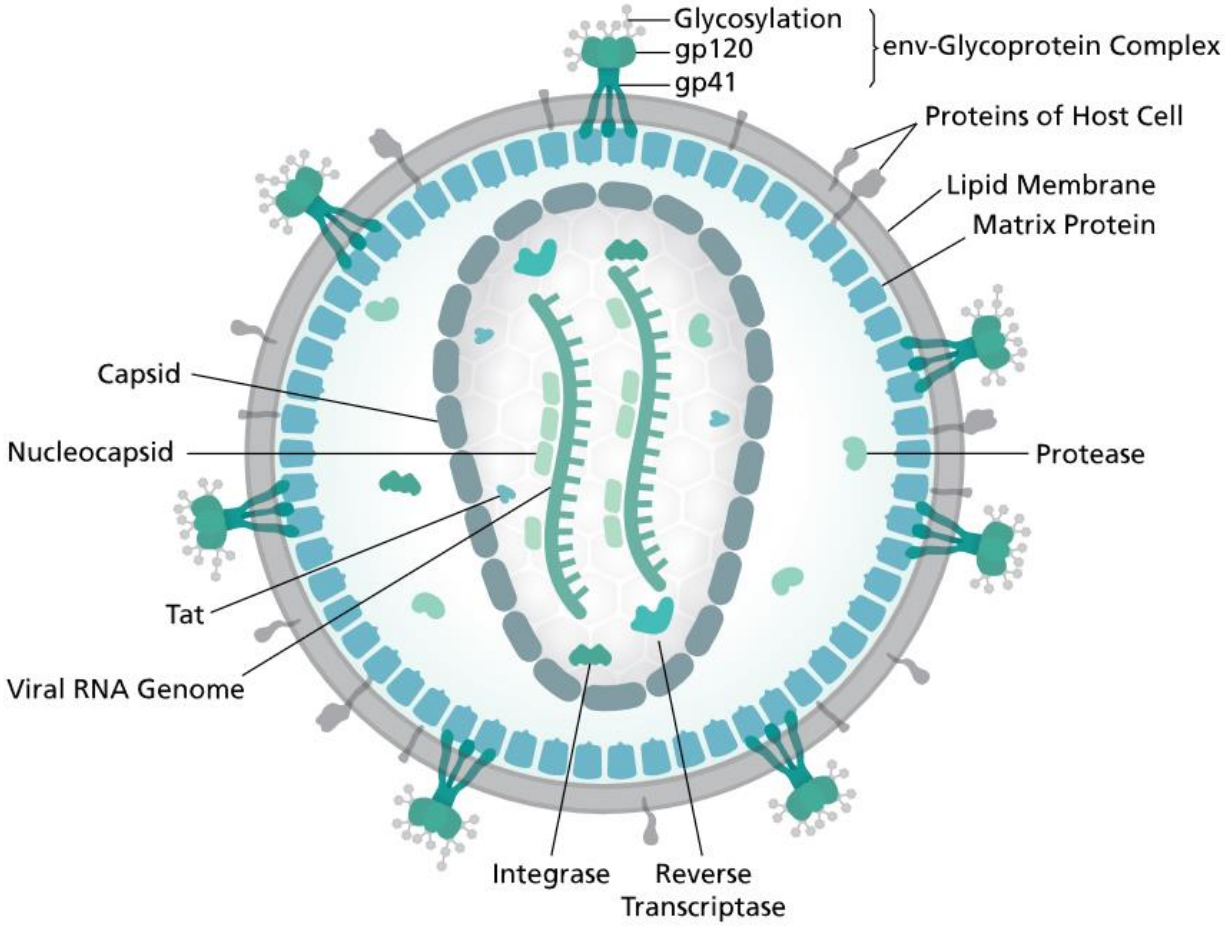


Figure 2.1: Diagram of HIV virion showing essential viral components [Credit: Thomas Splettstoesser (www.scistyle.com)].

Not surprising, it is these important enzymes and proteins essential to the infection, survival and spread of the virus which serve as targets for drug development in the treatment of HIV.

2.1.4 HIV Treatment

Interventions to the HIV pandemic have taken diverse approaches — from drug treatments, to social intervention campaigns toward protection from infection mainly, through sexual intercourse, to prevention of mother-to-child transmission (Cohen *et al.*, 2011). The first drug to be approved for HIV treatment was zidovudine (AZT) in 1987, a nucleoside reverse transcriptase inhibitor (NRTI), which was a high-dose drug that presented problems of inconvenience and intolerance to patients, as well as non-compliance to treatment (Pavlos & Phillips, 2012).

As other drugs (mainly NRTIs) were sequentially developed and used as monotherapy over time, there emerged the issue of HIV drug resistance (O'brien, 2000). HIV replicates at an unusually high rate, and coupled with the error-prone nature of its reverse transcriptase, it evolves at an extraordinarily fast rate (Greene, 2007; Sharp *et al.*, 2001). This attribute implies that occasionally some 'fortunate' viral replicates would possess a mutation that essentially made them resistant to a particular drug and gave them a selective advantage, in spite of treatment with that drug (O'brien, 2000). This is further complicated by high recombination rates between different variants of the virus simultaneously infecting the same cell (Greene, 2007; Tang & Shafer, 2012). This development necessitated the need for more aggressive drug treatments— one that would slow down replication drastically and thereby reduce the rate of development of resistant strains— and this set the stage for highly active antiretroviral therapy, HAART, beginning from 1996 when the first non-nucleoside reverse transcriptase inhibitor, Nevirapine (marketed as Viramune®) was approved (Gulick, 2007; Pennings, 2013).

HAART treatment typically involves the use of three or more antiretroviral drugs from two or more ART classes, commonly from nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) and/or protease inhibitors (PIs) (Battegay,

2014). Other classes of ART are introduced only upon signs of resistance by the virus to the initial drugs and also depending on the specific circumstances of the patient being treated (Gulick, 2007).

By the end of 2016, 30 individual antiretroviral drugs and 14 fixed-dose combination (FDC) drugs from seven ART classes had been approved for clinical use in the treatment of HIV (Figure 2.2).

With the advent of HAART, HIV-1 infection is now manageable as a chronic disease in patients who have access to medication and who achieve durable virologic suppression (Palella *et al.*, 1998; Scourfield *et al.*, 2011), but this success story is not without challenges. While majority of people tolerate current ARTs without any harm, a minority develop some mild to severe adverse drug reactions which remain of real concern to stakeholders in HIV clinical care (Dalal *et al.*, 2015; Tozzi, 2010).

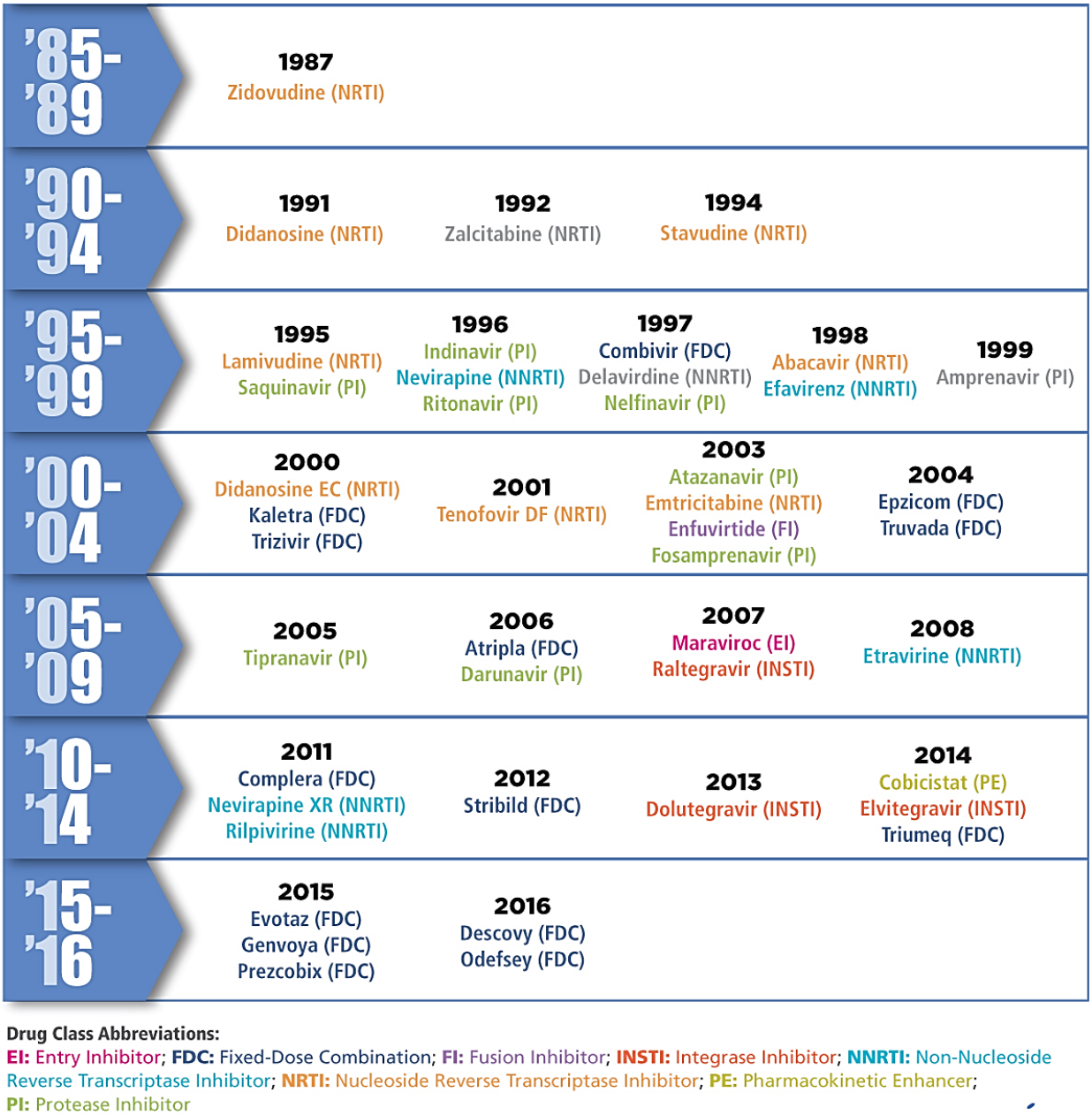


Figure 2.2: A timeline of all FDA approved antiretrovirals by end of 2016, categorized by drug class. (Adapted from https://aidsinfo.nih.gov/understanding-hiv-aids/infographics/25/fda-approval-of-hiv-medicines).

2.2 NEVIRAPINE

2.2.1 Nevirapine Structure and Mechanism of Action

Nevirapine (NVP) is a nonnucleside reverse transcriptase inhibitor (NNRTI) drug used in HAART treatment of HIV-1. It was the first of the NNRTI antiretroviral class to be approved for use in HIV clinical care in 1996 (marketed under the name Viramune®) (Popovic *et al.*, 2010) — to be administered *only* together with at least two other ART drugs. It was not approved for monotherapy because short-term treatment resulted in resistance (Richman *et al.*, 1994), with the most common observed mutation occurring at the reverse transcriptase amino acid residue 181, from tyrosine to cytosine (Havlir *et al.*, 1996).

Nevirapine is a benzodiazepine NNRTI (Pubchem, 2017) (Figure 2.3). In general, NNRTIs are small (<600 Da) hydrophobic molecules that inhibit HIV-1 reverse transcriptase (RT) enzyme noncompetitively by binding to a specific site on the p66 subunit of the HIV-1 RT known as the NNRTI binding pocket (Sluis-Cremer *et al.*, 2004).

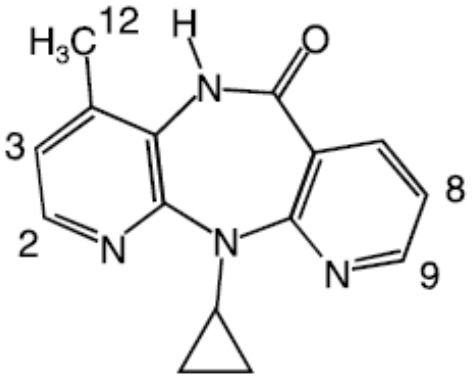


Figure 2.3: Structure of Nevirapine (11-cyclopropyl-4-methyl-5H-dipyrido [2,3-e:2',3'-f][1,4]diazepin-6-one) (Popovic *et al.*, 2010).

Specifically, NVP binds to tyrosines at amino acid residues 181 and 188 of the RT enzyme, inducing a conformational change comprising both short-range and long-range distortions of the RT structure, and thereby limiting its activity (Sluis-Cremer *et al.*, 2004). Nevirapine does not inhibit human DNA polymerase-α, -β, -δ, or -γ (Asmuth & Pollard, 2007), nor is it effective against HIV type 2 because the pocket of HIV-2 reverse transcriptase has a different structure, which confers intrinsic resistance to the drug (Nevimune® product insert, n.d.).

2.2.2 Nevirapine Pharmacokinetics

Nevirapine, following oral administration, is readily absorbed into the bloodstream with over 90% bioavailability (Lamson *et al.*, 1999). It is lipophilic (60% dissolved in plasma) and well distributed in body tissues, freely crossing both the blood brain barrier and the placenta into the brain and breast milk respectively (Nevimune® product insert). NVP pharmacokinetic profile permits a once-daily dosing; however, the established therapeutic dose is 200 mg twice daily, following a 14-day lead-in dose of 200 mg a day (Boehringer-Ingelheim, 2017). Peak plasma concentrations are generally achieved within 4 hours following a single 200 mg dose administration (Lamson *et al.*, 1999), although multiple secondary peaks have also been reported (Cheeseman *et al.*, 1993; Havlir *et al.*, 1995). Steady-state plasma trough concentrations of about 4.0 μg/ml are usually achieved after a period of 28 days and are stable for at least one year of therapy (Chen *et al.*, 2008).

NVP is extensively biotransformed via cytochrome P450 (oxidative) metabolism to 2-, 3-, 8-, and 12- hydroxynevirapine chiefly by CYP 3A4 and 2B6, while other CYP enzymes, CYP3A5, CYP2C9 and CYP2D6 play a minor role (Whirl-Carrillo *et al.*, 2012) (see Figure 2.4).

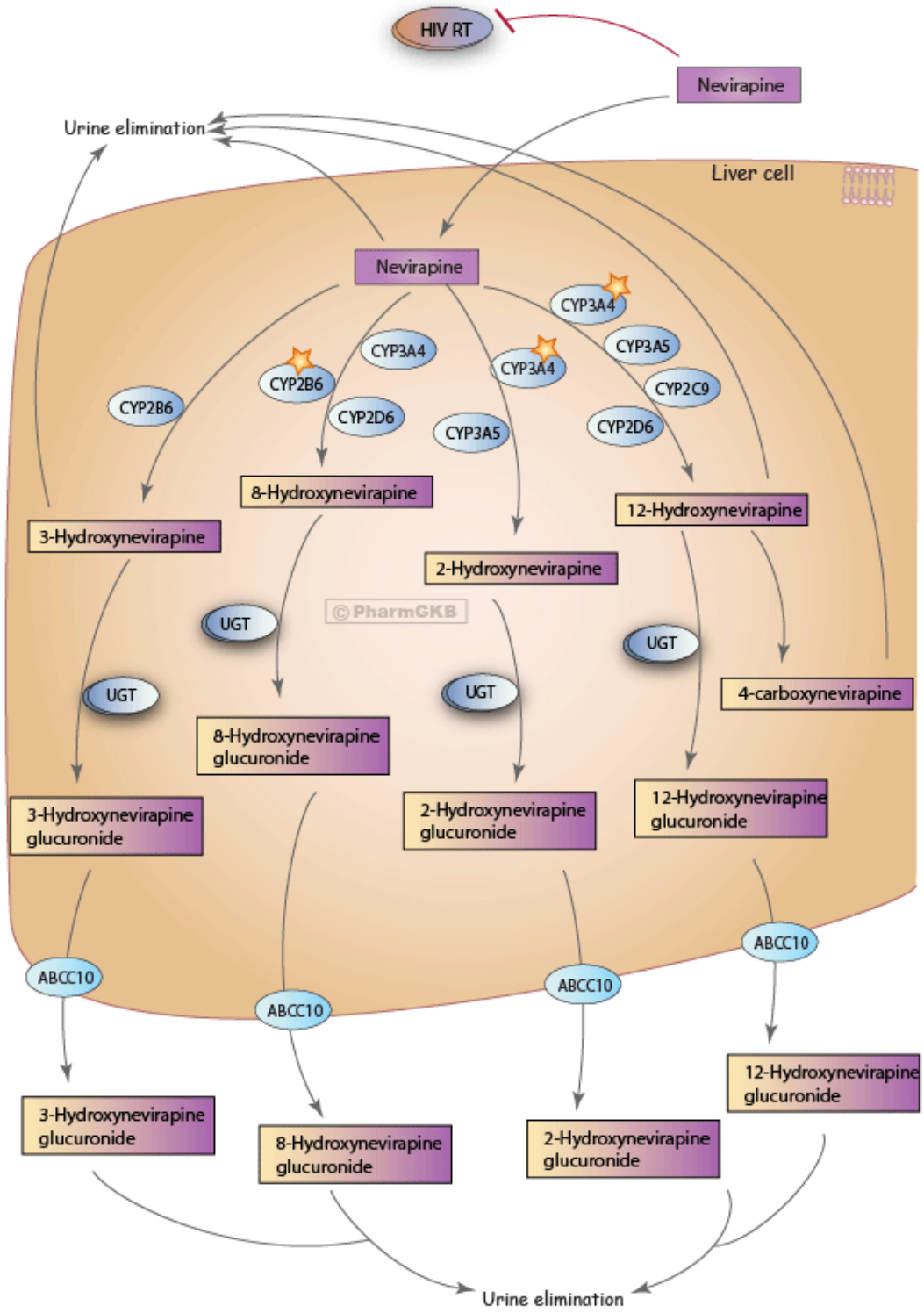


Figure 2.4: Nevirapine metabolic pathway showing key metabolizing genes and their resulting metabolites. [Copyright: PharmGKB (https://www.pharmgkb.org/pathway/PA165950411). Used with permission from PharmGKB and Stanford University.]

This is followed by glucuronidation of these hydroxyl metabolites by UGT enzymes and urinary excretion, which is the principal means of elimination with more than 80% of the urinary elimination in the form of glucuronide conjugates of hydroxylated metabolites (Riska *et al.*, 1999). Only a small fraction (< 5%) is excreted unchanged in urine (representing < 3% of the total dose) (Riska *et al.*, 1999).

Although little is known about the transporters that influence the disposition of NVP, drug transporters expressed in key target tissues, such as *ABCB1* and *ABCC10* (ATP-binding cassette proteins) have been implicated in the efflux process of NVP metabolites (Ciccacci *et al.*, 2010; Liptrott *et al.*, 2012).

2.2.3 Nevirapine Hypersensitivity

Nevirapine, like other NNRTIs, is known to cause some varied adverse drug reactions (ADRs) in a minority of patients, collectively termed nevirapine hypersensitivity reactions (HSRs). Prior to marketing, Boeringer-Ingelheim performed a series of clinical trials which concluded that NVP can lead to both skin rash and liver toxicities in patients (Pollard *et al.*, 1998). The frequencies of these ADRs were later found to be slightly different after post-markerting reports revealed that the prevalence of skin rash was found to be lower than observed prior to approval (decreased from 17 to 9%), while that of liver toxicity was slightly higher (increased from 1 to 3%) (Popovic *et al.*, 2010).

The most frequent reported NVP-related side effects in clinical trials include rash, allergic reactions, hepatitis, abnormal liver function tests, nausea, vomiting, diarrhoea, abdominal pain, fatigue, fever, headache and malygia (Boeringer-Ingelheim, 2017). While rash, for example, tends

to be very common among NVP patients (13.6%), post marketing experience has shown that the most serious adverse events are Stevens-Johnson syndrome and/or toxic epidermal necrolysis (SJS/TEN) which do occur in up to 2% of patients (Mittmann *et al.*, 2012).

SJS/TEN are delayed type mucocutaneous immune reactions with the involvement of widespread keratinocyte apoptosis (Roujeau & Stern, 1994). The clinical difference between SJS and TEN depends on the extent of epidermal detachment. SJS, SJS/TEN overlap, and TEN can be classified as the degree of skin detachment involving less than 10%, 10–30%, and greater than 30% of body surface area, respectively (Bastuji-Garin *et al.*, 1993). The manifestations of SJS/TEN range from mild exanthematous skin rashes to a large amount of bullae and extensive mucocutaneous sloughing (Roujeau, 1994). In spite of the low incidence, the mortality rate is 5–10% for SJS, about 30% for SJS/TEN overlap, and nearly 50% for TEN (Dao *et al.*, 2015).

NVP hypersensitivity reactions have been shown to be genetically mediated in recent years. Both immune (HLA-dependent) and non-immune (mostly Cytochrome P450 SNPs) pharmacogenetic associations have been linked to NVP-induced rash, SJS/TEN and hepatotoxicity by several studies. The most significant findings on NVP hypersensitivity pharmacogenetics are summarised in Table 2.1.

It is unclear whether the apparent complexity of NVP-HSR pharmacogenetics is a result of spurious associations or a reflection of population genetic differences as African populations, for example, are known to be genetically very diverse (Campbell & Tishkoff, 2008). Nevertheless, as it stands, no NVP-hypersensitivity pharmacogenetics association has yet translated into clinical use.

Table 2.1: Pharmacogenetic associations between various SNPs/HLA alleles and Nevirapine-induced hypersensitivity reactions.

SNP/Allele	ADR Type	Population	Reference
*HLA-B*35:05*	Rash	Thai	Chantarangsu *et al.* (2009)
*HLA-B*35*	Cutaneous ADR	Asian (mainly Thai)	Yuan *et al.* (2011)
	Hepatotoxicity	Asian	
*HLA-B*58:01*	Hepatotoxicity	South African	Phillips *et al.* (2013)
*HLA-C*04*	Rash	Thai	Likanonsakul *et al.* (2009)
	Cutaneous ADR	Asian, Black, White	Yuan *et al.* (2011)
	Hypersensitivity	Han Chinese	Gao *et al.* (2012)
*HLA-C*04:01*	SJS	Malawian	Carr *et al.* (2013)
	SJS/TEN	Malawian	Carr *et al.* (2017)
*HLA-DRB1*01*[a]	Rash	French	Vitezica *et al.* (2008)
	Hepatotoxicity	Whites	Yuan *et al.* (2011)
*HLA-DRB1*01:01*	Hypersensitivity	Western Australian	Martin *et al.* (2005)
*HLA-DRB1*01:02*	Hepatotoxicity	South African	Phillips *et al.* (2013)
CYP2B6 516TT	Hypersensitivity	Blacks	Yuan *et al.* (2011)
CYP2B6 516G>T	Hepatotoxicity	Asian, Black, White	Yuan *et al.* (2011)
		Mozambican	Ciccacci *et al.* (2013)
CYP2B6 983T>C	SJS/TEN	Mozambican	Ciccacci *et al.* (2013)
		Malawian & Ugandan	Carr *et al.* (2014)
ABCB1 3435C>T	Hepatotoxicity	Mozambican	Ciccacci *et al.* (2010)
TRAF31P2	SJS/TEN	Mozambican	Ciccacci *et al.* (2015)

[a]*HLA-DRB1*01* association by Vitezica *et al.* (2008) involved cutaneous hypersensitivity reactions resulting from both NVP and efavirenz use.

2.3 HUMAN LEUCOCYTE ANTIGEN (HLA)

2.3.1 The Major Histocompatibility Complex

Genes encoding the human leucocyte antigens (HLA) lie in a region on the short arm of Chromosome 6 (6p21.3) referred to as the Major Histocompatibility Complex (MHC) (Erlich *et al.*, 2001). The MHC region is divided into three distinct classes: Class I, II and III, with the HLA genes located in the class I and II regions (Mayor *et al.*, 2015); although, for all practical purposes, the term 'HLA' is often used to refer to the 'human MHC' (Robinson *et al.*, 2016). HLAs are so named because collectively they encode *a*ntigens that predominantly line (reside on) the surfaces of *h*uman *l*eucocytes and play key roles in the adaptive immune defense system (Paunić *et al.*, 2012). This region, approximately 4Mb, is by far the most polymorphic region in the human genome (Juhos *et al.*, 2016) with over 220 genes (Robinson *et al.*, 2014) and 16, 933 HLA and related alleles as of March, 2017 (http://hla.alleles.org/alleles/index.html).

2.3.2 Classification and Structure of HLA

The HLA complex encode two main distinct classes of highly polymorphic cell surface molecules that bind and present surface antigens in the form of peptides to T-lymphocytes (Erlich *et al.*, 2001). Genes from this region are thought to have arisen from multiple gene duplication and crossover events due to their shared extensive sequence homology (Lazaro *et al.*, 2013).

Classical HLA class I molecules, HLA-A, -B, and -C, are present as transmembrane glycoproteins on the surface of all nucleated cells and present mostly endogenously derived antigens (e.g. viral peptides) to CD8+ T cells (Erlich *et al.*, 2001; Robinson *et al.*, 2014). Class I molecules consist of an HLA encoded polymorphic glycoprotein chain (α-chain) associated with a monomorphic β2-

microglobulin (Erlich *et al.*, 2001; Robinson *et al.*, 2014). As of March 2017, there were a total of 12,351 known alleles encoded by this HLA class alone, although not all of these alleles are functional (http://hla.alleles.org/alleles/index.html) (see Figure 2.5).

HLA class II molecules, on the other hand, present mostly exogenously derived antigens (e.g. bacterial peptides) to CD4+ T helper cells. Class II molecules consist of HLA-encoded α and β glycoprotein chains associated as heterodimers on the cell surface of antigen presenting cells such as B cells, macrophages, and dendritic cells (Erlich *et al.*, 2001). Most important Class II loci include HLA-DRA, -DRB1, -DRB3, -DRB4, -DRB5, -DQA1, -DQB1, -DPA1, and -DPB1 (see Figure 2.5) (Lazaro *et al.*, 2013). While most HLA genes are present in a diploid state, individuals vary in the number of HLA-DRB genes carried; from two (DRB1 loci, one on each copy of chromosome 6) to four (additional loci being DRB3, DRB4, or DRB5) (Lazaro *et al.*, 2013). In general, the DRB3 locus is found on haplotypes where DRB1 is *03, *11,*12, *13, or *14; the DRB4 locus is found on haplotypes where DRB1 is *04, *07, and *09; and the DRB5 locus is found on haplotypes where DRB1 is 15 or 16 (*15 and *16 are subtypes of the serological type DR2). DRB1*01, *08, and *10 haplotypes typically have only the DRB1 locus (Erlich *et al.*, 2001).

A relatively less defined MHC Class III region harbours the tumour necrosis factor (TNF) superfamily of cytokines which are involved in systemic inflammation and programmed cell death (specifically apoptosis) (Hajeer & Hutchinson, 2001; Wajant *et al.*, 2003).

Figure 2.5 depicts a simplified MHC map showing some of the most important HLA loci associated with HLA typing in clinical practice.

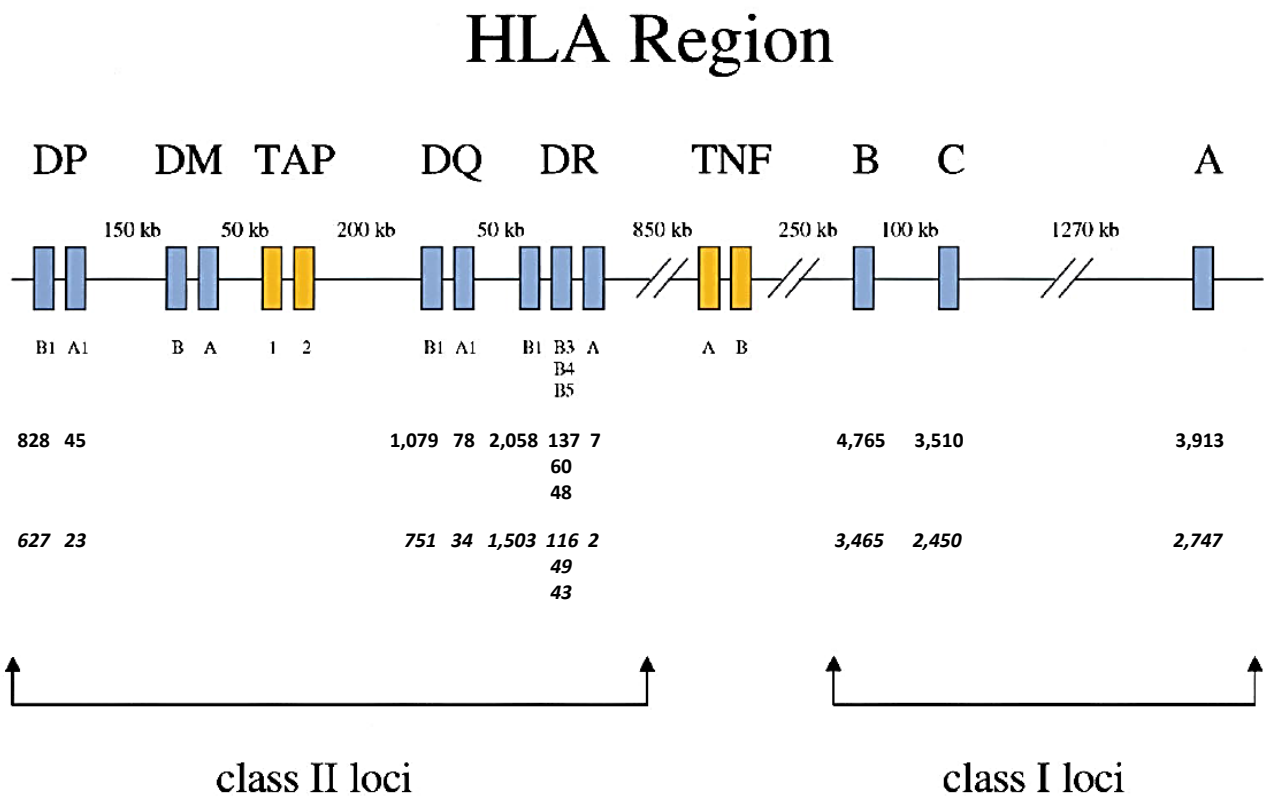


Figure 2.5: Simplified map of the human MHC region.

The HLA class I and class II loci are shown in blue. Kilobase distances between loci are shown above the colored boxes. For class II MHC loci, α (A) and β (B) chain descriptors are shown below the colored boxes. The number of alleles at each MHC locus (as of March, 2017) is shown in bold below the α and β chain descriptors, and the number of proteins encoded at each locus is shown in italics. HLA DM loci are relatively non-polymorphic regions that are typically not involved in clinical HLA typing. TNF belongs to MHC class III, while TAP-1 and -2 encode ATP binding cassette proteins which are non-HLA molecules found in the MHC class II region [Adapted from Erlich *et al.* (2001) with updates from http://hla.alleles.org/nomenclature/stats.html (Robinson *et al.*, 2015)].

2.3.3 HLA Nomenclature System

Early recognition of the highly polymorphic nature of HLA genes made it apparent that there was the need to develop a systematic method of naming and this has been the core function of the WHO

Nomenclature Committee for Factors of the HLA System since 1968. This committee meets regularly to assign official names to newly discovered HLA alleles which are then reported in the Nomenclature for Factors of the HLA System series (Robinson *et al.*, 2014).

By convention, each named allele begins with an 'HLA-' prefix followed by the name of the particular locus (A, B, DRB1 etc), an asterisk separator (*), and a set of digits describing the specific attributes of that allele. Previous nomenclature systems employed a unique four, six or eight digit system to describe each HLA allele (Bodmer *et al.*, 1999). The first two digits denote the type or allele group, which often corresponds to the serological antigen carried by an allotype, while the third and fourth digits are used to describe the subtypes, which refer to alleles that differ in one or more nucleotide substitutions that change the amino acid sequence of the encoded protein (i.e. non-synonymous nucleotide substitutions). The fifth and sixth digits describe nucleotide substitutions in the coding regions that are synonymous, while the final seventh and eighth digits describe changes in nucleotide sequences in the non-coding introns or untranslated regions. Numbers are assigned in the order in which DNA sequences have been determined. For example alleles *HLA-A*2701* and *HLA-A*2702* belong to the same HLA type (*HLA-A*27*) but differ in the primary protein sequence and thus belong to different subtypes, whereas *HLA-B*580102* and *HLA-B*580103* belong to the same HLA type and subtype but each possesses different nucleotide substitutions that do not change their primary amino acid sequences and thus essentially code for the same functional antigen. Alleles *HLA-B*58010302* will differ from *HLA-B*580103* above by mutations or substitutions outside the coding region.

This nomenclature system was ideal until the explosion in HLA allele discoveries in the early 2000s when DNA sequencing became the mainstay in HLA typing. A two-digit stepwise system such as above meant that each 'field' could only hold a maximum of 99 variants, above which

would be problematic — and indeed this was the case when scores of HLA-B and -DRB1 family of alleles were discovered beyond hundreds (Marsh *et al.*, 2005). Initially, a rollover system was adopted to address this problem (Marsh & WHO Nomenclature Committee for Factors of the HLA System, 2003), but it was soon realised that even this system was not capable of accommodating all the possible alleles at a given locus, considering the rate at which new alleles were being sequenced (Marsh *et al.*, 2010).

This called for a revision of the naming system and following the 14th and 15th International HLA and Immunogenetics Workshops in December 2005 in Melbourne, Australia, and in September 2008, in Buzios, Brazil, respectively, came the current Nomenclature for Factors of the HLA System, 2010 (Marsh *et al.*, 2010). This system also makes use of an 'HLA-' prefix, followed by the HLA locus and asterisk separator and a unique number corresponding to up to four sets of digits that describe the particular allele. All alleles receive at least a four-digit name, which corresponds to the first two sets of digits, and longer names are only assigned when necessary (http://hla.alleles.org/nomenclature/naming.html). The major revision under this system, however, was the introduction of colons (:) into the allele names to act as delimiters of the separate fields. In this system, a field is not limited to two digits, essentially removing the limitation of the previous system and the need for a rollover system, allowing the assignment of hundreds of allele variants without restriction; such as *HLA-A*02:101* for example. This system also allowed easy conversion of the old nomenclature to the new system, thus *HLA-B*580102* becomes *HLA-B*58:01:02* and *HLA-A*02010102L* becomes *HLA-A*02:01:01:02L*, for example. Where conversion was not obvious, as in the case of a rolled over allele, the Committee also laid down guiding principles for such name conversations [see Marsh *et al.* (2010) for the full report].

Finally, in addition to the prefix descriptors and unique allele numbers, there are additional optional suffixes that may be added to an allele to indicate its expression status. Alleles that have been shown not to be expressed (i.e. '**N**ull' alleles) have been given the suffix '**N**' while alleles that have been shown to be alternatively expressed have be given the suffix '**L**' (for '**L**ow' cell surface expression, compared to normal levels), '**S**' (for '**S**ecreted' molecule but not present on the cell surface), and '**Q**' (for an allele whose expression is '**Q**uestionable', given that the mutation seen in the allele has been shown to affect normal expression levels in other alleles). Other suffixes '**C**' (for alleles that produce proteins that are present in the '**C**ytoplasm' and not on the cell surface) and '**A**' (for an '**A**berrant' expression where there is some doubt as to whether a protein is actually expressed) exist, but have not yet been assigned to any allele as of March, 2017. (http://hla.alleles.org/nomenclature/naming.html). The current nomenclature system based on the 2010 naming convention for HLA alleles as defined by this committee is briefly illustrated below in Figure 2.6.

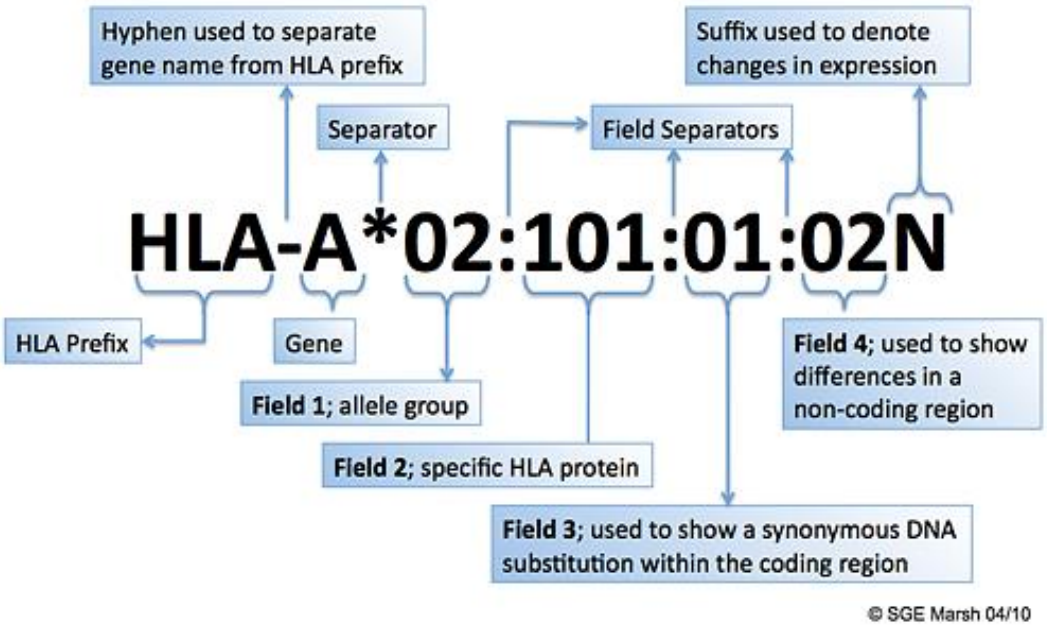


Figure 2.6: Current naming convention for HLA alleles (Marsh *et al.*, 2010).

2.3.4 HLA Typing

HLA typing refers to the process of identifying the particular alleles or groups of alleles a person carries at a particular HLA locus. The importance of knowing the HLA type of an individual has long been recognized as crucial to the success of whole organ (historically in kidney transplant patients) (Dupont & Svejgaard, 1977; Marsh *et al.*, 2000) and hematopoietic stem cell or bone marrow transplantations (Flomenberg *et al.*, 2004). Matching donor to recipient HLA types significantly increases the success rate of allogenic organ transplantations, while reducing the occurrence of graft rejection and/or graft-versus-host disease (Ichiki *et al.*, 2006; Vadakekolathu & Rutella, 2017).

Specific HLA alleles have also been found to be important markers for determining the occurrence or exclusion of some autoimmune disorders (such as multiple sclerosis, type 1 diabetes mellitus, ankylosing spondylitis, psoriasis, etc.) (Lin & Gong, 2017; Marsh *et al.*, 2000; Zambrano-Zaragoza *et al.*, 2013) and recently, some adverse drug reactions (such as abacavir hypersensitivity, carbamazepine-induced SJS/TEN) (Fricke-Galindo *et al.*, 2017). Traditionally, HLA typing is usually restricted to the classical HLA genes, with the strongest determinants being HLA-A, -B, -C and HLA-DRB1 (Fürst *et al.*, 2013; Lee *et al.*, 2007). These loci are therefore the most important for matching donor and recipient.

Allele-level resolution is considered ideal for most histocompatibility testing requirements but high level resolution methods is often restricted by cost and time limitations (Mayor *et al.*, 2015). While historically, low to intermediate level typing resolutions have been accepted as sufficient for matching donor to recipients in solid organ transplantations, with high resolution typing often reserved for hematopoietic stem cell transplantations (Erlich *et al.*, 2001), growing recent evidence suggests that allele-level typing offers a better approach to identifying suitable donors for

sensitized patients (Hahn *et al.*, 2015). Nonetheless, low level resolution typing is still relatively in common use in the field of HLA-drug association studies (Gao *et al.*, 2012; Likanonsakul *et al.*, 2009; Vitezica *et al.*, 2008; Yuan *et al.*, 2011).

Several methods have been employed in the field of HLA typing over the years with technology advancement and the need for high level typing resolutions being the main driving forces. The earliest HLA typing method used was serologic, making use of sera derived from multiparous women or individuals who had received multiple blood transfusions (Erlich *et al.*, 2001). This method of typing was, however, limited to only the range of diversity that could be identified by HLA derived antibodies (referred to as serotypes) and was thus not specific enough to distinguish allelic differences. Serologic typing was hence often augmented with cellular typing methods such as mixed lymphocyte reaction (MLR) (Erlich *et al.*, 2001).

These methods, were soon replaced by DNA typing methods first by restriction fragment length polymorphisms (RFLPs), and then subsequently by Sanger sequence-based DNA typing (SBT) following PCR using locus-specific primers (Erlich, 2012; Fasano *et al.*, 2015; Lazaro *et al.*, 2013). SBT methods have traditionally focused on sequencing only the core exons that code for the antigen recognition site of HLA molecules (exons 2 and 3 for HLA class I genes and exon 2 only for class II genes) since these exons harbour majority of the genetic variation found in the classical HLAs (Erlich *et al.*, 2001; Lazaro *et al.*, 2013). This approach has been used extensively to save time and cost, as opposed to full sequencing of an entire HLA locus, which can be quite expensive particularly by Sanger sequencing (Mayor *et al.*, 2015).

However, due to the high polymorphic and co-dominant nature of HLA loci, SBT methods using locus-specific primers often yield mixed DNA sequences which are not always suitable for discriminating specific HLA alleles and often lead to ambiguous typing results (Ehrenberg *et al.*,

2014; Lazaro *et al.*, 2013; Mayor *et al.*, 2015). Locus-specific PCR SBT methods are thus *usually* augmented with the use of sequence-specific primers (SSPs) (which involves using allele sequence-specific primers for PCR) and/or sequence-specific oligonucleotide probes (SSOPs) (which involves the use of oligonucleotide probes that hybridize specifically to a particular allele after PCR with locus-specific primers) (Fasano *et al.*, 2015; Lazaro *et al.*, 2013) — and this had long been regarded as the 'gold standard' for HLA typing for decades (Osoegawa *et al.*, 2016).

However, in recent times, the explosion in numbers of HLA allele discoveries coupled with the inherent inability of Sanger SBT methods to determine the phase of observed DNA sequence polymorphisms, has called for the need for other DNA typing methods that would be better suited to reducing typing ambiguities (Erlich, 2012, 2015; Juhos *et al.*, 2016; Mayor *et al.*, 2015). Next generation sequencing (NGS) systems provide hope of addressing this problem due to their ability to determine the phase of polymorphisms by generating large numbers of 'clonal' sequence reads allowing each read to be assigned to a single allele, resulting in HLA types with fewer ambiguities (Osoegawa *et al.*, 2016). NGS-based methods also lend themselves to various adaptations for sequencing of whole HLA genes, allowing for accurate allele calling (Mayor *et al.*, 2015) and hence higher resolution typing. NGS is thus fast becoming the preferred DNA technology for HLA typing, but its adoption is severely hampered by high equipment cost and available expertise.

2.4 PHARMACOGENETICS AND THE ROLE OF DNA SEQUENCING

The concept of personalized medicine is not new to the world of medical care. Sir William Osler (1849 – 1919) once made the profound statement, 'the good physician treats the disease; the great physician treats the patient who has the disease' (Centor, 2007) — suggesting individualization of medical care as being superior to generalization. However, for decades, this idea was far from reality as the tools to access the often not so obvious individual differences were not yet available (McCarthy *et al.*, 2013). Little progress was thus made in this field of medicine, until the advent of recombinant DNA technology in the 1970s and most importantly DNA sequencing (Kim, 2002).

Now, with the various DNA sequencing methods available, scientists have direct access to one of the major factors influencing individualization of medical care — genetic variability. Upon the completion of the human genome project in 2003 (Consortium, 2004), the world of biomedical science has since seen an enormous turnout of molecular genetic data. This coupled with a rapid developing computing landscape, has emerged a thriving field of bioinformatics by which useful genetic information are mined from these big biological data (Kane & Brewer, 2007). Today, significant strides have been made in the field of personalised medicine, ranging from risk assessment in healthy individuals to genome-guided treatment in patients with complex diseases (McCarthy *et al.*, 2013).

Pharmacogenetics, the study of the relationship between the genetic make-up of an individual and their response to a particular drug or drugs, has since emerged as a prospective branch of personalised medicine and by far the biggest beneficiary of these recent advancements in medical science. A lot of specific drug responses, from dose-dependent efficacy and/or toxicity responses to idiosyncratic adverse drug reactions have been linked to the presence or absence of specific individual gene mutations/polymorphisms over the last few decades (reviewed in McCarthy *et al.*,

2013). However, this field of medicine, really came to the limelight 10 years ago when the United States Food and Drug Administration (FDA) updated the package insert for warfarin to include dosing guidance based on pharmacogenetics data (Bottorff *et al.*, 2017). It is now standard practice for drug companies to include a 'pharmacogenetic warning' section in the drug product inserts while drug authority governing bodies such as the US FDA and the Clinical Pharmacogenetics Implementation Consortium (CPIC) also actively update consumers with useful guidelines on these as well as other pharmacogenetic warnings (Bottorff *et al.*, 2017; Kusinitz *et al.*, 2017). Although these findings are not always profound as expected, and often struggle to gain clinical significance (Pavlos & Phillips, 2012), there still remains a lot to be discovered in this relatively young and exciting field.

With the continual advancement of DNA sequencing methods, from the early Sanger dideoxy chain termination sequencing methods using gel electrophoresis, to automated capillary electrophoretic chain termination methods, and now to high throughput next generation sequencing, one can be hopeful that DNA sequencing will continue to play a vital role in the science of personalized medicine now and in the distant foreseeable future.

CHAPTER 3

3.0 MATERIALS AND METHODS

3.1 PATIENTS RECRUITMENT AND ETHICS STATEMENT

The study was conducted at the Juaben Municipal Hospital in the remote town of Juaben, in the Ejisu-Juaben District, of the Ashanti Region of Ghana. The HIV patients recruited in the study were accessed from the Antiretroviral Centre of the hospital. The hospital was chosen owing to a preliminary investigation, which revealed a 'frequent' incidence of HAART adverse events of rash and other severe hypersensitivity reactions including SJS/TEN. Prior to the start of the research, ethical approval from Juaben Municipal Hospital and the Committee on Human Research, Publication and Ethics (CHRPE) (Ref: CHRPE/AP/392/15), Kwame Nkrumah University of Science and Technology, School of Medical Sciences were obtained. Written informed consent was sought from all participants prior to their enrolment into the study (shown in appendix 1).

3.2 STUDY DESIGN

This study follows a case-control design involving HIV-1 patients accessing NVP-based combination therapy at the Juaben Municipal Hospital. Patients who developed rash and/or SJS/TEN symptoms within the first three months of NVP-based combination therapy were recruited as cases, while those who had been on NVP-based combination therapy for at least three months and shown no signs of both rash and SJS/TEN constituted the controls. Cases were selected based on the clinicians' diagnosis, as recorded in the hospital's antiretroviral centre database, while controls were selected by simple randomization from the lot of NVP-tolerant patients recorded in the same database.

3.3 STUDY POPULATION

3.3.1 Sample Size

A minimum estimated sample size of 56 participants was calculated following the case-control design using the OpenEpi version 3 online sample size calculator available at http://www.openepi.com/SampleSize/SSCC.htm with the following parameters: Confidence Interval = 95%, Power = 80%, Ratio of controls to cases = 2.0, Hypothetical proportion of controls with exposure = 30%, Hypothetical proportion of cases with exposure = 70% (after Kelsey *et al.*, 1996). A total of 58 participants were enrolled into the study out of which two were excluded, following application of the exclusion criteria. A subpopulation of 30 individuals was selected for HLA typing and association studies comprising 10 patients with NVP-induced rash (cases) and 20 tolerant patients (controls). Participants in the typing population were all female. This was done to minimize possible confounding factors in order to avoid spurious associations since the initial population was heavily gender-biased (with women accounting for all cases) and given that women have been previously reported to have a slightly increased risk for developing NVP-related hypersensitivity reactions (Bersoff-Matcha *et al.*, 2001).

3.3.2 Inclusion Criteria

- All patients who received Nevirapine as part of their HAART treatment drug combination.
- Patients who developed NVP-induced rash and/or SJS/TEN during the course of treatment.

3.3.3 Exclusion Criteria

- Unavailability of blood samples due to either patient's subsequent death since treatment initiation, transfer or travel to a different location, or unwillingness to participate in the study.

- Diagnosis with HIV type 2.
- Coinfection with hepatitis B virus.
- Use of immunomodulatory medications within the first 8 weeks of Nevirapine therapy.
- Initiation of abacavir or trimethoprim/sulfamethoxazole within 2 weeks prior to or within 8 weeks after initiating Nevirapine.

3.4 DATA COLLECTION AND PATIENT INFORMATION MANAGEMENT

The principal means of data collection for the selection of both cases and controls was the Juaben Municipal Hospital ART Centre's electronic database which keeps records of all patients who have been treated at the Centre since its inception in 2008 to date. However, this was aided by a questionnaire-type data collection sheet (shown in appendix 2) administered by the researcher to access other information relevant to the research that were not captured in the Centre's database. These included for example the ethnicity of patients, height and weight (for BMI calculation), whether patient was taking other drugs prior to and during the first weeks of NVP treatment, etc. The same data collection sheet was used to record other relevant data obtained from the patient's folder but not available from the Centre's database.

Each participant was given a unique *anonymous* identification code only traceable to the patient by the researcher and not available either to the general public or for any publication purposes. All patient-sensitive data from the study are being kept secure until the life span of the research after which they shall be destroyed.

No names were collected nor used in this study and no minors were recruited.

3.5 BLOOD SAMPLE COLLECTION AND PREPARATION

Approximately 5ml of venous blood were collected from each participant into BD Vacutainer® blood collection tubes with K_3 EDTA anticoagulant and quickly transferred onto ice to prevent clotting and potential DNA damage. Samples were transferred to the Cancer and Infection Research Facility (CIRF), School of Medical Sciences, KNUST and kept at 4°C for laboratory analysis.

3.6 LABORATORY ANALYSIS

3.6.1 DNA Isolation and Purification

Deoxyribonucleic acid (DNA) was extracted from whole blood samples at the Cancer and Infection Research Facility using a non-enzymatic salting out DNA isolation protocol modified after Suguna *et al.* (2014). Briefly, 900 µl of low salt buffer (TKM 1) and 50 µl of 1x Triton-X were added to 300 µl (more added on modification) of well mixed whole blood samples in an autoclaved 1.5 ml eppendorf tube. This was incubated at 37°C for 5 mins to lyse red blood cells (RBCs) and then centrifuged at 8000 rpm for 3 mins to pellet down unlysed cells while discarding lysed RBCs in the supernatant. This process was repeated 3 to 4 times with decreasing amount of Triton-X, to ensure complete lysis of all RBCs and to obtain a clear whitish pellet of white blood cells (WBCs).

White blood cell lysis was performed by adding 300 µl of high salt buffer (TKM 2) and 40 µl of 10% SDS to the resulting pellet from the RBC lysis, mixed thoroughly and incubated at 37 °C for 5 mins. 100 µl of 6 M NaCl was then added and vortexed to precipitate proteins. The resulting

mixture was then centrifuged at 8000 rpm for 5 mins and the supernatant transferred into a new 1.5 ml eppendorf tube containing 300 µl of absolute isopropanol.

DNA was precipitated by inverting the eppendorf tube slowly. The eppendorfs were then centrifuged at 8000 rpm for 10 mins to pellet down DNA. Discarding the supernatant, the DNA pellet was washed slowly with 70% ethanol to remove excess salts. The resulting mixture was finally centrifuged at 8000 rpm for 5 mins to re-pellet down the DNA, supernatant was discarded and the DNA air-dried. The DNA was then resuspended in 50 µl of sterile double distilled water.

Quality of isolated DNA was verified by the NanoDrop® ND-1000 Spectrophotometer.

3.6.2 Polymerase Chain Reaction (PCR)

3.6.2.1 Primer Design

PCR primers for target gene sequences (i.e. HLA-A and HLA-B loci) were designed using the NCBI Primer-BLAST online tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye *et al.*, 2012) from the Genome Reference Consortium Human Build 38 patch release 2 (GRCh38.p2). Exon-specific primers were designed for exons 1, 2, 3, 4, 5 & 6 combined, and 8 respectively, for both HLA-A and -B loci. This was done to eliminate the need for amplification of the often long intronic regions of HLA-class I molecules (Lazaro *et al.*, 2013) whose genetic variation have a lower likelihood of having a significant impact on gene function due to mRNA splicing.

In the present work, PCR (and subsequent DNA sequencing) was performed for only exons 2 and 3 regions of both HLA-A and -B loci. This is because these exons encode virtually all the genetic variation found in HLA class I antigens and it is standard practice in Sanger sequence-based HLA typing to sequence only these exons as opposed to an entire gene locus to reduce the cost, time and labour required for HLA typing (Erlich, 2012; Lazaro *et al.*, 2013).

All primers were ordered from and sequenced by Integrated DNA Technologies, Coralville, Iowa, United States. Table 3.1 shows the primer sequences used for PCR in the current work (i.e. HLA-A and -B exons 2 and 3) and their respective properties.

Table 3.1: Primer Sequences used for PCR.

Locus	Direction	Sequence (5′>3′)	Length (bp)	GC (%)	*Tm* (°C)
		HLA-A			
Exon 2	Forward	AGG GTC GGG CAG GTC TC	17	70.6	59.7
	Reverse	GGG GAT GAG GGG TCG TGA	18	66.7	59.2
Exon 3	Forward	CTG GGG GGA CTG GGC TG	17	76.5	61.1
	Reverse	TAG GCG ATC AGG GAG GCG	18	66.7	59.4
		HLA-B			
Exon 2	Forward	GGG AGG AGC GAG GGG AC	17	76.5	60.5
	Reverse	TAC GTG GGG GAT GGG GAG	18	66.7	59.0
Exon 3	Forward	AAA ATC CCC GCG GGT TGG	18	61.1	58.9
	Reverse	TCC TCC TCT TCT CGT GGG AG	20	60.0	57.8

GC, GC-content; *Tm*, melting temperature

3.6.2.2 PCR Conditions

All PCR reactions were performed using the Thermo Scientific DreamTaq Green PCR Master Mix (2X). A total reaction volume of 20 µl (comprising 10 µl DreamTaq Green PCR Master Mix, 1 µl DNA template (see Table 4.2 for their respective concentrations), 1 µl each of primer pairs at 10 µM, and 7 µl of sterile double-distilled water) was run for all selected samples, using the following cycling conditions: an initial denaturation at 95°C for 10 mins; 30 cycles of 95°C for 1 min, (Tm^{a*}-

[*] Tm^{a} = Adjusted melting temperature of primer with final $MgCl_2$ and dNTP concentrations.

5°C) for 1 min, and 68°C for 2 min; a final extension of 68°C for 10 min; and a holding temperature of 4°C until end of run.

Resulting PCR products were stored at 4°C for downstream processing.

3.6.3 Agarose Gel Electrophoresis and Visualization

PCR products were run on a 1% (w/v) agarose gel (Fisher BioReagents® Genetic Analysis Grade, broad separation range for DNA/RNA) dissolved in 1x Trisacetate EDTA (TAE) stained with Ethidium Bromide. 5 µl each of PCR product was run for a maximum of 1 hour alongside either a 1 kb or 100 bp molecular weight DNA ladder.

Bands were visualized using a FOTODYNE® Incorporated FOTO/UV 15 transilluminator connected to a regular computer.

Un-run PCR products were stored at 4°C for DNA sequencing.

3.6.4 DNA Sequencing

DNA sequencing was done by the automated Sanger sequencing method (Sanger *et al.*, 1977). Approximately 15 µl each of PCR product was supplied alongside 50 µl of primer (forward primer only) for each set of reaction (i.e. for each exon PCR) to the commercial sequencing company of choice, Inqaba Biotech™, Pretoria, South Africa.

Product purification and sequencing reactions were performed, using the following standard sequencing parameters: ABI 3500XL Genetic Analyzer, POP7™, BigDye® 3.1, according to the sequencing company.

3.7 BIOINFORMATICS AND DNA DATA MINING

DNA sequencing results were delivered as electropherograms in the ABI (.ab1) format by the sequencing company indicated. Using Chromas software version 2.6.4, Technelysium Pty Ltd, South Brisbane, Queensland, Australia, the DNA sequences from the electropherograms (or chromotagrams) were analysed and converted into FASTA (.fasta) formats, which were subsequently used for HLA typing and allele variant calling.

Gapped BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1997) analyses were performed using the concatenated DNA sequences of exons 2 and 3 respectively for each locus. All BLAST analyses were performed in the specialist IPD-IMGT/HLA database (Release 3.28.0, 2017-04-13) which curates all known HLA alleles named by the WHO Nomenclature Committee for Factors of the HLA System to date (Robinson *et al.*, 2015).

BLAST analyses were performed using NCBI nucleotide BLAST version 2.2.29+ (BLASTN 2.2.29+) hosted at http://www.ebi.ac.uk/ipd/imgt/hla/index.html. BLAST parameters were as follows:

- Databank – IMGT/HLA
- Program – blastn
- Task – blastn
- Match/mismatch scores – 1,-3
- Gap Open Penalty – 5
- Gap Extend Penalty – 2
- Expectation Threshold – 10 (default)
- Filter – Yes
- Dropoff – 0 (default)
- Scores – 50 (default)
- Alignments – 50 (default)
- Sequence range (START-END)
- Gap Align – True
- Alignment Views – Flat query-anchored identities

BLAST hits producing the most significant alignments on the bases of highest bit scores were used for genotyping.

3.8 STATISTICAL ANALYSIS

Pearson Chi-square was used to assess the study population for any association between socio-demographic or clinical characteristics and NVP-induced rash. Fisher's exact test was used to analyse the association between alleles observed and NVP-induced rash.

Frequency distribution analyses were performed to assess typing resolution by comparing the proportions of specific versus ambiguous alleles/genotypes typed.

All statistical analyses were performed using IBM SPSS Statistics for Windows, Version 20.0 (Armonk, NY: IBM Corp).

CHAPTER 4

4.0 RESULTS

4.1 BASELINE CHARACTERISTICS OF STUDY POPULATION

The total number of participants enrolled in the study was 56, comprising ten cases (NVP-related rash) and 46 controls (tolerant group) — representing a prevalence of 17.9% NVP-related rash cases among patients accessing NVP-based combination therapy. The selected population varied widely across the socio-demographic and clinical baseline characteristics measured (see appendix 3), but was notably biased towards gender, with females being the majority and accounting for all NVP-related rash cases. Accordingly, this was factored into the selection of the subpopulation of 30 samples used for DNA sequencing and HLA typing by selecting an all-female typing population.

Pearson Chi-square analysis revealed no significant association between most of the socio-demographic/clinical baseline characteristics measured and NVP-induced rash (at *p*-value $\alpha = 0.05$); except for ethnic group (*p*-value = 0.003), suggesting that the ethnic group of a patient significantly influences their likelihood of developing NVP-rash.

Table 4.1 shows the clinical and socio-demographic baseline characteristics of the main population and their association to NVP-induced rash.

Table 4.1: Baseline characteristics of study participants and their association to NVP-rash.

Characteristic	Cases	Controls	All	*p* (cases *vs.* controls)
Age[1], mode (%)	>50 (60.0)	30-40 (43.5)	30-40 (41.1)	0.759
Sex, *n* (%)				0.104
Male	0 (0.0)	10 (21.7)	10 (17.9)	
Female	10 (100.0)	36 (78.3)	46 (82.1)	
Pregnant women[2], *n* (%)	4 (40.0)	16 (44.4)	20 (35.7)	0.802
Ethnic group, mode (%)	Asante (40.0)	Asante (78.3)	Asante (71.4)	**0.003**
CD4 cell count[3], (cells/µl)	361 (177-375)	313 (169-389)	319 (171-387.8)	0.227
Women >250cells/µl, *n* (%)	3 (60.0)	16 (55.2)	19 (55.9)	0.329
Men >400 cells/µl, *n* (%)	N/A	None	None	N/A
Body mass index[3]	24.0 (19.9-25.8)	23.6 (21.8-25.7)	23.7 (20.3-23.8)	0.241

p-value significant at <0.05; N/A, Not Applicable; Significant associations are reported in bold.
[1]Age groups recorded range from between 20-30 years and above 50 years.
[2]Women pregnant at treatment initiation.
[3]Values for CD4 cell count and body mass index are reported as median and interquartile range in parenthesis.

4.2 DNA EXTRACTS

The DNA isolation protocol (Suguna *et al.*, 2014) produced appreciably satisfactory DNA purity and, upon slight modification, high DNA yields as well.

With a minimum of 1.81 and maximum of 1.97, all DNA purity results obtained were well within the 1.8-2.0 range for absorbance ratio 260/280 required to infer good quality DNA by UV spectrophotometry.

DNA yield was also high enough for PCR standards with an average of 375.17 ng/µl, ranging from 40.92 ng/µl to as high as 2,270.9 ng/µl.

Table 4.2 shows the yield and purity values from the samples selected for PCR using the NanoDrop™ ND-1000 Spectrophotometer.

Table 4.2: DNA yield and purity check by NanoDrop™.

Sample ID	Yield (ng/µl)	A260	A280	260/280	260/230
1	98.96	1.979	1.022	1.94	1.86
2	53.98	1.080	0.547	1.97	1.95
4	163.68	3.274	1.667	1.96	1.88
6	498.07	9.961	5.243	1.90	2.01
11	111.78	2.236	1.183	1.89	2.13
12	519.09	10.382	5.651	1.84	2.20
13	99.79	1.996	1.076	1.86	1.99
14	508.15	10.163	5.484	1.85	2.09
17	1335.91	26.718	14.587	1.83	1.84
19	83.80	1.676	0.888	1.89	2.13
20	91.62	1.832	0.979	1.87	2.29
21	220.75	4.415	2.379	1.86	1.92
22	126.81	2.536	1.335	1.90	2.20
25	2720.90	54.418	29.906	1.82	2.00
26	40.92	0.818	0.443	1.85	1.18
27	266.94	5.339	2.825	1.89	2.09
28	119.80	2.396	1.218	1.97	2.01
30	70.61	1.412	0.745	1.89	1.85
31	257.47	5.149	2.780	1.85	1.25
34	820.03	16.401	8.507	1.93	1.95
37	741.68	14.834	7.582	1.96	1.85
38	128.05	2.561	1.379	1.86	1.27
39	239.47	4.789	2.391	1.92	1.92
41	103.46	2.069	1.140	1.81	1.39
43	228.96	4.579	2.510	1.82	1.25
44	428.28	8.566	4.582	1.87	1.70
52	172.49	3.450	1.869	1.85	1.64
53	274.34	5.487	2.949	1.86	1.78
57	141.39	2.828	1.494	1.89	1.70
58	587.90	11.758	6.313	1.86	2.13

A260, Absorbance at 260nm; A280, Absorbance at 280nm; 260/280, Ratio of A260 to A280; 260/230, Ratio of A260 to A230.

4.3 DNA AMPLICONS

Products from PCR were visualized by running a 1% (w/v) agarose gel stained with ethidium bromide at 100 V for 1 hour for all products.

Selected gel profiles are presented in Figures 4.1 to 4.4. Reactions that failed at first attempt were repeated at least once to confirm results.

Legend: Starting from the left,

M_1 = 1kb DNA ladder (Thermo Scientific GeneRuler #SMO311)

M_2 = 100 kb DNA ladder (Thermo Scientific GeneRuler #SMO241)

Numbers 13, 14, 17, 22, 30, 31, 38, 52, 57, 58 = Case samples

C1 = negative control 1 (contains primers but no template DNA)

C2 = negative control 2 (contains neither primers nor template DNA)

Numbers 1, 2, 4. 6, 11, 12, 19, 20, 21, 26, 27, 28, 34, 36, 37, 39, 41, 43, 44, 53 = Control samples.

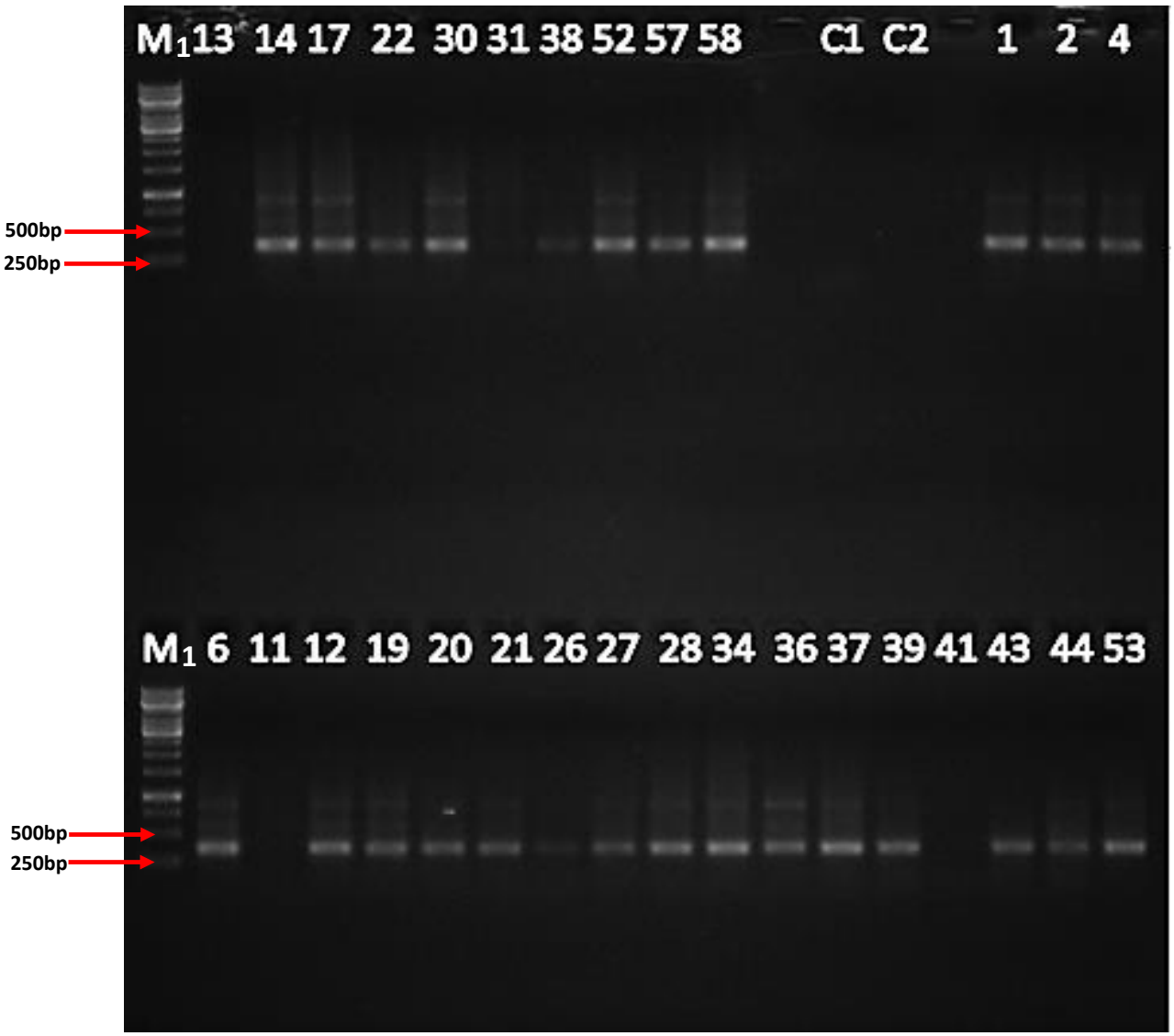


Figure 4.1: Gel electropherogram of amplified DNA from HLA-A exon 2 locus. Expected band size is 351 bp.

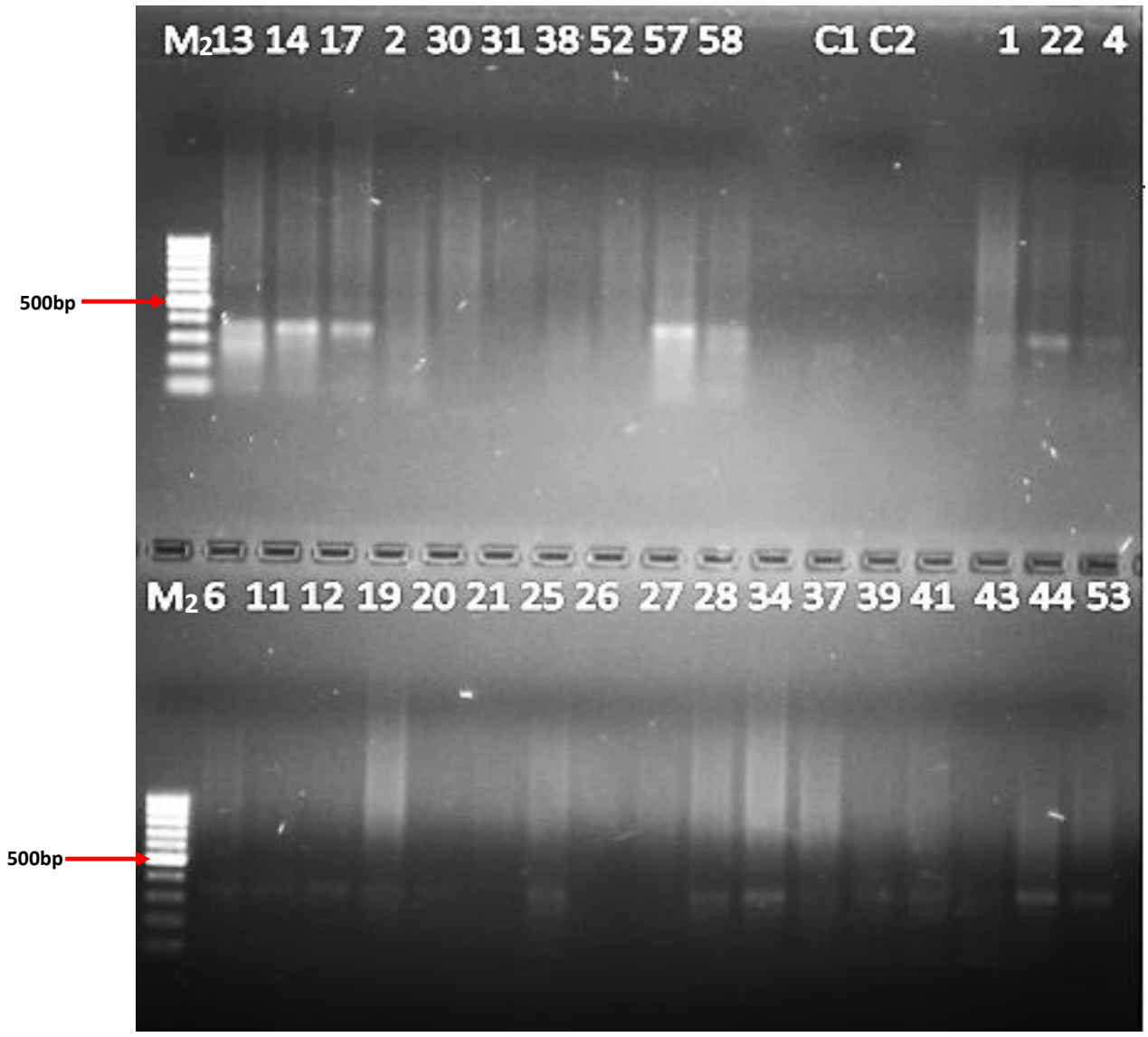


Figure 4.2: Gel electropherogram of amplified DNA from HLA-A exon 3 locus. Expected band size is 348 bp.

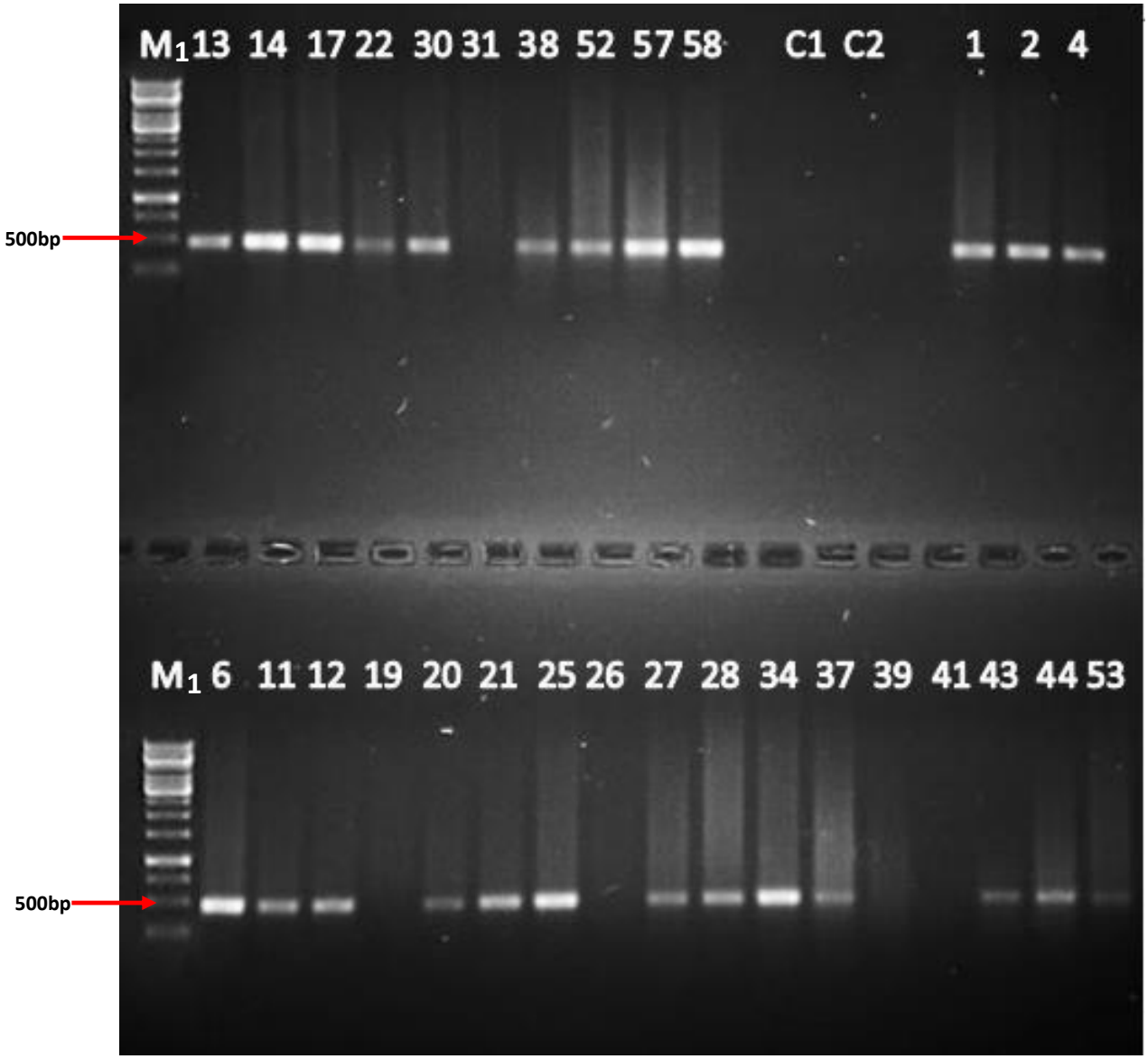


Figure 4.3: Gel electropherogram of amplified DNA from HLA-B exon 2 locus. Expected band size is 415 bp.

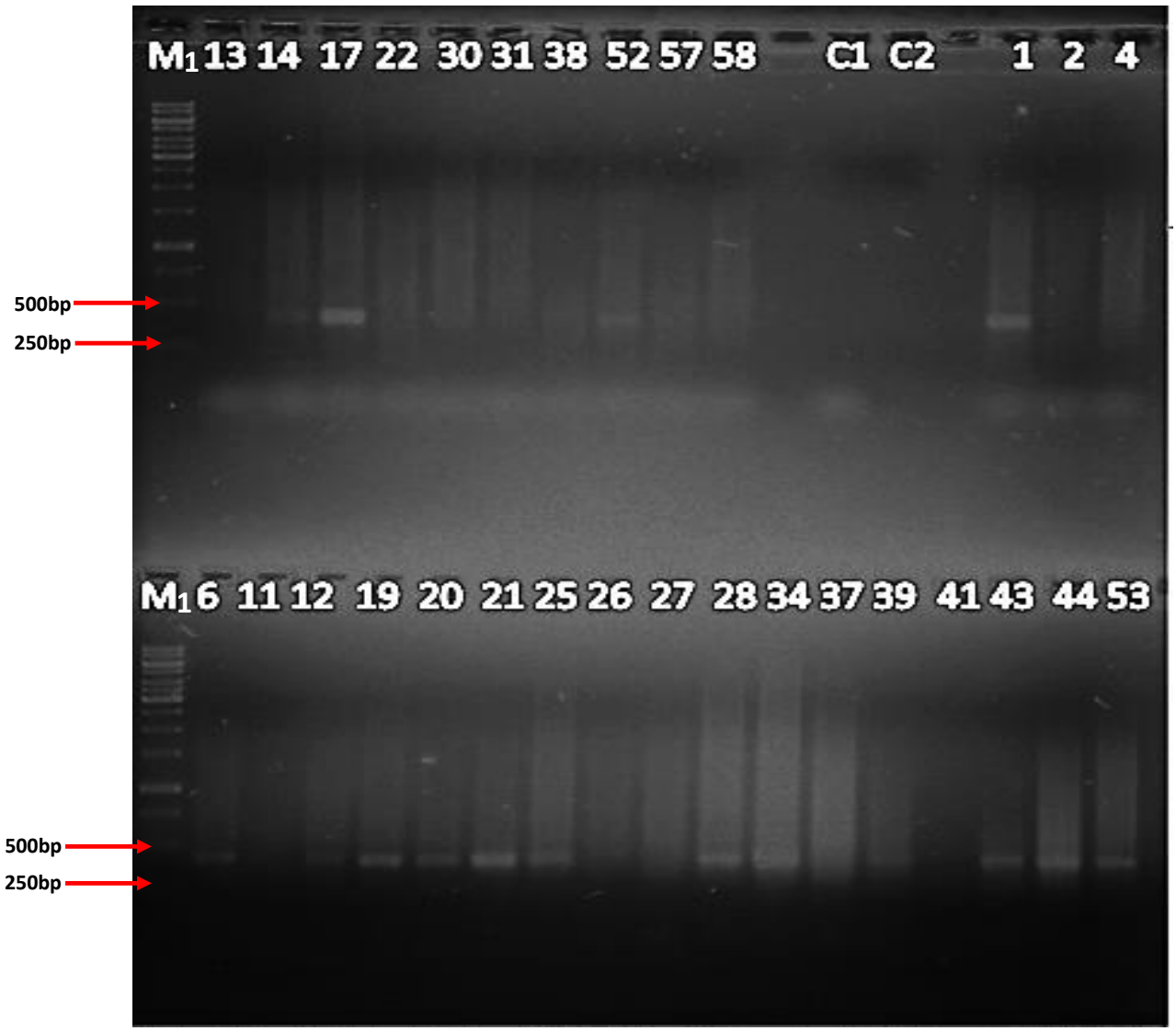


Figure 4.4: Gel electropherogram of amplified DNA from HLA-B exon 3 locus. Expected band size is 420 bp.

4.4 DNA SEQUENCING ANALYSIS

4.4.1 Chromatogram Data

Sequencing data resulting from the Sanger DNA sequencing were of varying degrees of quality — ranging from very high to poor quality. While some chromatograms showed high quality base scores with evenly spaced peaks and minimal background noise, others showed mixed sequences of poorly spaced peaks with very low quality base scores and very prominent and interfering background noise. Yet others showed no sequencing data at all. Average signal intensity ranged from as high as between 500 to 1000 for very high quality data to as low as between 150 to 300 for poor sequencing data.

As expected with automated Sanger sequencing, good quality sequences usually began at roughly 50-60 bp upstream at the 5' end (earlier or later depending on the quality of the particular sequencing data), while the best quality sequence calls were found midway of the chromatograms, with the 3' ends always ending with an 'A' flair.

Figures 4.5 to 4.8 show printouts of selected chromatograms representing the various qualities of sequencing data obtained. All chromatogram printouts were generated using Chromas software version 2.6.4, Technelysium Pty Ltd, South Brisbane, Queensland, Australia.

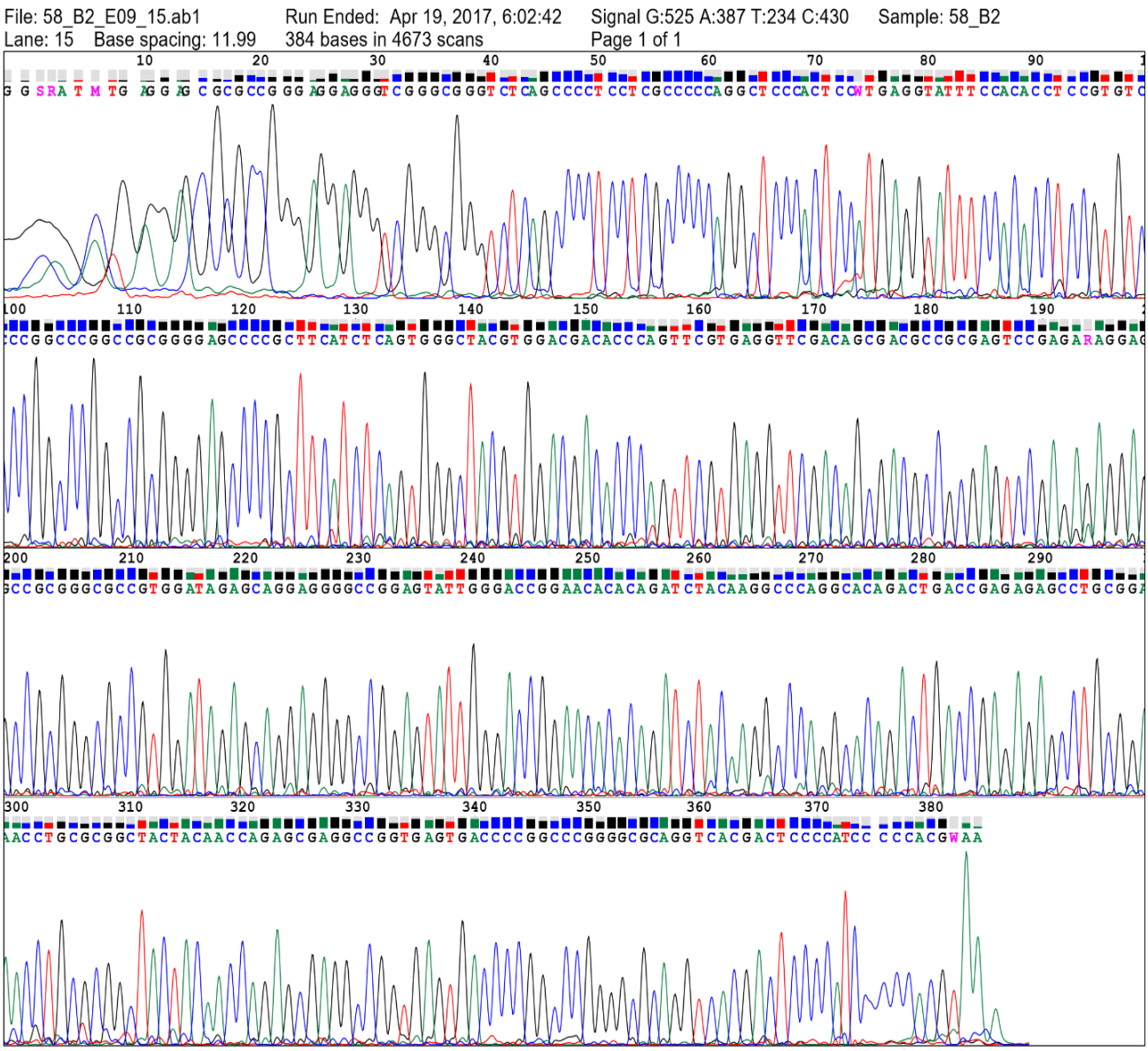


Figure 4.5: Chromatogram printout of HLA-B exon 2 from sample 58.

An example of good sequencing data showing evenly spaced peaks with minimal background noise. Basecall quality scores are high and there are no apparent peaks under peaks visible. However, two potentially miscalled bases can be gleaned at positions 74 and 194.

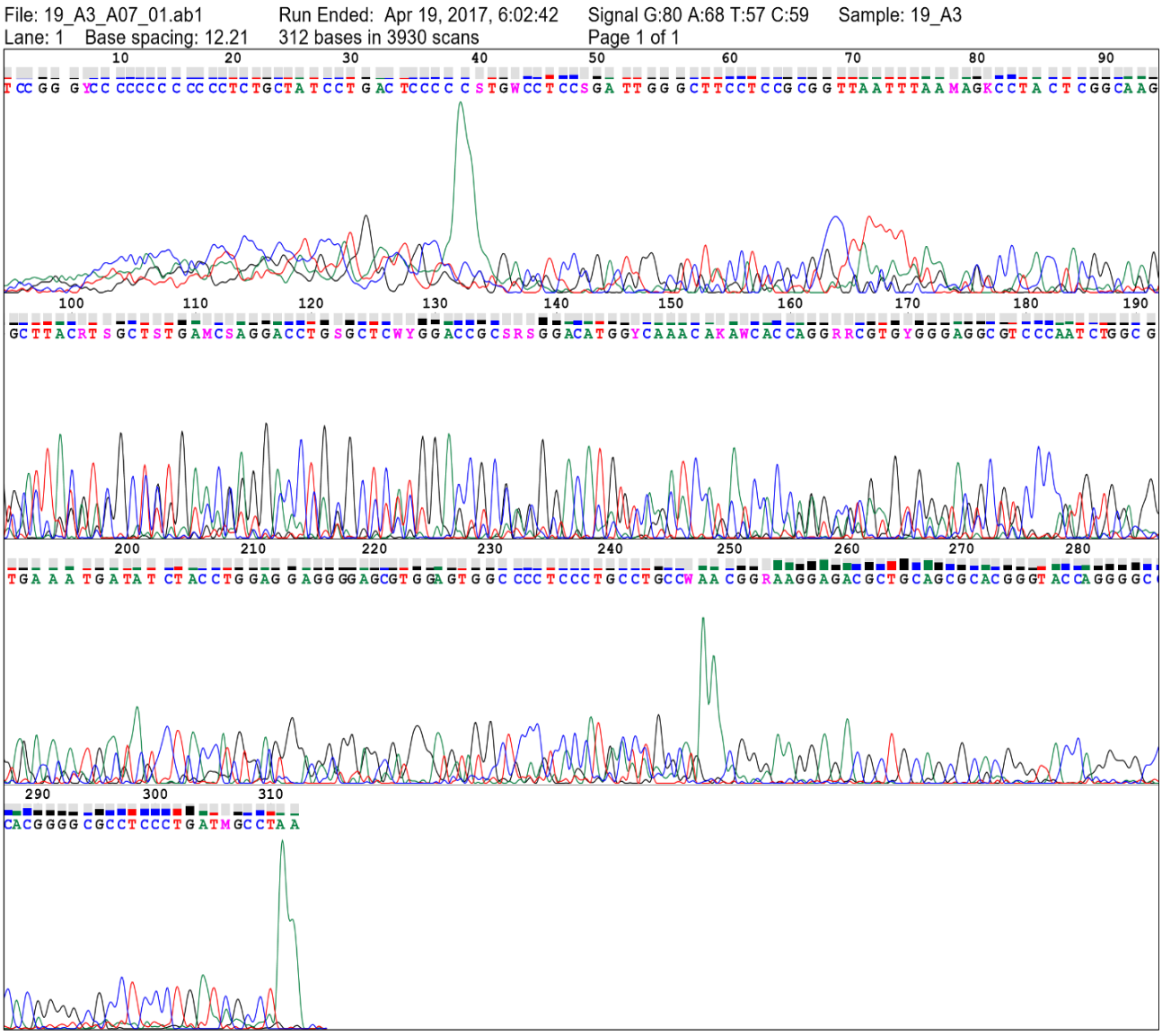
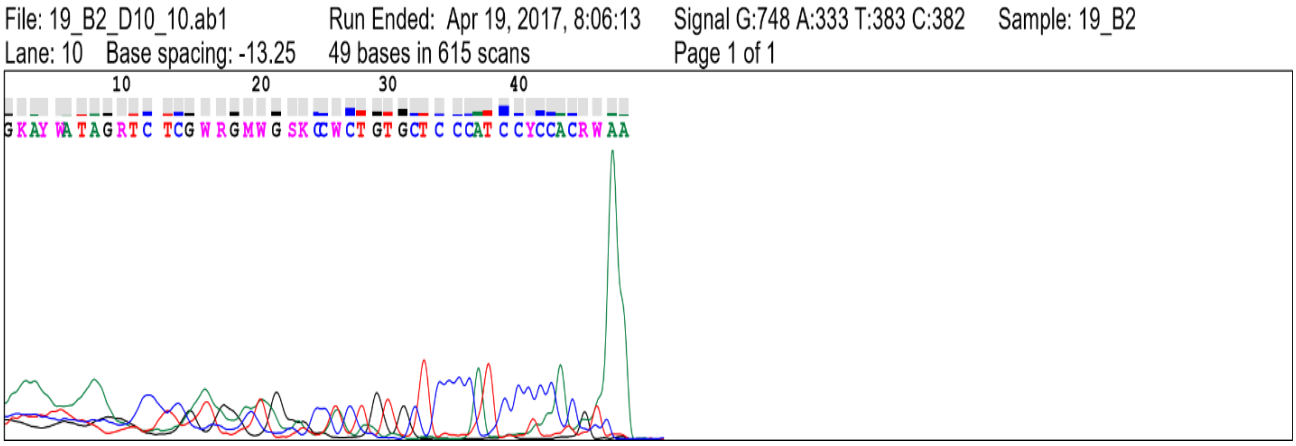


Figure 4.6: Chromatogram printout of HLA-A exon 3 from sample 19.

Mixed sequence data. Signals intensities are low and basecall quality scores poor. Several potential peaks under peaks can be seen from this chromatogram. Dye blob artifacts can be seen in the regions of position 40 and 70. Sequences from this data are not likely to produce 'true' alignments from BLAST analysis.

A



B

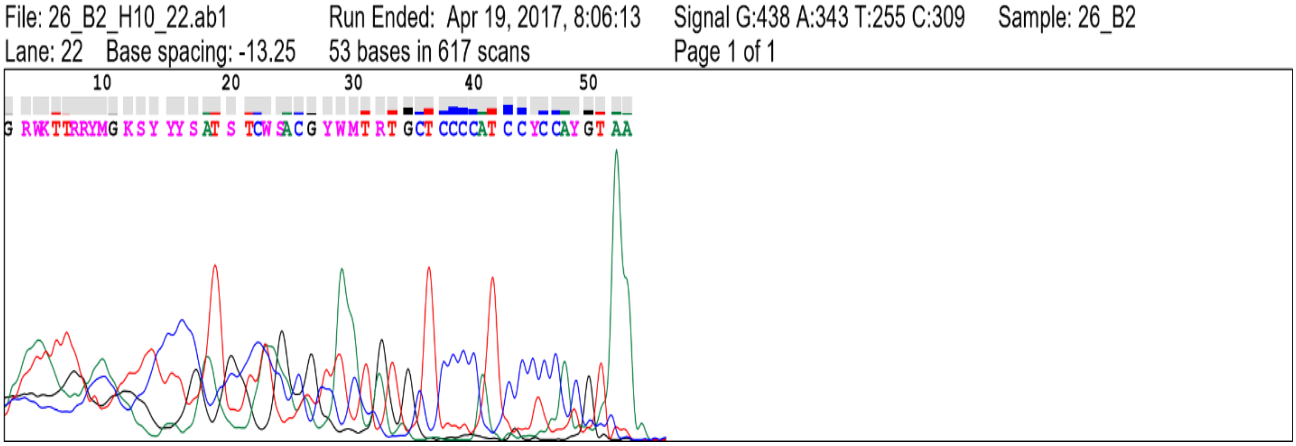


Figure 4.7: Chromatogram printouts of HLA-B exon 2 from samples 19 (**A**) and 26 (**B**).

No sequencing data. This could potentially be due to a lack of primer binding during cycle sequencing, not enough or no DNA/primer in sequencing sample or the presence of an inhibitory contaminant during sequencing. Sequences from such chromatograms are not usable.

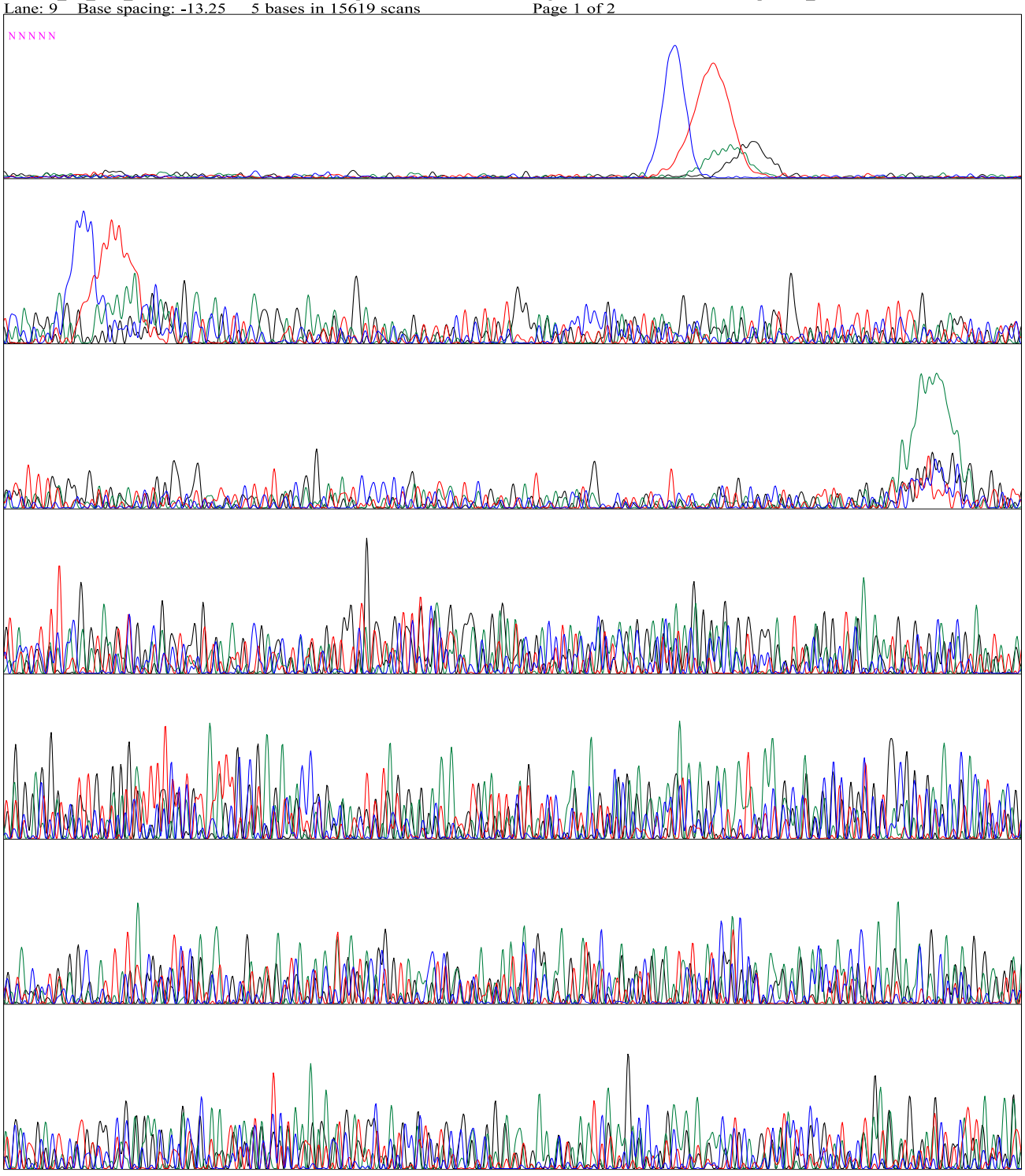


Figure 4.8: Chromatogram printout of HLA-B exon 3 from sample 26.

Failed sequencing reaction. A possible contamination in PCR template product or cycle sequencing product could be inhibiting the sequencing reaction.

4.4.2 DNA Sequences

It is important and hence customary to examine chromatogram data resulting from automated Sanger sequencing prior to exporting the DNA sequences contained in them to ensure quality. This is because no matter how sensitive an automated sequencer is designed to be, it is not always 100% accurate in base calling often due to the nature of samples being sequenced. This problem is further compounded when PCR products that contain mixed products (either more than one amplicon, or heterozygote products) are sequenced. The following steps were thus taken to ensure quality of the sequencing data used for BLAST analysis.

All chromatogram data were manually checked for miscalled bases and edited where found. Very few miscalled single peak bases were found and where there were double peaks, the base caller accurately assigned the right redundant IUPAC codes as follows: K= G or T, M= A or C, R= A or G, S= C or G, W= A or T and Y= C or T; where A, C, G and T represent Adenine, Cytosine, Guanine and Thymidine nucleotides respectively. Remarkably, there were no unassigned basecalls ('N') made in any of the chromatogram data.

The ends of all chromatogram data were subsequently trimmed to get rid of regions of poor quality. Given that good quality sequences usually began at about 50 bp upstream of the chromatograms, all chromatogram data were trimmed of 50 bp at the 5' end and the sequence limited to a specified length of 206 bp for HLA-A exons, and 330 bp for HLA-B exons. This was done to ensure both quality and uniformity of sequencing data.

Appendix 4 and 5 show the resulting DNA sequences after editing and trimming using Chromas software.

4.5 HLA TYPING ANALYSIS

4.5.1 BLAST Analysis and Genotype Assignment

All BLAST alignments produced scores with significant Expect values (E-values) at the threshold of 1E-10. The least significant E-value recorded was 2.30E-77 while most averaged around 1E-150. Identity scores recorded were quite high ranging from a least of 78.2% to as high 100%.

Surprisingly, some HLA-B samples produced hits in HLA-C regions. While this was unexpected, it could be attributable to either very close sequence similarities between the HLA-B alleles under query and that of the HLA-C alleles observed, or simply a case of mixed sequences from unspecific PCR amplification/cycle sequencing or other sequencing artifacts at those loci.

Appendix 6 and 7 show the top 5 BLAST hits (selected from the default top 50 hits) for each sample sequenced. These are BLAST hits that produced significant alignments and were selected on the bases of highest bit scores. Not all exons were successfully sequenced and where only one exon sequence was available, this is shown in parenthesis next to the sample ID in both tables. Two HLA-B samples (i.e. 26 and 31) had no sequencing data for both exons and hence no BLAST results. These are represented by 'N/A' for 'Not Available' in Appendix 7.

Given that HLA genes are co-dominantly expressed with heterozygous genotypes particularly favoured in a population for a higher repertoire of antigen presentation (Mayor *et al.*, 2015; Paunić *et al.*, 2012), a BLAST analysis for HLA genotype assignment, will produce two perfectly best matched alleles (or allele groups) for a heterozygous genotype and only one for a homozygous genotype, at a given locus (Lazaro *et al.*, 2013). Organising the BLAST results obtained on two levels of resolution (i.e. allele-level ('two field') resolution and allele group level ('one field') resolution) yields the genotypes presented in Table 4.3 below.

Table 4.3: Typing results from BLAST analysis in IPD-IMGT/HLA database.

	Allele-level genotype		Allele group genotype	
	HLA-A	**HLA-B**	**HLA-A**	**HLA-B**
1	A/G	A/G	*A*31; A*24*	B*57
2	A*01:19; A*01:173	*B*27:05*[1]	A*01	*B*27*
4	A*01:19; A*01:173	A/G	A*01	*B*53; B*51*
6	A/G	A/G	A*69; A*68	A/G
11	*A*03:222*	A/G	*A*03*	B*58
12	A/G	A/G	*A*02*	A/G
13	A/G	A/G	A*30	A/G
14	A/G	B*51:158[1]	A/G	B*51
17	A/G	HLA-C*	A/G	HLA-C*
19	A/G	HLA-C*	A*03; A*02	HLA-C*
20	A/G	A/G	A/G	A/G
21	A/G	A/G	A*02	B*53; B*51
22	A/G	A/G	*A*31; A*24*	*B*49; B*44*
25	A/G	A/G	A*31; A*24	B*45; B*44
26	A/G	N/A	A*03; A*02	N/A
27	A/G	A/G	*A*69; A*68*	*B*53; B*51*
28	A/G	A/G	A*30	*B*53; B*51*
30	A/G	*B*15:01*[1]	*A*31; A*24*	*B*15*
31	A/G	N/A	*A*02*	N/A
34	A/G	A/G	A*31; A*24	B*15
37	A/G	A/G	A*31; A*24	*B*53; B*51*
38	A/G	*B*07:273*	*A*69; A*68*	*B*07*
39	A/G	HLA-C*	*A*30*	HLA-C*
41	A/G	HLA-C*	A/G	HLA-C*
43	A/G	B*07:133	*A*30*	B*07
44	A/G	B*57:85	A/G	B*57
52	A/G	HLA-C*	*A*02*	HLA-C*
53	A/G	B*07:133	A*30	B*07
57	A/G	A/G	A*31; A*30	A/G
58	A/G	B*42:02[1]	A*30	B*42

A/G, ambiguous genotype; HLA-C*, alleles from HLA-C locus; N/A, no available sequencing data. Homozygous genotypes are presented by a single allele while heterozygous genotypes are presented as two alleles separated by a semicolon**.**

Alleles in italics were typed with only one exon.

[1]Alleles typed to more than two fields, refer to Appendix 7 for details.

4.5.2 Typing Resolution and Genotypic Frequencies

Frequency distribution analysis of successful versus ambiguous genotypes produced an expected pattern of genotyping results. While, the current approach, failed to produce enough high resolution (allele-level) genotypes with only 28.25% overall successfully typed samples at this resolution, it performed better at the allele group level resulting in 73.8% of successfully typed samples for both loci.

Genotypic frequencies from the typing results are presented in Figures 4.9 through 4.12.

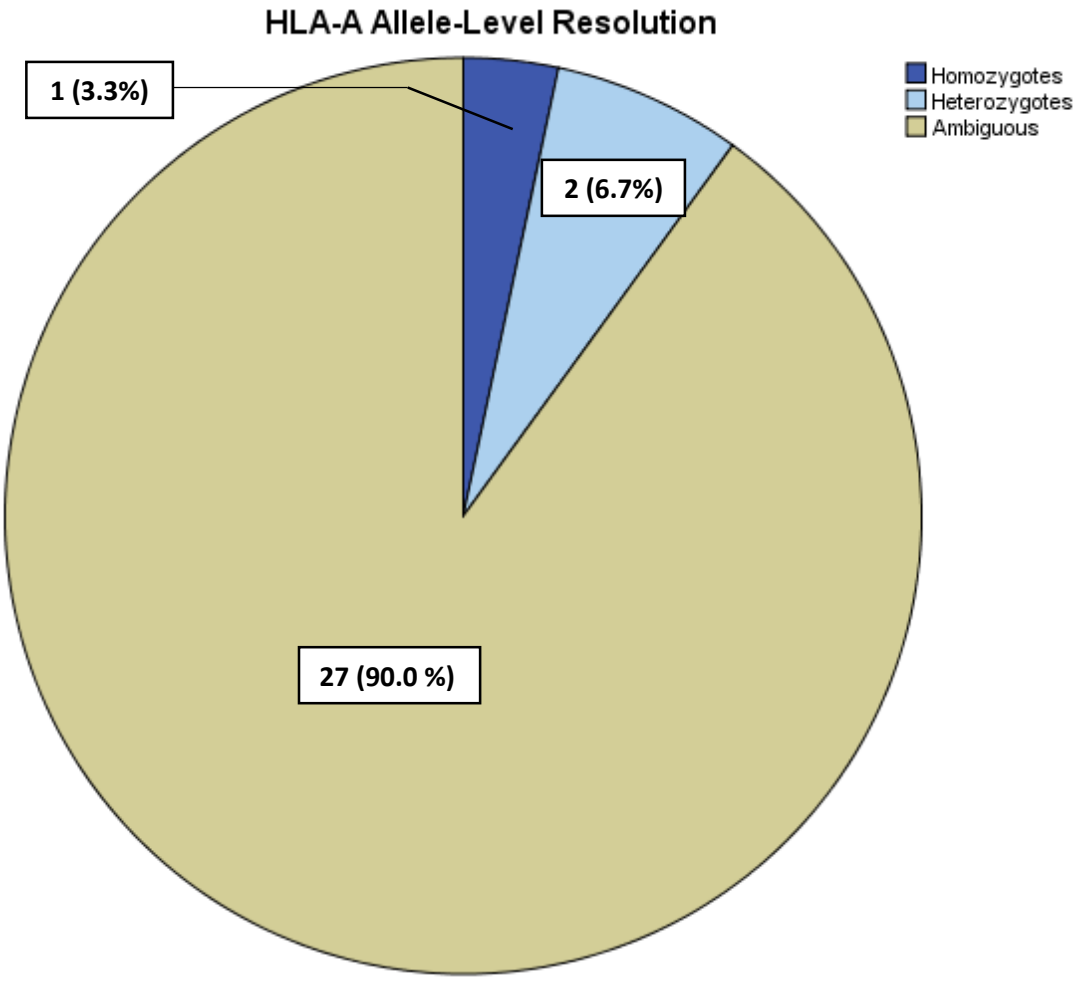


Figure 4.9: Allele-level genotypic frequencies of HLA-A samples typed.

Only one out of the 30 samples (representing 3.3%) was adequately resolved to this level as homozygous at this locus. Two samples (6.7%) were resolved as heterozygotes at this locus, while majority (90%) of the samples were ambiguous at this level of typing resolution with the current approach.

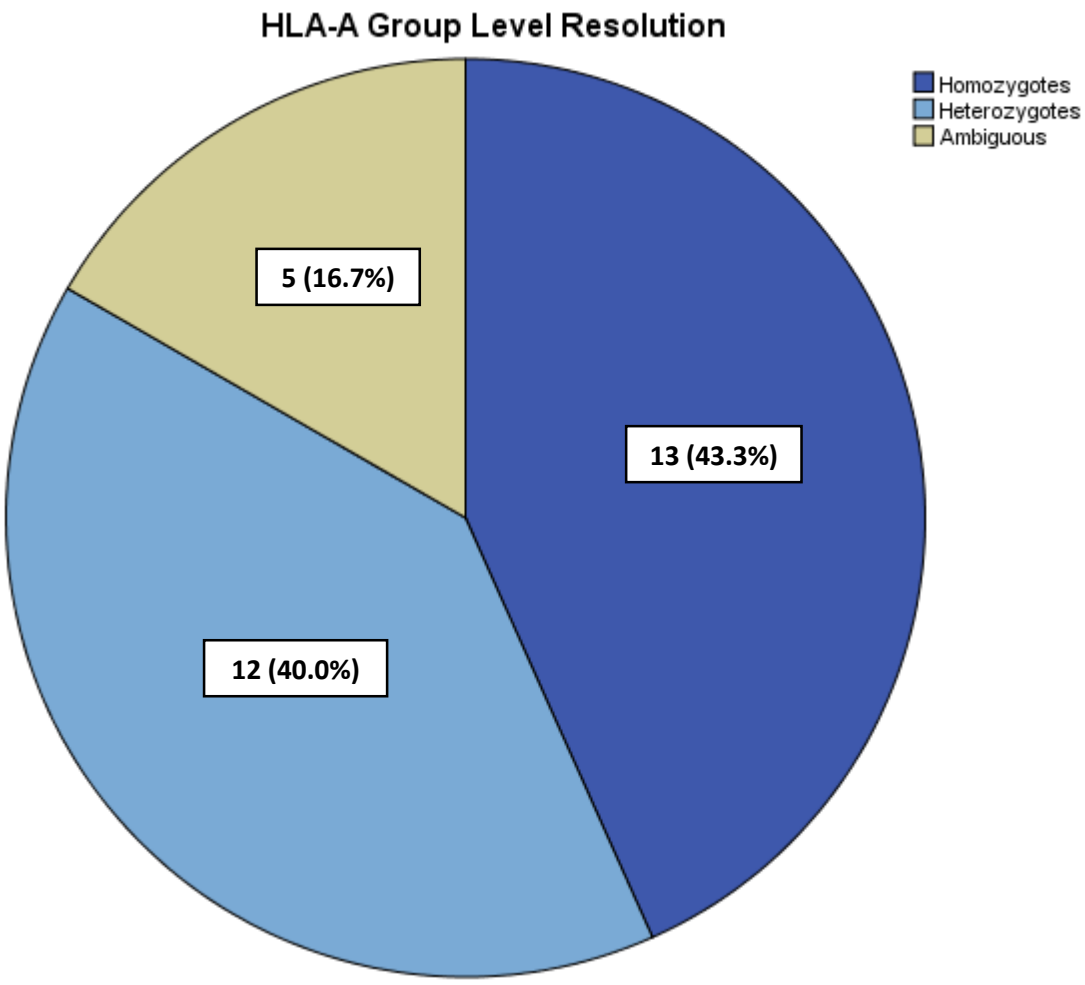


Figure 4.10: Allele group-level genotypic frequencies of HLA-A samples typed.

13 out of the 30 samples (representing 43.3%) yielded allele combinations that were homozygous at this level of resolution, while 12 samples (40.0%) of the alleles typed were heterozygous. 5 typed sample (16.7%) genotypes were ambiguous beyond resolution at this level.

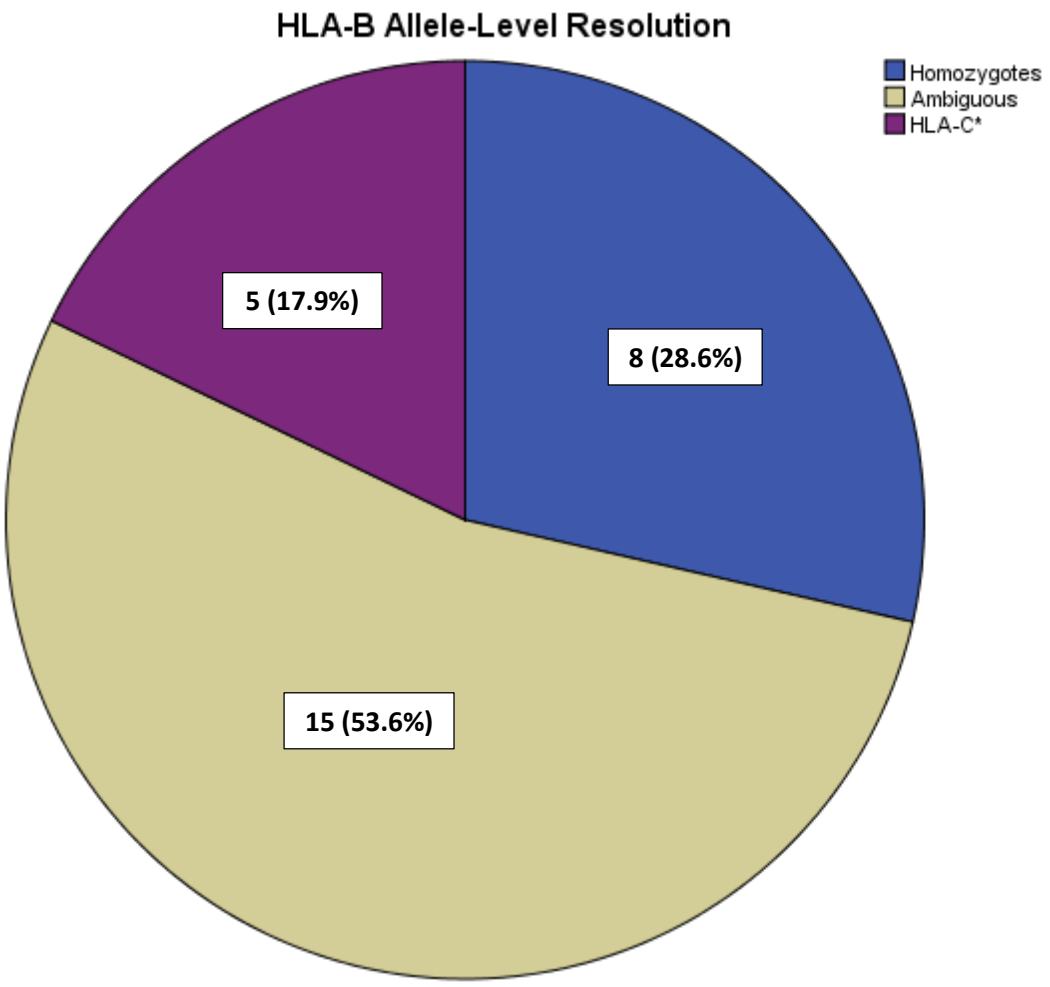


Figure 4.11: Allele-level genotypic frequencies of HLA-B samples typed.

8 samples out of a possible 28 (28.6%) produced genotypes that were homozygous at the allele-level for HLA-B locus, while 15 samples (53.6%) produced ambiguous results. 5 typed samples yielded HLA-C alleles as the best matched BLAST hits. Two samples had no sequencing data.

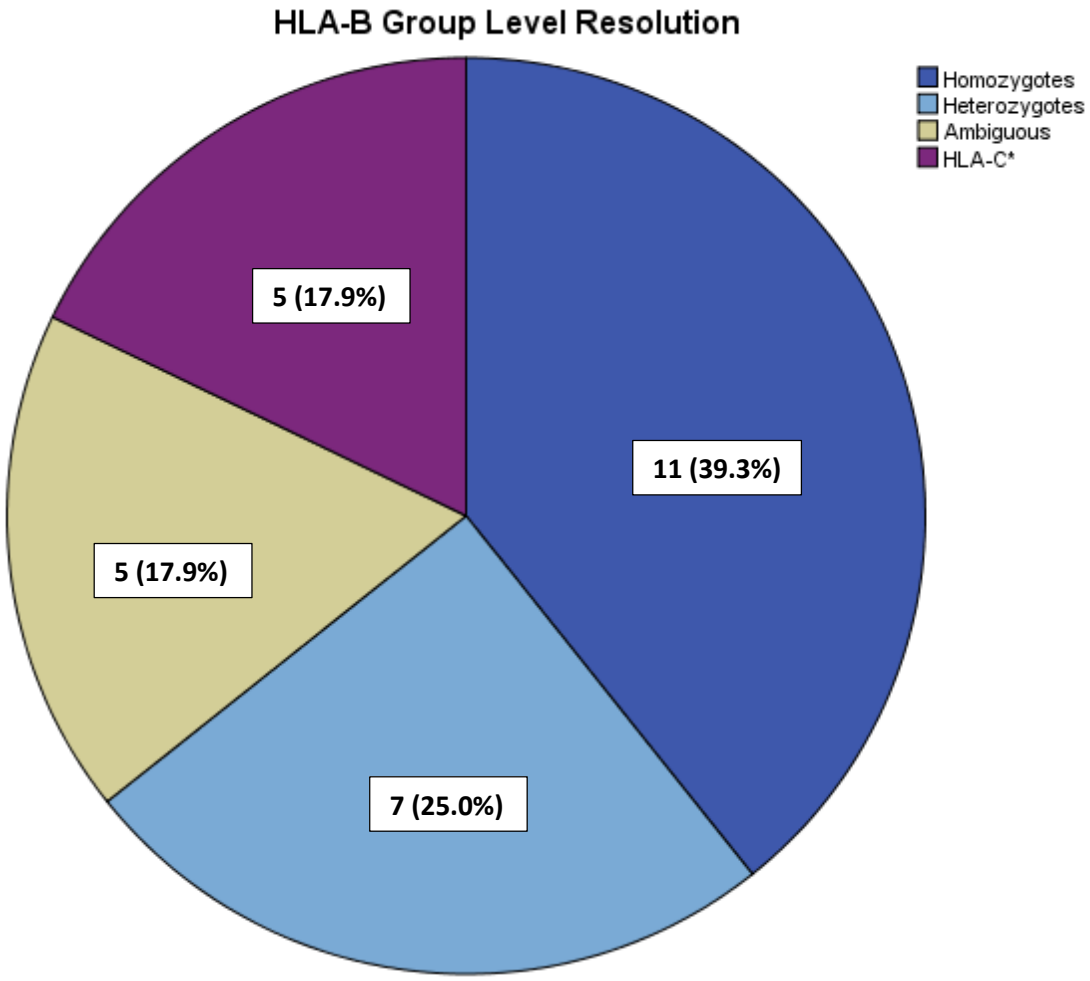


Figure 4.12: Allele group-level genotypic frequencies of HLA-B samples typed.

11 samples out 28 (39.3%) were homozygous, 7 (25%) were heterozygous, and 5 (17.9%) were ambiguous. Five (17.9%) typed samples produced hits in the HLA-C region outside the target sequence of interest. Two samples had no sequencing data.

4.5.3 Allele Frequencies and Associations

A total of 25 out of 30 samples (83.3%) were successfully typed to the allele group level for the HLA-A locus, comprising 8 (80%) cases and 17 (85%) tolerant patients, while 18 samples (60%) were successfully typed to this level for the HLA-B locus, comprising 5 (50%) rash cases and 13 (65%) tolerant patients. Fisher's exact test was used to analyse associations between the alleles observed and the occurrence of NVP-rash among the study population. No significant associations were found.

Table 4.4 shows the allele frequency distribution in the two groups of participants and their corresponding *p*-values.

Table 4.4: Allele distribution and their association to rash among typing population.

Allele(s)	**Rash, *n* (%)**	**Tolerant, *n* (%)**	**All, *n* (%)**	***p* (rash *vs.* tolerant)**
*A*01*	0 (0.0)	2 (11.8)	2 (8.0)	0.453
*A*02, A*24*	2 (25.0)	4 (23.5)	6 (24.0)	0.651
*A*03*	0 (0.0)	3 (17.6)	3 (12.0)	0.296
*A*30, A*31*	3 (37.5)	4 (23.5)	7 (28.0)	0.393
*A*68, A*69*	1 (12.5)	2 (11.8)	3 (12.0)	0.704
*B*07*	1 (20.0)	2 (15.4)	3 (16.7)	0.650
*B*15, B*44*	1 (20.0)	1 (7.7)	2 (11.1)	0.490
*B*27, B*45, B*58*	0 (0.0)	1 (7.7)	1 (5.6)	0.722
*B*42, B*49*	1 (20.0)	0 (0.0)	1 (5.6)	0.278
*B*51*	1 (20.0)	5 (38.5)	6 (33.3)	0.439
*B*53*	0 (0.0)	5 (38.5)	5 (27.8)	0.150
*B*57*	0 (0.0)	1 (7.7)	1 (5.6)	0.722

p-values significant at <0.05.
Alleles of the same frequencies are grouped together for simplification, separated by commas.

CHAPTER 5

5.0 DISCUSSION

The present study sought to genotype HLA-A and HLA-B alleles in selected HIV patients on NVP-based combination therapy in Ghana and to use the resulting data to identify any possible association (be it novel or known) between the HLA alleles observed and NVP-induced rash — as there is currently no known published data in this regard from Ghana. The majority of NVP-HLA pharmacogenetics have revealed an association between notably HLA-C*04, HLA-B*35 and HLA-DRB1*01 alleles and NVP-rash or hypersensitivity reactions among different ethnic populations (reviewed in section 2.2.3). However, given the high genetic diversity among black African populations (Campbell & Tishkoff, 2008) and the variety of associations made, it is possible that different polymorphisms and hence different HLA alleles could be responsible for similar NVP-induced adverse drug events in a Ghanaian population, hence this objective.

While there was no significant association in sex and CD4 cell count between the NVP-rash and tolerant groups, suggesting that these characteristics have no influence on the development of NVP-mediated rash contrary to what others have found (Bersoff-Matcha *et al.*, 2001; Lyons *et al.*, 2006; Phillips *et al.*, 2007; Wit *et al.*, 2008), it was not totally unexpected as such similar insignificant associations have been reported by various studies as well (Ciccacci *et al.*, 2013, 2015; Vitezica *et al.*, 2008; Yuan *et al.*, 2011). This owes largely to the fact that such associations generally require very large sample sizes to be detectible by the statistics employed.

There was however, a significant association between ‘ethnic group’ and NVP-rash, suggesting that the ethnic group of a patient influences their likelihood of developing NVP-induced rash. This is likely, considering the high genetic diversity within and among African populations (Campbell

& Tishkoff, 2008). However, given the relatively small sample size and the fact that the ethnic group data was heavily dominated by one group (i.e. 'Asantes', accounting for over 70% of all ethnic groups observed), it remains to be proven whether this is a true association or simply a reflection of ethnic group frequencies in the study population.

The modified DNA isolation protocol used (Suguna *et al.*, 2014) produced considerably satisfactory results. DNA purity ranges were normal for PCR while DNA yields were relatively high compared to the yields of the original authors primarily because of a modification at the first step of the process, by repeatedly adding more blood samples with each RBC wash (up to 900 ml) as opposed to a one time addition of 300ml as stipulated in the original protocol.

Initial polymerase chain reactions required several optimisations until final reaction conditions were attained. Results from HLA-A and -B exon 2 regions appeared much better, while exons 3 proved difficult. As to whether the latter was as a result of some specific primer properties (as suspected) or in-house laboratory techniques, requires further testing.

Employment of the services of a commercial sequencing company adds complexity to the assessment of sequencing results. Although, some results were of good quality, the same cannot be said of others, and while this is admittedly a limitation to the success of typing, it is difficult to point out the source of these disparities as they can range anywhere from between PCR product quality to product purification and cycle sequencing procedures. Nevertheless, the choice of sequencing company (and technology for that matter) remained a question of availability and accessibility.

Sanger sequence-based typing (SBT) of the core exons encoding the HLA antigen recognition site (exons 2 and 3 for HLA class I, and exon 2 for class II molecules) has been considered the 'gold

standard' of HLA typing for decades. However, the inability of this approach to fully assign all genotypes in question has been well noted alongside some measures to help resolve the problem (Erlich, 2012; Erlich *et al.*, 2001; Lazaro *et al.*, 2013). For example, Lazaro *et al.* (2013) reported the approximate percentage of samples that result in final HLA assignments following locus-specific PCR to be 60% with the remaining 40% being ambiguous genotypes that require further testing methods such as the use of SSP (sequence-specific primers), or SSOP (sequence-specific oligonucleotide probes). While, they were not explicit on the level of typing resolution being alluded to, they however, noted that these percentages will vary depending on the frequency of specific HLA alleles in the population and the characteristics of the alleles.

A major reason for ambiguous genotyping results from SBT is the inherent inability of Sanger sequencing to detect the phase of polymorphisms, thereby yielding mixed sequence data (Juhos *et al.*, 2016; Lazaro *et al.*, 2013; Mayor *et al.*, 2015). Additionally, some Class I alleles tend to be perfectly identical over exons 2 and 3, and therefore require sequencing of additional exons to discriminate between them. Unfortunately, following an explosion of HLA allele discovery in recent years, the number of ambiguous genotypes that result from traditional SBT methods have continued to rise, gradually making SBT a less preferable choice for high resolution HLA typing in favour of newer methods which make use of next generation sequencing technology (Ehrenberg *et al.*, 2014; Erlich, 2015; Juhos *et al.*, 2016; Lazaro *et al.*, 2006; Mayor *et al.*, 2015).

Results from this study can thus be said to be reflective of the current state of SBT following locus-specific PCR as the current method clearly performed better at the allele group level resolution compared with the high resolution allele-level, as follows. At the allele group level resolution, 83.3% of HLA-A samples were successfully genotyped, while only 10% were successful at the allele-level resolution, leaving as high as 90% ambiguous genotypes at this level. Similarly, for

the HLA-B locus, only 28.6% of the samples were successfully genotyped to the allele-level resolution compared to 64.3% at the allele group level. While this rate of success of genotype assignments may not be satisfactory for donor-recipient matching requirements in hematopoietic stem cell transplantation, they are still potentially useful for other applications, as allele group level typing have been frequently reported in HLA matching for solid organ transplantations (Erlich *et al.,* 2001; Hahn *et al.,* 2015) and drug adverse events association studies (Gao *et al.,* 2012; Likanonsakul *et al.,* 2009; Vitezica *et al.,* 2008; Yuan *et al.,* 2011). Nonetheless, it is undeniable that allele-level (two field level) resolution remains ideal for most clinical applications (Hahn *et al.,* 2015; Mayor *et al.,* 2015).

Finally, allele distribution analysis revealed an interesting pattern of results. Despite the fact that there was no significant association between the observed alleles and NVP-induced rash in the current study population, this was not surprising as no previous associations have been reported between the alleles observed in this study and NVP-induced rash in the general literature. However, given that none of the commonly reported associated alleles was found in the NVP-rash group from this study, it remains possible that different sets of alleles may be involved in NVP-induced rash in the Ghanaian population, but this requires further studies.

CHAPTER 6

6.0 CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The resolving power of Sanger sequence-based typing of HLA has been dwindling in recent times due, in part, to the increasing numbers of HLA alleles being discovered by the day (Erlich, 2012, 2015) and the findings of this study is no different. Out of a possible 58 samples to be typed, only 11 (less than 19%) were successfully genotyped to allele-level resolution, whereas the vast majority of genotypes were ambiguous at this resolution. At the allele group (or serotypic) level however, results were much better, with 43 (over 70%) successful genotypes. While this rate of success of genotyping is considerably lower than expected, it reflects the current state of Sanger sequence-based HLA typing following locus-specific PCR, and depending on the application of typing information, may be considered undesirable or satisfactory.

This study found no significant association between sex/CD4 cell count and the development of NVP rash as well as between the HLA alleles observed and NVP-induced rash.

There was however, a significant association between ethnic group and the occurrence of NVP-induced rash, suggesting that the ethnic group of a patient had a significant influence on their likelihood of developing NVP-rash, but this requires further testing.

6.2 LIMITATIONS

A major limitation of the current work is the small sample size. Given the thousands of HLA alleles co-dominantly expressed, with heterozygous genotypes being strongly favoured in a population, a typing sample size of 30 severely limits the observable allele frequencies from reaching numbers high enough to make significant associations by many statistical tests, such as Chi-square, Fisher's exact test, or by logistic regression analysis.

Additionally, the current sequencing method used, Sanger sequence-based typing, has its own inherent limitations. Notably, sequencing of only exons 2 and 3 (which encode the antigen recognition site and harbours majority of the genetic variation in HLA Class I molecules) can lead to ambiguous genotyping results because some alleles are perfectly identical across these exons and thus require sequencing of additional exons to resolve these ambiguities. Also, the inherent inability of Sanger sequencing to detect the phase of two or more polymorphisms frequently leads to mixed sequence data resulting from locus-specific PCR and thus often requires more expensive and laborious further testing by means of SSP-PCR or SSOP-PCR. In the present work, both SSP- and SSOP-PCR were not accessible to the researcher and hence could not be used.

Finally, the issue of ethnicity of available reference sequence databases could be a major limiting factor in the present work. Due to inadequate sequencing data generated from sub-Saharan Africa, particularly Ghana, it is likely that some of the differences in the current sequencing data generated are simply not captured in the present IPD-IMGT/HLA database.

6.3 RECOMMENDATIONS

It is recommended that further studies be done in this field of HLA typing and NVP-pharmacogenetic association studies in Ghana to help pave the way for the envisaged future adoption of pharmacogenetic testing as a standard practice, prior to HIV treatment, as the world progresses toward a more personalised medical care system.

It is also recommended that future Sanger sequence-based typing of HLA genes should consider the use of sequence-specific primers and/or sequence-specific oligonucleotide probes to limit the number of mixed sequence data that result. Considerations should also be made to sequence other exons apart from those that encode the antigen recognition sites in future works to limit the number of ambiguities that result from sequencing only the ARS regions for HLA typing. Other HLA loci such as HLA-C and -DRB1, as well as cytochrome P450 enzymes, particularly CYP 2B6 should also be considered in future studies to attempt to replicate previously reported associations in NVP-induced hypersensitivity reaction studies.

Finally, where available and accessible, next generation sequencing should be opted for, in place of Sanger sequencing, to help address the problem of polymorphism phasing and to enable for higher resolution HLA typing.

REFERENCES

Aihara, M. (2011). Pharmacogenetics of cutaneous adverse drug reactions: Pharmacogenetics of cutaneous ADR. *The Journal of Dermatology,* 38 (3), 246–254.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25 (17), 3389–3402.

Asmuth, D. & Pollard, R. (2007). In: G. Skowron & R. Ogden (eds.). *Reverse Transcriptase Inhibitors in HIV/AIDS Therapy*. Infectious Disease. Springer Science & Business Media, 303–334.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, 220 (4599), 868–871.

Bastuji-Garin, S., Fouchard, N., Bertocchi, M., Roujeau, J.-C., Revuz, J. & Wolkenstein, P. (2000). SCORTEN: a severity-of-illness score for toxic epidermal necrolysis. *Journal of Investigative Dermatology,* 115 (2), 149–153.

Bastuji-Garin, S., Rzany, B., Stern, R.S., Shear, N.H., Naldi, L. & Roujeau, J.C. (1993). Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. *Archives of Dermatology,* 129 (1), 92–96.

Battegay, M. (2014). Choice of initial therapy. *Journal of the International AIDS Society*, 17 (4 Suppl 3), 19488.

Bersoff-Matcha, S.J., Miller, W.C., Aberg, J.A., van Der Horst, C., Hamrick Jr, H.J., Powderly, W.G. & Mundy, L.M. (2001). Sex differences in nevirapine rash. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America,* 32 (1), 124–129.

Bodmer, J.G., Marsh, S.G.E., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Hansen, J.A., Mach, B., Mayr, W.R., Parham, P., Petersdorf, E.W., Sasazuki, T., Schreuder, G.M.T., Strominger, J.L., Svejgaard, A. & Terasaki, P.I. (1999). Nomenclature for factors of the HLA system, 1998. *Tissue Antigens,* 53 (4pt2), 407–446.

Boehringer-Ingelheim (2017). *VIRAMUNE® Prescribing Information*. [Online]. Available from: http://docs.boehringer-ingelheim.com/Prescribing%20Information/PIs/Viramune/Viramune.pdf. [Accessed: 26 June 2017].

Bottorff, M.B., Bright, D.R. & Kisor, D.F. (2017). Commentary: Should Pharmacogenomic Evidence Be Considered in Clinical Decision Making? Focus on Select Cardiovascular Drugs. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. [Online]. Available from: http://doi.wiley.com/10.1002/phar.1979. [Accessed: 3 July 2017].

Campbell, M.C. & Tishkoff, S.A. (2008). AFRICAN GENETIC DIVERSITY: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics,* 9, 403–433.

Carr, D.F., Bourgeois, S., Chaponda, M., Takeshita, L.Y., Morris, A.P., Castro, E.M.C., Alfirevic, A., Jones, A.R., Rigden, D.J., Haldenby, S., Khoo, S., Lalloo, D.G., Heyderman, R.S., Dandara, C., Kampira, E., van Oosterhout, J.J., Ssali, F., Munderi, P., Novelli, G., Borgiani, P., Nelson, M.R., Holden, A., Deloukas, P. & Pirmohamed, M. (2017). Genome-wide association study of nevirapine hypersensitivity in a sub-Saharan African HIV-infected population. *The Journal of Antimicrobial Chemotherapy*, 72 (4), 1152–1162.

Carr, D.F., Chaponda, M., Cornejo Castro, E.M., Jorgensen, A.L., Khoo, S., Van Oosterhout, J.J., Dandara, C., Kampira, E., Ssali, F., Munderi, P., Lalloo, D.G., Heyderman, R.S. & Pirmohamed, M. (2014). CYP2B6 c.983T>C polymorphism is associated with nevirapine hypersensitivity in Malawian and Ugandan HIV populations. *Journal of Antimicrobial Chemotherapy,* 69 (12), 3329–3334.

Carr, D.F., Chaponda, M., Jorgensen, A.L., Castro, E.C., van Oosterhout, J.J., Khoo, S.H., Lalloo, D.G., Heyderman, R.S., Alfirevic, A. & Pirmohamed, M. (2013). Association of Human Leukocyte Antigen Alleles and Nevirapine Hypersensitivity in a Malawian HIV-Infected Population. *Clinical Infectious Diseases,* 56 (9), 1330–1339.

CDC (1981). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR. Morbidity and Mortality Weekly Report*, 30 (25), 305–308.

CDC (1982). Update on acquired immune deficiency syndrome (AIDS)--United States. *MMWR. Morbidity and Mortality Weekly Report*, 31 (37), 507–508, 513–514.

Center for Drug Evaluation and Research (2016). *Genomics - Table of Pharmacogenomic Biomarkers in Drug Labeling*. [Online]. Available from: https://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm. [Accessed: 14 June 2017].

Centor, R.M. (2007). To Be a Great Physician, You Must Understand the Whole Story. *Medscape General Medicine,* 9 (1), 59.

Chantarangsu, S., Mushiroda, T., Mahasirimongkol, S., Kiertiburanakul, S., Sungkanuparph, S., Manosuthi, W., Tantisiriwat, W., Charoenyingwattana, A., Sura, T., Chantratita, W. & Nakamura, Y. (2009). HLA-B*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in HIV-infected Thai patients: *Pharmacogenetics and Genomics,* 19 (2), 139–146.

Cheeseman, S.H., Hattox, S.E., McLaughlin, M.M., Koup, R.A., Andrews, C., Bova, C.A., Pav, J.W., Roy, T., Sullivan, J.L. & Keirns, J.J. (1993). Pharmacokinetics of nevirapine: initial single-rising-dose study in humans. *Antimicrobial Agents and Chemotherapy,* 37 (2), 178–182.

Chen, J., Mannargudi, B.M., Xu, L. & Uetrecht, J. (2008). Demonstration of the metabolic pathway responsible for nevirapine-induced skin rash. *Chemical Research in Toxicology,* 21 (9), 1862–1870.

Chung, W.-H., Hung, S.-I., Hong, H.-S., Hsih, M.-S., Yang, L.-C., Ho, H.-C., Wu, J.-Y. & Chen, Y.-T. (2004). A marker for Stevens– Johnson syndrome. *Nature,* 428, 486.

Ciccacci, C., Borgiani, P., Ceffa, S., Sirianni, E., Marazzi, M.C., Altan, A.M.D., Paturzo, G., Bramanti, P., Novelli, G. & Palombi, L. (2010). Nevirapine-induced hepatotoxicity and pharmacogenetics: a retrospective study in a population from Mozambique. *Pharmacogenomics,* 11 (1), 23–31.

Ciccacci, C., Di Fusco, D., Marazzi, M.C., Zimba, I., Erba, F., Novelli, G., Palombi, L., Borgiani, P. & Liotta, G. (2013). Association between CYP2B6 polymorphisms and Nevirapine-induced SJS/TEN: a pharmacogenetics study. *European Journal of Clinical Pharmacology,* 69 (11), 1909–1916.

Ciccacci, C., Rufini, S., Mancinelli, S., Buonomo, E., Giardina, E., Scarcella, P., Marazzi, M., Novelli, G., Palombi, L. & Borgiani, P. (2015). A Pharmacogenetics Study in Mozambican Patients Treated with Nevirapine: Full Resequencing of TRAF3IP2 Gene Shows a Novel Association with SJS/TEN Susceptibility. *International Journal of Molecular Sciences,* 16 (3), 5830–5838.

Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H. & Vogt, P. (1986). What to call the AIDS virus? *Nature,* 321 (6065), 10.

Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H. & others (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine,* 365 (6), 493–505.

Consortium, I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. *Nature,* 431 (7011), 931–945.

Dalal, B., Shankarkumar, A. & Ghosh, K. (2015). Individualization of antiretroviral therapy - Pharmacogenomic aspect. *The Indian Journal of Medical Research,* 142 (6), 663–674.

Dao, R.-L., Su, S.-C. & Chung, W.-H. (2015). Recent advances of pharmacogenomics in severe cutaneous adverse reactions: immune and nonimmune mechanisms. *Asia Pacific Allergy,* 5 (2), 59.

Daw, M.A., El-Bouzedi, A., Ahmed, M.O. & Dau, A.A. (2017). Molecular and epidemiological characterization of HIV-1 subtypes among Libyan patients. *BMC Research Notes*. [Online]. 10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5410017/. [Accessed: 3 July 2017].

Dupont, B. & Svejgaard, A. (1977). HLA and disease. *Transplantation Proceedings*, 9 (1), 1271–1274.

Durack, D.T. (1981). Opportunistic infections and Kaposi's sarcoma in homosexual men. *The New England Journal of Medicine,* 305 (24), 1465–1467.

Ehrenberg, P.K., Geretz, A., Baldwin, K.M., Apps, R., Polonis, V.R., Robb, M.L., Kim, J.H., Michael, N.L. & Thomas, R. (2014). High-throughput multiplex HLA genotyping by next-generation sequencing using multi-locus individual tagging. *BMC Genomics,* 15 (1), 864.

Erlich, H. (2012). HLA DNA typing: past, present, and future. *Tissue Antigens,* 80 (1), 1–11.

Erlich, H.A. (2015). HLA typing using next generation sequencing: An overview. *Human Immunology,* 76 (12), 887–890.

Erlich, H.A., Opelz, G. & Hansen, J. (2001). HLA DNA typing and transplantation. *Immunity,* 14 (4), 347–356.

Fasano, M.E., Dametto, E. & D'Alfonso, S. (2015). HLA Genotyping: Methods for the Identification of the HLA-DQ2,-DQ8 Heterodimers Implicated in Celiac Disease (CD) Susceptibility. *Methods in Molecular Biology*, 1326, 79–92.

Flomenberg, N., Baxter-Lowe, L.A., Confer, D., Fernandez-Vina, M., Filipovich, A., Horowitz, M., Hurley, C., Kollman, C., Anasetti, C., Noreen, H., Begovich, A., Hildebrand, W., Petersdorf, E., Schmeckpeper, B., Setterholm, M., Trachtenberg, E., Williams, T., Yunis, E. & Weisdorf, D. (2004). Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood*, 104 (7), 1923–1930.

Fricke-Galindo, I., LLerena, A. & López-López, M. (2017). An update on HLA alleles associated with adverse drug reactions. *Drug Metabolism and Personalized Therapy,* 32 (2), 73–87.

Fürst, D., Müller, C., Vucinic, V., Bunjes, D., Herr, W., Gramatzki, M., Schwerdtfeger, R., Arnold, R., Einsele, H., Wulf, G., Pfreundschuh, M., Glass, B., Schrezenmeier, H., Schwarz, K. & Mytilineos, J. (2013). High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. *Blood*, 122 (18), 3220–3229.

Gallo, R., Sarin, P., Gelmann, E., Robert-Guroff, M., Richardson, E., Kalyanaraman, V., Mann, D., Sidhu, G., Stahl, R., Zolla-Pazner, S., Leibowitch, J. & Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science,* 220 (4599), 865–867.

Gallo, R.C. & Montagnier, L. (2003). The discovery of HIV as the cause of AIDS. *New England Journal of Medicine,* 349 (24), 2283–2285.

Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J. & Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (New York, N.Y.)*, 224 (4648), 500–503.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. & Hahn, B.H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature,* 397 (6718), 436–441.

Gao, S., Gui, X.-E., Liang, K., Liu, Z., Hu, J. & Dong, B. (2012). HLA-dependent hypersensitivity reaction to nevirapine in Chinese Han HIV-infected patients. *AIDS Research and Human Retroviruses,* 28 (6), 540–543.

Ghana AIDS Commission (2016). *Ghana National HIV and AIDS Strategic Plan 2016 - 2020*. [Online]. Available from: http://www.ghanaids.gov.gh/gac1/pubs/COMPREHENSIVE%20NSP%202016-2020.pdf. [Accessed: 12 June 2017].

Greene, W.C. (2007). A history of AIDS: looking back to see ahead. *European journal of immunology,* 37 (S1), S94–S102.

Gulick, R.M. (2007). Antiretroviral management of treatment-naive patients. *Infectious Disease Clinics of North America,* 21 (1), 71–84.

Hahn, A.B., Bravo-Egana, V., Jackstadt, J.L., Conti, D.J. & Duquesnoy, R.J. (2015). HLA-A2 reactive antibodies in a patient who types as HLA-A2: The importance of high resolution typing and epitope-based antibody analysis. *Transplant Immunology,* 32 (3), 141–143.

Hahn, B.H., Shaw, G.M., De, K.M., Sharp, P.M. & others (2000). AIDS as a zoonosis: scientific and public health implications. *Science,* 287 (5453), 607–614.

Hajeer, A.H. & Hutchinson, I.V. (2001). Influence of TNFalpha gene polymorphisms on TNFalpha production and disease. *Human Immunology,* 62 (11), 1191–1199.

Havlir, D., Cheeseman, S.H., McLaughlin, M., Murphy, R., Erice, A., Spector, S.A., Greenough, T.C., Sullivan, J.L., Hall, D. & Myers, M. (1995). High-dose nevirapine: safety, pharmacokinetics, and antiviral effect in patients with human immunodeficiency virus infection. *The Journal of Infectious Diseases,* 171 (3), 537–545.

Havlir, D.V., Eastman, S., Gamst, A. & Richman, D.D. (1996). Nevirapine-resistant human immunodeficiency virus: kinetics of replication and estimated prevalence in untreated patients. *Journal of Virology,* 70 (11), 7894–7899.

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine,* 18 (3), 182–192.

Hemelaar, J., Gouws, E., Ghys, P.D. & Osmanov, S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS (London, England)*, 25 (5), 679–689.

Hung, S.-I., Chung, W.-H., Jee, S.-H., Chen, W.-C., Chang, Y.-T., Lee, W.-R., Hu, S.-L., Wu, M.-T., Chen, G.-S., Wong, T.-W. & others (2006). Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenetics and Genomics,* 16 (4), 297–306.

Hung, S.-I., Chung, W.-H., Liou, L.-B., Chu, C.-C., Lin, M., Huang, H.-P., Lin, Y.-L., Lan, J.-L., Yang, L.-C., Hong, H.-S. & others (2005). HLA-B* 5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proceedings of the National Academy of Sciences of the United States of America,* 102 (11), 4134–4139.

Ichiki, Y., Bowlus, C.L., Shimoda, S., Ishibashi, H., Vierling, J.M. & Gershwin, M.E. (2006). T cell immunity and graft-versus-host disease (GVHD). *Autoimmunity Reviews*, 5 (1), 1–9.

Juhos, S., Rigó, K. & Horváth, G. (2016). *On Genotyping Polymorphic HLA Genes — Ambiguities and Quality Measures Using NGS*. [Online]. Available from: http://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/on-genotyping-polymorphic-hla-genes-ambiguities-and-quality-measures-using-ngs. [Accessed: 23 June 2017].

Kane, M.D. & Brewer, J.L. (2007). An information technology emphasis in biomedical informatics education. *Journal of Biomedical Informatics*, 40 (1), 67–72.

Kelsey, J.L., Whittemore, A.S., Evans, A.S. & Thompson, W.D. (1996). Google-Books-ID: Xnz6VgL22osC. *Methods in Observational Epidemiology,* 2nd Ed. Oxford University Press.

Kim, J.H. (2002). Bioinformatics and genomic medicine. *Genetics in Medicine,* 4 (S6), 62S–65S.

Kusinitz, M., Braunstein, E. & Wilson, C.A. (2017). Advancing Public Health Using Regulatory Science to Enhance Development and Regulation of Medical Products: Food and Drug Administration Research at the Center for Biologics Evaluation and Research. *Frontiers in Medicine,* 4, 71.

Lamson, M.J., Sabo, J.P., Macgregor, T.R., Pav, J.W., Rowland, L., Hawi, A., Cappola, M. & Robinson, P. (1999). Single dose pharmacokinetics and bioavailability of nevirapine in healthy volunteers. *Biopharmaceutics & Drug Disposition*, 20 (6), 285–291.

Lazaro, A., Tu, B., Yang, R., Xiao, Y., Kariyawasam, K., Ng, J. & Hurley, C.K. (2013). Human Leukocyte Antigen (HLA) Typing by DNA Sequencing. In: A. A. Zachary & M. S. Leffell (eds.). *Transplantation Immunology,* [Online]. Totowa, NJ: Humana Press, 161–195. Available from: http://link.springer.com/10.1007/978-1-62703-493-7_9. [Accessed: 11 May 2017].

Lazaro, A.M., Cao, K., Masaberg, C., Steiner, N.K., Xiao, Y., Tu, B., Turner, V., Nickerson, P., Stoll, S., Schall, C., Valdez, R., Ng, J., Hartzman, R.J. & Hurley, C.K. (2006). Twenty-three novel HLA-B alleles identified during intermediate-resolution testing. *Tissue Antigens,* 68 (3), 245–248.

Lee, S.J., Klein, J., Haagenson, M., Baxter-Lowe, L.A., Confer, D.L., Eapen, M., Fernandez-Vina, M., Flomenberg, N., Horowitz, M., Hurley, C.K., Noreen, H., Oudshoorn, M., Petersdorf, E., Setterholm, M., Spellman, S., Weisdorf, D., Williams, T.M. & Anasetti, C. (2007). High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*, 110 (13), 4576–4583.

Likanonsakul, S., Rattanatham, T., Feangvad, S., Uttayamakul, S., Prasithsirikul, W., Tunthanathip, P., Nakayama, E.E. & Shioda, T. (2009). HLA-Cw*04 allele associated with nevirapine-induced rash in HIV-infected Thai patients. *AIDS Research and Therapy,* 6 (1), 22.

Lin, H. & Gong, Y.-Z. (2017). Association of HLA-B27 with ankylosing spondylitis and clinical features of the HLA-B27-associated ankylosing spondylitis: a meta-analysis. *Rheumatology International*. [Online]. Available from: http://link.springer.com/10.1007/s00296-017-3741-2. [Accessed: 8 June 2017].

Liptrott, N.J., Pushpakom, S., Wyen, C., Fätkenheuer, G., Hoffmann, C., Mauss, S., Knechten, H., Brockmeyer, N.H., Hopper-Borge, E., Siccardi, M., Back, D.J., Khoo, S.H., Pirmohamed, M. & Owen, A. (2012). Association of ABCC10 polymorphisms with nevirapine plasma concentrations in the German Competence Network for HIV/AIDS: *Pharmacogenetics and Genomics,* 22 (1), 10–19.

Lyons, F., Hopkins, S., Kelleher, B., McGeary, A., Sheehan, G., Geoghegan, J., Bergin, C., Mulcahy, F.M. & McCormick, P.A. (2006). Maternal hepatotoxicity with nevirapine as part of combination antiretroviral therapy in pregnancy. *HIV Medicine,* 7 (4), 255–260.

Macher, A.M. & Goosby, E.P. (2016). Acquired immune deficiency syndrome (AIDS). *Access Science*. [Online]. Available from: http://accessscience.com.sci-hub.cc/content/acquired-immune-deficiency-syndrome-aids/007175. [Accessed: 6 June 2017].

Mallal, S., Phillips, E., Carosi, G., Molina, J.-M., Workman, C., Tomazic, J., Jägel-Guedes, E., Rugina, S., Kozyrev, O., Cid, J.F., Hay, P., Nolan, D., Hughes, S., Hughes, A., Ryan, S., Fitch, N., Thorborn, D., Benbow, A. & PREDICT-1 Study Team (2008). HLA-B*5701 screening for hypersensitivity to abacavir. *The New England Journal of Medicine,* 358 (6), 568–579.

Marsh, S.G.E., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Fernández-Viña, M., Geraghty, D.E., Holdsworth, R., Hurley, C.K., Lau, M., Lee, K.W., Mach, B., Maiers, M., Mayr, W.R., Müller, C.R., Parham, P., Petersdorf, E.W., Sasazuki, T., Strominger, J.L., Svejgaard, A., Terasaki, P.I., Tiercy, J.M. & Trowsdale, J. (2010). Nomenclature for factors of the HLA system, 2010. *Tissue Antigens,* 75 (4), 291–455.

Marsh, S.G.E., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Geraghty, D.E., Hansen, J.A., Hurley, C.K., Mach, B., Mayr, W.R., Parham, P., Petersdorf, E.W., Sasazuki, T., Schreuder, G.M.T., Strominger, J.L., Svejgaard, A., Terasaki, P.I. & Trowsdale, J. (2005). Nomenclature for factors of the HLA system, 2004. *Tissue Antigens,* 65 (4), 301–369.

Marsh, S.G.E., Parham, P. & Barber, L.D. (2000). 13 - HLA and Disease. In: *The HLA FactsBook*. Factsbook. [Online]. London: Academic Press, 79–83. Available from: http://www.sciencedirect.com/science/article/pii/B9780125450256501416. [Accessed: 23 June 2017].

Marsh, S.G.E. & WHO Nomenclature Committee for Factors of the HLA System (2003). Nomenclature for factors of the HLA system, update December 2002. *Tissue Antigens,* 61 (5), 412–413.

Martin, A.M., Nolan, D., James, I., Cameron, P., Keller, J., Moore, C., Phillips, E., Christiansen, F.T. & Mallal, S. (2005). Predisposition to nevirapine hypersensitivity associated with HLA-DRB1*0101 and abrogated by low CD4 T-cell counts. *AIDS (London, England)*, 19 (1), 97–99.

Mayor, N.P., Robinson, J., McWhinnie, A.J.M., Ranade, S., Eng, K., Midwinter, W., Bultitude, W.P., Chin, C.-S., Bowman, B., Marks, P., Braund, H., Madrigal, J.A., Latham, K. & Marsh, S.G.E. (2015). HLA Typing for the Next Generation. *PLoS ONE*, 10 (5), e0127153.

McCarthy, J.J., McLeod, H.L. & Ginsburg, G.S. (2013). Genomic Medicine: A Decade of Successes, Challenges, and Opportunities. *Science Translational Medicine,* 5 (189), 189sr4-189sr4.

Mittmann, N., Knowles, S.R., Koo, M., Shear, N.H., Rachlis, A. & Rourke, S.B. (2012). Incidence of toxic epidermal necrolysis and Stevens-Johnson Syndrome in an HIV cohort: an observational, retrospective case series study. *American Journal of Clinical Dermatology,* 13 (1), 49–54.

National HIV/AIDS/ STI Control Programme (2008). *Guidelines for antiretroviral therapy in Ghana.* [Online]. Available from: http://www.who.int/hiv/pub/guidelines/ghana_art.pdf. [Accessed: 20 May 2017].

Nevimune® product insert. *Summary of Product Characteristics/Package Leaflet: Information for the User*.

O'brien, W.A. (2000). Resistance against reverse transcriptase inhibitors. *Clinical Infectious Diseases,* 30 (Supplement 2), S185–S192.

Osoegawa, K., Mack, S.J., Udell, J., Noonan, D.A., Ozanne, S., Trachtenberg, E. & Prestegaard, M. (2016). HLA Haplotype Validator for quality assessments of HLA typing. *Human Immunology,* 77 (3), 273–282.

Palella, F.J.J., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., Aschman, D.J. & Holmberg, S.D. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine,* 338 (13), 853–860.

Paunić, V., Gragert, L., Madbouly, A., Freeman, J. & Maiers, M. (2012). Measuring Ambiguity in HLA Typing Methods. *PLoS ONE*. [Online]. 7 (8). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3430707/. [Accessed: 23 June 2017].

Pavlos, R. & Phillips, E.J. (2012). Individualization of antiretroviral therapy. *Pharmacogenomics and Personalized Medicine,* 5, 1–17.

Pennings, P.S. (2013). HIV drug resistance: problems and perspectives. *Infectious Disease Reports*, 5 (1S), 5.

Phillips, E., Bartlett, J.A., Sanne, I., Lederman, M.M., Hinkle, J., Rousseau, F., Dunn, D., Pavlos, R., James, I., Mallal, S. & Haas, D.W. (2013). Associations between HLA-DRB1*0102, HLA-B*5801 and Hepatotoxicity during Initiation of Nevirapine-Containing Regimens in South Africa. *Journal of Acquired Immune Deficiency Syndromes*, 62 (2), e55–e57.

Phillips, E., Gutiérrez, S., Jahnke, N., Yip, B., Lima, V.D., Hogg, R.S., Harrigan, P.R. & Montaner, J.S.G. (2007). Determinants of nevirapine hypersensitivity and its effect on the association between hepatitis C status and mortality in antiretroviral drug-naive HIV-positive patients. *AIDS (London, England)*, 21 (12), 1561–1568.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D. & Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences,* 77 (12), 7415–7419.

Pollard, R.B., Robinson, P. & Dransfield, K. (1998). Safety profile of nevirapine, a nonnucleoside reverse transcriptase inhibitor for the treatment of human immunodeficiency virus infection. *Clinical Therapeutics,* 20 (6), 1071–1092.

Popovic, M., Sarngadharan, M.G., Read, E. & Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science (New York, N.Y.)*, 224 (4648), 497–500.

Popovic, M., Shenton, J.M., Chen, J., Baban, A., Tharmanathan, T., Mannargudi, B., Abdulla, D. & Uetrecht, J.P. (2010). Nevirapine Hypersensitivity. In: J. Uetrecht (ed.). *Adverse Drug Reactions*. [Online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 437–451. Available from: http://link.springer.com/10.1007/978-3-642-00663-0_15. [Accessed: 8 May 2017].

Pubchem (2017). *Nevirapine | C15H14N4O - PubChem*. [Online]. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/nevirapine. [Accessed: 20 June 2017].

Richman, D.D., Havlir, D., Corbeil, J., Looney, D., Ignacio, C., Spector, S.A., Sullivan, J., Cheeseman, S., Barringer, K. & Pauletti, D. (1994). Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *Journal of Virology,* 68 (3), 1660–1666.

Riska, P., Lamson, M., MacGregor, T., Sabo, J., Hattox, S., Pav, J. & Keirns, J. (1999). Disposition and biotransformation of the antiretroviral drug nevirapine in humans. *Drug Metabolism and Disposition*, 27 (8), 895–901.

Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P. & Marsh, S.G.E. (2015). The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research,* 43 (Database issue), D423-431.

Robinson, J., Halliwell, J.A. & Marsh, S.G.E. (2014). IMGT/HLA and the Immuno Polymorphism Database. *Methods in Molecular Biology (Clifton, N.J.)*, 1184, 109–121.

Robinson, J., Soormally, A.R., Hayhurst, J.D. & Marsh, S.G.E. (2016). The IPD-IMGT/HLA Database - New developments in reporting HLA variation. *Human Immunology,* 77 (3), 233–237.

Roujeau, J.C. (1994). The spectrum of Stevens-Johnson syndrome and toxic epidermal necrolysis: a clinical classification. *The Journal of Investigative Dermatology,* 102 (6), 28S–30S.

Roujeau, J.-C., Allanore, L., Liss, Y. & Mockenhaupt, M. (2009). Severe cutaneous adverse reactions to drugs (SCAR): definitions, diagnostic criteria, genetic predisposition. *Dermatol Sinica*, 27 (2), 203–209.

Roujeau, J.C. & Stern, R.S. (1994). Severe Adverse Cutaneous Reactions to Drugs. *New England Journal of Medicine,* 331 (19), 1272–1285.

Rufini, S., Ciccacci, C., Politi, C., Giardina, E., Novelli, G. & Borgiani, P. (2015). Stevens-Johnson syndrome and toxic epidermal necrolysis: an update on pharmacogenetics studies in drug-induced severe skin reaction. *Pharmacogenomics,* 16 (17), 1989–2002.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America,* 74 (12), 5463–5467.

Schüpbach, J., Popovic, M., Gilden, R.V., Gonda, M.A., Sarngadharan, M.G. & Gallo, R.C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science (New York, N.Y.)*, 224 (4648), 503–505.

Scourfield, A., Waters, L. & Nelson, M. (2011). Drug combinations for HIV: what's new? *Expert Review of Anti-infective Therapy,* 9 (11), 1001–1011.

Sharp, P.M., Bailes, E., Chaudhuri, R.R., Rodenburg, C.M., Santiago, M.O. & Hahn, B.H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 356 (1410), 867–876.

Sharp, P.M. & Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine,* 1 (1), a006841–a006841.

Shenton, J.M., Popovic, M. & Uetrecht, J.P. (2007). Nevirapine hypersensitivity. In: *Drug Hypersensitivity*. [Online]. Karger Publishers, 115–128. Available from: http://www.karger.com/Article/Abstract/104195. [Accessed: 6 June 2017].

Sluis-Cremer, N., Temiz, N.A. & Bahar, I. (2004). Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Current HIV Research,* 2 (4), 323–332.

Somkrua, R., Eickman, E.E., Saokaew, S., Lohitnavy, M. & Chaiyakunapruk, N. (2011). Association of HLA-B*5801 allele and allopurinol-induced Stevens Johnson syndrome and toxic epidermal necrolysis: a systematic review and meta-analysis. *BMC Medical Genetics,* 12, 118.

Suguna, S., Nandal, D., Kamble, S., Bharatha, A. & Kunkulol, R. (2014). Genomic DNA isolation from human whole blood samples by non enzymatic salting out method. *Int J Pharm Pharm Sci,* 6, 198–199.

Tang, M.W. & Shafer, R.W. (2012). HIV-1 antiretroviral resistance. *Drugs*, 72 (9), e1–e25.

Tassaneeyakul, W., Jantararoungtong, T., Chen, P., Lin, P.-Y., Tiamkao, S., Khunarkornsiri, U., Chucherd, P., Konyoung, P., Vannaprasaht, S., Choonhakarn, C., Pisuttimarn, P., Sangviroon, A. & Tassaneeyakul, W. (2009). Strong association between HLA-B*5801 and allopurinol-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in a Thai population. *Pharmacogenetics and Genomics,* 19 (9), 704–709.

Tozzi, V. (2010). Pharmacogenetics of antiretrovirals. *Antiviral Research,* 85 (1), 190–200.

UNAIDS Fact Sheet (2016). *Fact sheet - Latest statistics on the status of the AIDS epidemic | UNAIDS*. [Online]. 2016. Available from: http://www.unaids.org/en/resources/fact-sheet. [Accessed: 15 June 2017].

Vadakekolathu, J. & Rutella, S. (2017). T-Cell Manipulation Strategies to Prevent Graft-Versus-Host Disease in Haploidentical Stem Cell Transplantation. *Biomedicines,* 5 (2).

Vitezica, Z.G., Milpied, B., Lonjou, C., Borot, N., Ledger, T.N., Lefebvre, A. & Hovnanian, A. (2008). HLA-DRB1*01 associated with cutaneous hypersensitivity induced by nevirapine and efavirenz. *AIDS (London, England)*, 22 (4), 540–541.

Wajant, H., Pfizenmaier, K. & Scheurich, P. (2003). Tumor necrosis factor signaling. *Cell Death & Differentiation*, 10 (1), 45–65.

Whirl-Carrillo, M., McDonagh, E., Hebert, J., Gong, L., Sangkuhl, K., Thorn, C., Altman, R. & Klein, T. (2012). Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology and Therapeutics,* 92 (4), 414–417.

WHO (2017). *WHO | HIV/AIDS*. [Online]. 2017. WHO. Available from: http://www.who.int/gho/hiv/en/. [Accessed: 3 June 2017].

Wit, F.W.N.M., Kesselring, A.M., Gras, L., Richter, C., van der Ende, M.E., Brinkman, K., Lange, J.M.A., de Wolf, F. & Reiss, P. (2008). Discontinuation of Nevirapine Because of Hypersensitivity Reactions in Patients with Prior Treatment Experience, Compared with Treatment-Naive Patients: The ATHENA Cohort Study. *Clinical Infectious Diseases,* 46 (6), 933–940.

Wong, K.C., Kennedy, P.J. & Lee, S. (1999). Clinical manifestations and outcomes in 17 cases of Stevens-Johnson syndrome and toxic epidermal necrolysis. *The Australasian Journal of Dermatology,* 40 (3), 131–134.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134.

Yuan, J., Guo, S., Hall, D., Cammett, A.M., Jayadev, S., Distel, M., Storfer, S., Huang, Z., Mootsikapun, P., Ruxrungtham, K., Podzamczer, D. & Haas, D.W. (2011). Toxicogenomics of nevirapine-associated cutaneous and hepatic adverse events among populations of African, Asian, and European descent: *AIDS*, 25 (10), 1271–1280.

Zambrano-Zaragoza, J.F., Agraz-Cibrian, J.M., González-Reyes, C., Durán-Avelar, M. de J. & Vibanco-Pérez, N. (2013). Ankylosing Spondylitis: From Cells to Genes. *International Journal of Inflammation*. [Online]. 2013. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3736459/. [Accessed: 8 June 2017].

APPENDICES

APPENDIX 1

Participant Information Leaflet and Consent Form

This leaflet must be given to all prospective participants to enable them know enough about the research before deciding to or not to participate

Title of Research: Investigating the pharmacogenetic basis of Nevirapine-induced SJS/TEN among selected HIV patients in Ghana.

Name(s) and affiliation(s) of researcher(s):
This study is being conducted by the under listed persons:

1. **Dr. Kwabena Owusu-Danquah,** Lecturer/Principal Investigator (PI).
   Department of Medical Laboratory Technology, School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
   Contact: +233 233 224 002, or +233 543 913 833
2. **Dr. Fareed K.N. Arthur,** Lecturer/co-investigator.
   Department of Biochemistry and Biotechnology, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
3. **Mr. Ali Baba Issah,** Student /co-investigator.
   Department of Biochemistry and Biotechnology, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

Background (Please explain simply and briefly what the study is about):
One of the most important and effective drugs used in HIV treatment in Ghana is Nevirapine, a non-nucleoside reverse transcriptase inhibitor (NNRTI). Despite its effectiveness, the use of Nevirapine is often associated with some disturbing adverse effects such as hypersensitivity reactions including mild to serious skin rash and/or hepatic (liver) disorder. This skin rash condition, often referred to as Stevens-Johnson syndrome or toxic epidermal necrolysis (SJS/TEN) depending on the severity, poses serious health dangers to patients on the drug causing them in many cases to discontinue treatment, and thereby threatening the continual and effective use of this drug. In this study, we seek to identify what genetic factors predispose a person, specifically the Ghanaian patient, to developing the SJS/TEN condition as a result of Nevirapine treatment so as to help prevent future occurrences.

Purpose(s) of research:
This research is aimed at alleviating the unfortunate conditions of Stevens-Johnson syndrome and/or toxic epidermal necrolysis that some HIV patients on highly active antiretroviral therapy (HAART), specifically those including Nevirapine, develop upon treatment. We hope to learn about the causes of this condition and how to prevent them in future treatment.
You have been selected as a possible subject in this study because you have been clinically diagnosed with HIV type 1 and you received Nevirapine as part of your HAART treatment.

Procedure of the research, what shall be required of each participant and approximate total number of participants that would be involved in the research:
There will be approximately between 200 and 300 persons enrolled in this study. If you decide to participate in this study, we will collect some health information from you and your healthcare provider (i.e. your hospital) such as CD4 cell count, liver condition and other general health information. We will also collect personal information such as your sex, age, weight, height and ethnic group (but NOT your name) to aid with our research. Additionally, we will also collect a 10ml blood sample from you from which genetic material for the research will be extracted. We may also take a photograph of you showing the symptoms of SJS/TEN without showing your face or any other identifiable body part of yours, or any part of your body which you do not wish to be taken.
All information/samples taken from you will be made available only to investigators of this study for research purposes and none other. Information/samples taken will be stored on an encrypted computer/secure location where only authorised persons have access to.
There are no third party institutions/grantors/sponsors/funders of this research and therefore we are not bound by any such parties to release any of your information to them.

Risk(s):
There are no known risks associated with this study.

Benefit(s):
You will not (directly) benefit from taking part in this research study.
There may be no direct personal benefit to you in taking part in this study other than a sense of helping the public at large. Knowledge may be gained from your participation that will benefit others. Indeed, the results of this research may prevent future HIV patients in Ghana from having to suffer from the devastating effects of SJS/TEN and help medical science to gain further understanding on how to treat SJS/TEN in future.

Confidentiality:
All information collected in this study will be given code numbers. No name will be recorded. Data collected cannot be linked to you in anyway. No name or identifier will be used in any publication or reports from this study. However, as part of our responsibility to conduct this research properly, we may allow officials from the Food and Drugs Board (for clinical trials) and ethics committees to have access to your records.

Voluntariness:
Taking part in this study should be out of your own free will. You are not under obligation to. Research is entirely voluntary.

Alternatives to participation:
If you choose not to participate, this will not affect your treatment in this hospital/institution in any way.

Withdrawal from the research:
You may choose to withdraw from the research at any time without having to explain yourself. You may also choose not to answer any question you find uncomfortable or private.

Consequence of Withdrawal: There will be no consequence, loss of benefit or care to you if you choose to withdraw from the study. Please note however, that some of the information that may have been obtained from you without identifiers (name etc), before you chose to withdraw, may have been modified or used in analysis reports and publications. These cannot be removed anymore. We do promise to make good faith effort to comply with your wishes as much as practicable.

Costs/Compensation: You will not receive any monetary compensation or gift for participating in this research study. You should not expect anyone to pay you for pain, worry, lost income, or non-medical care costs that occur from taking part in this research study.
Materials obtained from you in this research may NOT be used for commercial purposes and hence you should not expect any form of compensation in that regard.

Contacts: If you have any questions concerning this study, please do not hesitate to contact Dr. Kwabena Owusu-Danquah on 0233 224 002 or 0543 913 833.

Further, if you have any concern about the conduct of this study, your welfare or your rights as a research participant, you may contact:

The Office of the Chairman
Committee on Human Research and Publication Ethics
Kumasi
Tel: 03220 63248 or 020 5453785

CONSENT FORM

Statement of person obtaining informed consent:

I have fully explained this research to the undersigned participant and have given sufficient information about the study, including that on procedures, risks and benefits, to enable the prospective participant make an informed decision to or not to participate.

DATE: _____ NAME: **ALI BABA ISSAH (Co-Researcher)**

Statement of person giving consent:

I have read the information on this study/research or have had it translated into a language I understand. I have also talked it over with the interviewer to my satisfaction.

I understand that my participation is voluntary (not compulsory).

I know enough about the purpose, methods, risks and benefits of the research study to decide that I want to take part in it.

I understand that I may freely stop being part of this study at any time without having to explain myself.

I have received a copy of this information leaflet and consent form to keep for myself.

PATIENT ID: _____

DATE: _____ SIGNATURE/THUMB PRINT: _____

Statement of person witnessing consent (Process for Non-Literate Participants):

I _____ (Name of Witness) certify that information given to

_____ (Patient ID of Participant), in the local language, is a true reflection of what l have read from the study Participant Information Leaflet, attached.

WITNESS' SIGNATURE (maintain if participant is non-literate): _____

MOTHER'S SIGNATURE (maintain if participant is under 18 years): _____

MOTHER'S NAME: _____

FATHER'S SIGNATURE (maintain if participant is under 18 years): _____

FATHER'S NAME: _____

APPENDIX 2

Data Collecting Sheet

This form should be filled out by the Research Assistant for each Patient enrolled in the Nevirapine SJS/TEN study.

Date: [ ] [ ]/ [ ] [ ]/ [ ] [ ][ ] [ ] Patient ID: [ ] [ ][ ] [ ]

Section A: Please record the demographic data of the patient

1. Please indicate the current age of patient.
   [ ] 15 yrs but less than 20yrs
   [ ] 20 yrs but less than 30 yrs
   [ ] 30 yrs but less than 40 yrs
   [ ] 40 yrs but less than 50 yrs
   [ ] more than 50 yrs

2. Please indicate the Sex of the patient.
   [ ] male
   [ ] Female
   Pregnancy status Yes [ ] No [ ]

3. Please indicate the weight of the patient:
   [ ] Kg

4. Please indicate the height of the patient
   [ ] m

5. Please indicate the ethnic group of the patient:

   ……

6. Please indicate the date of Initial HIV diagnosis
   [ ] [ ]/ [ ] [ ]/ [ ] [ ][ ] [ ]

Section B: Please record Physical Evaluation of the Patient

7. Please mark one answer for each question in the table below.

Statement	Week 0	Week 2	Week 4
a) Did the patient have fever on this visit?	[ ] Yes	[ ] Yes	[ ] Yes
	[ ] No	[ ] No	[ ] No

b) Did the patient have rash on this visit?	[ ] Yes	[ ] Yes	[ ] Yes
	[ ] No	[ ] No	[ ] No

Statement	Week 6	Week 8	Week 10	Week 12
c) Did the patient have fever on this visit?	[ ] Yes	[ ] Yes	[ ] Yes	[ ] Yes
	[ ] No	[ ] No	[ ] No	[ ] No
d) Did the patient have rash on this visit?	[ ] Yes	[ ] Yes	[ ] Yes	[ ] Yes
	[ ] No	[ ] No	[ ] No	[ ] No

Section C: Please record Clinical Evaluation of the patient

8. Confirmatory HIV status:

   [ ] Type 1

   [ ] Type 2

Test	Day 0	Week 12
CD4 count		
Viral Load		

9. What combination of antiretroviral drugs was patient given?

   1……

   2……

   3……

   4……

10. Hepatitis B status of patient
    [ ] Positive
    [ ] Negative

11. (For positive SJS/TEN cases only)
    a. How long was patient on Nevirapine before onset of SJS/TEN symptoms?

       …… days

    b. What measures were taken after SJS/TEN onset? (E.g. whether there was withdrawal of any of the drugs and what changes were made to the combination.)

       ……

       ……

       ……

       ……

       ……

    c. What is the current SJS/TEN status of patient? How well has patient recovered? Please tick.

       Not at all [ ] Fairly Good [ ]

       Good [ ] Very Well [ ]

12. (For negative SJS/TEN cases only)
    a. How long has patient been on Nevirapine?

       …… Days

    b. Briefly state what complications (if any) patient has developed since being on Nevirapine.

       ……

       ……

       ……

       ……

       ……

13. What other drugs is patient taking besides antiretroviral therapy? Please list.

1. ……

2. ……

3. ……

4. ……

5. ……

APPENDIX 3

Socio-demographic and clinical characteristics of patients enrolled in study.

Patient ID	Age	Sex	Pregnancy	Ethnic group	Weight	Height	BMI	CD4 cell count	Rash
1	30-40	Female	No	Asante	64	1.6	25.00	774	No
2	40-50	Female	No	Fafra	68	1.5	30.22	408	No
3	30-40	Female	Yes	Asante	49	1.59	19.38	500	No
4	20-30	Female	Yes	Asante	55	1.5	24.44	755	No
5	30-40	Female	No	Asante	54	1.56	22.19	206	No
6	30-40	Female	Yes	Asante	71	1.57	28.80	167	No
7	>50	Female	No	Asante	51	1.4	26.02	N/A	No
8	>50	Male	--	Asante	73	1.63	27.48	313	No
9	>50	Male	--	Asante	49	1.5	21.78	185	No
10	30-40	Female	No	Asante	51	1.53	21.79	N/A	No
11	30-40	Female	No	Asante	74	1.68	26.22	380	No
12	>50	Female	No	Asante	79	1.59	31.25	445	No
13	40-50	Female	No	Basare	62	1.55	25.81	149	Yes
14	30-40	Female	No	Ewe	66	1.53	28.19	375	Yes
15	40-50	Male	--	Asante	61	1.7	21.11	N/A	No
16	40-50	Female	No	Asante	58	1.45	27.59	N/A	No
17	20-30	Female	Yes	Asante	62	1.55	25.81	N/A	Yes
18	20-30	Female	Yes	Asante	69	1.63	25.97	431	No
19	30-40	Female	Yes	Asante	40	1.58	16.02	350	No
20	30-40	Female	No	Asante	54	1.61	20.83	185	No
21	20-30	Female	Yes	Krobo	48	1.45	22.83	249	No
22	>50	Female	No	Asante	50	1.6	19.53	361	Yes
23	>50	Male	--	Asante	70	1.77	22.34	194	No
24	40-50	Female	No	Asante	62	1.6	24.22	29	No
25	30-40	Female	Yes	Kotokoli	57	1.54	24.03	335	No
26	20-30	Female	No	Asante	50	1.49	22.52	124	No
27	>50	Female	No	Asante	51	1.43	24.94	169	No
28	30-40	Female	No	Asante	50	1.4	25.51	183	No
29	30-40	Male	--	Fafra	56	1.45	26.63	206	No
30	>50	Female	Yes	Gruma	75	1.6	29.30	621	Yes
31	>50	Female	No	Asante	57	1.56	23.42	177	Yes
32	30-40	Female	Yes	Asante	63	1.7	21.80	427	No
33	40-50	Female	No	Asante	64	1.57	25.96	N/A	No
34	>50	Female	Yes	Asante	52	1.47	24.06	189	No
35	30-40	Female	Yes	Kussasi	50	1.4	25.51	1022	No
36	40-50	Female	No	Asante	53	1.56	21.78	350	No
37	>50	Female	No	Asante	55	1.5	24.44	154	No
38	30-40	Female	Yes	Gruma	54	1.48	24.65	N/A	Yes

39	30-40	Female	No	Dagomba	58	1.5	25.78	122	No
40	30-40	Female	Yes	Asante	51	1.56	20.96	384	No
41	40-50	Female	No	Ewe	66	1.52	28.57	325	No
42	>50	Male	--	Asante	59	1.63	22.21	350	No
43	40-50	Female	No	Asante	60	1.54	25.30	163	No
44	30-40	Female	Yes	Asante	54	1.58	21.63	129	No
45	30-40	Female	Yes	Asante	53	1.6	20.70	N/A	No
46	20-30	Female	Yes	Busanga	54	1.59	21.36	N/A	No
47	30-40	Female	Yes	Asante	64	1.63	24.09	389	No
48	30-40	Male	--	Asante	73	1.78	23.04	108	No
49	>50	Male	--	Asante	48	1.63	18.07	120	No
51	>50	Male	--	Asante	58.5	1.64	21.75	350	No
52	40-50	Female	Yes	Gruma	49	1.66	17.78	N/A	Yes
53	40-50	Female	No	Nzema	61	1.62	23.24	493	No
54	40-50	Female	Yes	Asante	51	1.51	22.37	N/A	No
55	30-40	Male	--	Fafra	60	1.75	19.59	N/A	No
57	>50	Female	No	Asante	47	1.6	19.07	N/A	Yes
58	30-40	Female	No	Ewe	54	1.6	21.09	N/A	Yes

N/A, not available

APPENDIX 4

DNA sequences from HLA-A locus after chromatogram trimming. Trimming was done by taking out 50 bases from the 5' end and limiting the sequence to 260 bp.

	HLA-A	
#	**Exon 2**	**Exon 3**
1	GCGGGGAGCCCCGCTTCATCGCMGTGGGCTACG TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA CGCCGCGAGCCAKAGGATGGAGCCGCGGGCGCC GTGGATAGAGCAGGAGGGGCCGGAGTATTGGGA CGAGGAGACAGGGAAAGTGAAGGCCCACTCACA GACTGACCGAGAGAACCTGSGGATCSYGCTCCGC TACTACAACCAGAGCGAGGCCGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCCC	TTKGAAYMTACCGCCAGCTCCGCCTCCCGGAAA CCTGSGGGTYSG
2	CGCGGGGAGCCCCGCTTCATCGCCGTGGGCTACG TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA CGCCGCGAGCCAGAAGATGGAGCCGCGGGCGCC GTGGATRGAGCAGGAGGGGCCGGAGTATTGGGA CCAGGAGACACGGAATATGAAGGCCCACTCACAG ACTGACCGAGCGAACCTGGGGACCCTGCGCGGCT ACTACAACCAGAGCGAGGACGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCC	GGGCGCTTCCTCCGCGSCSTWYYTGGCAGGACG CCTACGACGGCAAGGATTACATCGCCCTGAACG AGGACCTGCGCTCTTGGACCGCGGCGGACATGG CAGCTCAGATCACCAAGCGCAAGTGGGAGGCGG YCCRTGCGGCGGAGCAGCGGAGAGYCTACCTGG AGGGCCGGTGCGTGGACGGGCTCCGCAGATACC TGGAGAACGGGAAGGAGACGCTGCAGCGCACG GGTACCAGGGGCCACGGGGCSCCTCCCTGA
4	CGCGGGGAGCCCCGCTTCATCGCCGTGGGCTACG TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA CGCCGCGAGCCAKAAGATGGAGCCGCGGGCGCC GTGGATRGAGCAGGAGGGGCCGGAGTATTGGGA CCAGGAGACACGGAATATGAAGGCCCACTCACAG ACTGACCGAGYGAACCTGGGGACCCTGCGCGGCT ACTACAACCAGAGCGAGGACGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCC	GCTTCCTCCGCGGGCSCGGTGGACGCTTACGAC GGWAGGATTACTCGCCCTGAACGAGGACCTGC GCTCTTGGACCGCGGGGACATGGCGGCTCAKAT CACCCAGCGCAAGTGGGAGACGGYCCATGAGGC GGAGCAGTKGAGAGCCTACCTGGAGGGCCCGTG CGTGGAGYGGCTCCGCASATACCTGGAGAACGG RAAGGAMACGCTGCAGCGCACGG
6	CGGGGAGCCCCGCTTCATCGCMGTGGGCTACGT GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC GCCGMGAGCCAGAGGATGGAGCCGCGGGCGCC GTGGATRGAGCAGGAGGGKCCGGAGTATTGGGA CSGGRASACACGGAAWGTGAAGGCCCASTCACA GACTSACCGAGTGGACCTGGGGACCCTGCGCGGC TACTACAACCAGAGCGAGGCCGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCCTC	GCSCTTCCTCCGCGCCCTCCATAAGKACGCCTACS ACGGCAAGGMTTACATCGCCCTGAMMSAGGAC CTGSGCTCTTGGACCGCGRSGGACATGKYARCTC AKACCACCARGCMCAAGTGGGACGCGKCCCATG TGGMGGAGCASWGGASAGCCTACCTGGAGGG MRCGGGSGWGGAGWGGCTCCCCCSATACCMG SAKAACGGGAAGGAGACGCTGCAGCGCACGGG TACCAGGGGCCACGGGKCGCCTCCCTGATCGC
11	GGGGGGCGGGTGGTTGKTCTGTGWMCGACYCM TCATACMCT	CGGGCGCTTCCTCCGCGCCTATCGGSAGGACGCC TACGACGGCAGGATTACATCGCCCTGAACGAGG ACCTGCGCTCTTGGACCGCGGCGGACATGGCGG

		CTCAGATCACCAAGCGCAAGTGGGAGGCGGCCC ATGAGGCGGAGCARTTGAGASCCTACCTGGATG GCACGTGCGTGGAGTGGCTCCGCAGATACCTGG ARAACGGGAAGGAGACGCTGCAGCGCACGGGT ACCAGGGGCCACGGGGCGCCTCCCTGATC
12	CGGGGAGCCCCGCTTCTCGCAGAGGGCTACGTGT MCRACACGCATTTCGGGCGGTTCGACAGCKAAGC CTCGAGCCGTAKGATGGASCCGCGKGCGCCGTGR ATKGAGCAGGAGGGWCCGGAGTATTGGGACGG GMAGACACGGAAWGTGAAGGYCCAYTCACAGAT TGAYCGAGTGGACCTGKGGACCCTGCGAGAYTAC TACAACCAGAGCGAGGMCGGTGAGTGACCCCGG MCCGGRGCGCAGGTCACGACCCCCTC	TKRWGGMTTCWWGCATCCTCCGCCTCTAGGGC TGCTKAWGGGTC
13	MGYGGAGAGMCCCGCTTCATCGCMGTGGGCTA CGTGGACGACACGCAGTTCGTGCGGTTCGACAGC GACGCCGCGAGCCAKAGGATGGAGCCGCGGGCG CCGTGGATAGAGCAGGAGRGGCCKGAGTATTGG GACSAGGAGACASGGAAWGTGAAGGCCCACTCA CAGACTGACCGAGAGAACCTGSGGAYCSYGCKCS GCTACTACAACCAGAGCGAGGCCGGTGAGTGACC CCGSCCSGGGGCGCAGGTCACGACCCC	WCCTCCCCTGTAAAATTAACAGKCCACCACGGW GGSCTTATGTGGCCGTGTCCCAGGMCCTGKGCC TACGGACCGCGRGGGACATGTTCAATCATCTCCC CGGGCGAGWGTCGSTCTCCTCCCCATCTCAMYT GAACAGCCYTCCCTACCACTCCTGMRCGGAGGT GGASWMTGCKCATCCCTGATCGSCTAACGGGAA GGAGACGCTGCAGSGCACGGGKACCAGGGGCC ACGGKKCGCCTCCCTGATCGCCTAA
14	GCGGGGAGCCCCGCTTCATCGCMGTGGGCTACG TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA CGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCC GTGGATAGAGCAGGAGGGKCCGGAGTATTGGGA CSGGRASACACGGAAWGTGAAGGCCCACTCACA GACTSACCGAGTGGACCTGGGGACCCTGCGCGGC TACTACAACCAGAGCGAGGMCGGTGAGTGACCC CGGCCCGGGGCGCAGGTYACGACCCCC	GCTWCCTCCTGGTTACGGTAARACTCTTACKACG GAATCTTACRTCGCCCTGAMCCAGGACCTGCGC TCTTGGACCGCGRSGGACATGGYSKATCMKATCA CCSRGCGMGAGYGGGAGACGTCCCATGAGGCG GAGAAYATGAGAGCCTACCTGGAGGGCACGGGS GTGGAGTGGCTCCCYCCATRCCTGSAGAACGGG AAGGAGACGCTGCAGCGCACGGGTACCAGGGG CCACGGGGCGCCTCCCTGATCRMCTAA
17	GCGGGGAGCCCCGCTTCATCGCMGTGGGCTACG TGGACGACWCGCAGTTCGTGCRGTTCGACAGCG ACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGC CGTGGATAGAGCAGGAGRGCCGGAGTATTGGGA CSAGGAGACACGGAATGTGAAGGCCCASTCACAG ACTGACCGAGTGRACCTGGGGACCCTGCGCGGCT ACTACAACCAGAGCGAGGMCGGTGAGTGACCCC GGCCSGGGGCGCAGGTYACGACCCCT	CGCTACCTCCGTGTTTTTGTGTGATCGTTTACTAC GTCAGTATTACTCGCCCTGAACGAGGACCTGCGC TCYTGGACCGGRCGGACATGGCGGATCAGATCA CCAAGCGCAAGTGGGAGGCGGCCCGTGRGGCG GAGCAGYTGARAGCCTACCTGGAKGGCRMGKG CGTGGAGTGGCTCCGCASATACCTGGAGAACGG GAAGGAGACGCTGCAGCGCACGGGTACCAGGG GCCACGGGSCGCCTCCCTGATCRCCTAA
19	GCGGGGAGCCCCGCTTCATCGCAGTGGGCTACGT GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC GCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCG TGGATAGAGCAGGAGTCCGGAGTATTGGGACGG GGAGACACGGAAAGTGAAGGCCCACTCACAGAC	ATTGGGCTTCCTCCGCGGTTAATTTAAMAGKCCT ACTCGGCAAGGCTTACRTSGCTSTGAMCSAGGAC CTGSGCTCWYGGACCGCSRSGGACATGGYCAAA CAKAWCACCAGGRRCGTGYGGGAGGCGTCCCA ATCTGGCGTGAAATGATATCTACCTGGAGGAGG

	TCACCGAGTGGACCTGGGGACCCTGCGCGGCTAC TACAACCAGAGCGAGGCCGGTGAGTGACCCCGG CCCGGGGCGCAGGTYACGACCCCC	GGAGCGTGGAGTGGCCCCTCCCTGCCTGCCWAA CGGRAAGGAGACGCTGCAGCGCACGGGTACCA GGGGCCACGGGGCGCCTCCCTGATMGCCT
20	CGCGGGGAGCCCCGCTTCATCGCMGTGGGCTAC GTGGACGACACGCAGTTCGTGCGGTTCGACAGCG AMGCCGCGAGCCRGAGGATGGAGCCGCGGGCG CCGTGGATAGAGCAGGAGGGKCCGGAGTATTGG GACSRGGAGACACGGAAWGTGAAGGCCCACTCA CAGACTSACCGAGTGGACCTGGGGACCCTGCGCG GCTACTACAACCAGAGCGAGGCCGGTGAGTGACC CCGGCCSGGGGCGCAGGTYACGACCCC	GGCGCTTCCTCCGCGCGCATSGTAAGGACGCCTA CGACGGCAAGGATTACATCGCCCTGAACGAGGA CCTGCGCTCTTGGACCGCGGCGGACATGGCGG MTCAGATCACCAAGCGCAAGTGGGAGGCGGCC CATGAGGCGGAGCAGTTGAGAGCCTACCTGGAK GGCACGTGCGTGGAGTGGCTCCGCMSATACCTG GAGAACGGGAAGGAGACGCTGCAGCGCACGGG TACCAGGGGCCACGGGGCGCCTCCCTGATMG
21	GCGGGGAGCCCCGCTTCATCGCAGTGGGCTACGT GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC GCCGMGAGCCGGAGGATGGAGCCGCGGGCGCC GTGGATRGAGCAGGAGGGKCCGGAGTATTGGGA CSGGGAGACACGGAAAGTGAAGGCCCACTCACA GACTCACCGAGTGRACCTGGGGACCCTGCGCGGC TACTACAACCAGAGCGAGGCCGGTGAGTGACCCC GGCCCGGGGCGCAGGTYACGACCCCT	TGAASAGGGSTAGAGTAGCTCCGCCTCCCTGAAA CCTATC
22	GCGGGGAGCCCCGCTTCATCGCMGTGGGCTACG TGKACGACACGCAGTTCGTGCGGTTCGACMGSGA CGYCGYGAGCCAKASGATGGAGCCGCGGGYGCCG TGGATAGAKCASGAGGGSCCGGAGTATTGGGACG AGGAGACWGGSAAAGTGAAGGCCCACTCACAGA CTGACCGAGAGAACCTGCGGATCGCGCTCCGCTA CTACAACCARAGCGAGGCCGGTGAGTGACCCCGG CCCGGGGCGCAGGTCACGACCCCT	TKTSAAAATTKTGTCACCTACGCCTCCCTGAAGCC TKAAGKSTYC
25	GCRGGGAGCCCCGCTTCATCGCCGTGGGCTACGT GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC GCCGCGAGCCAKAGGATGGAGCCGCGGGCGCCG TGGATAGAGCAGGAGGGGCCGGAGTATTGGGAC GAGGAGACAGGGAAAGTGAAGGCCCACTCACAG ACTGACCGAGAGAACCTGCGGATCGCGCTCCGCT ACTACAACCAGAGCGAGGCCGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCCC	GGSGCCTCCTCCSCTGCTAAATTAAGACTCCAACT CGGYWAGTSTTACGTCGCCTTGAMCCAGGACCT GYGCTCTCGGACCGCCKSGGACATGGYCWAACC TATCMCCCGGSGAGWGYGGGAGGCGGCCCGTG TGGSGKAACAATTGATAGCCTACCTGGAGGGCA CGTGCGKGGASKGGCTCCTCCCAGCCTGCCTAAA CGGRAAGGAGACGCTGCAGCGCACGGGTACCA GGGGCCACGGKGCGCCTCCCTGATCGCC
26	GCRGGGAGCCCCGCTTCATCGCAGTGGGCTACGT GKWCGACACGCAGTTCGTGCGGTTCGACAGSGA CGCCGYGAGCCAKAGGATGGAGCCGCGGGCGCC GTGGATAGAGCAGGAGGGYCCGGAGTATTGGGA CGGGRAGACWCGGAAWGTGAAGGCCCACTCAC AGACTSACCGAGTGSACCTGGGGACCCTGCGCGG	NNNNN

	CTACTACAACCAGAKMGAGGMCGGWGAGTGAC CCCGGSCCGGRGSGSWGGTYRCGACCCCT	
27	YGGRGAGCCCCGCTTCATCGCMGTGGGCTACGTG GACGACACGCAGTTCGTGCGGTTCGACAGCGACG CCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGT GGATRGAGCAGGAGRGGCCKGAGTATTGGGACC RGAASACACGGAWTGTSAAGGCCCASKCACAGAC TGACCGAGWGRACCTGGGGACCCTGCGCGGCTA CTACAACCAGAGCGAGGCCGGTGAGTGACCCCG GCCSGGGGCGCAGGTCACGACCCCTC	CGTTTATGAASTCTGTAGCTTCCGCCTMCCTGAT CGMCTAAGTG
28	TGGMATCYCGGCACTTTTCCAGGCTGAGGGGTAC GTGTACCACGATGTTGGCAATCCGTTCGCCAGCG ACGCCTAGAGMCWGAGGATGGAKCCGCGGGCG CCGTGGATRGAGCAGGAGAGGCCTGAKTATTGG GACCAGGAGACMYAGAATGTGAAGGCCCASTCA CAGACTGACCGAGWGRACCTGGGGACCCTGCGC GGCTACTACAACCAGAGCGAGGCCGGTGAGTGA CCCCGGCCGGGGGCGCAGGTCACGACCC	GGGCGTTTCCTCCGYGGCTTGAACAGCACKCCTA CCACGGCAAGGCTTACRTCGCCCTGAACSAGGAC CTGCGCTCTTGGACCGCSRSGGACATGGCGGATC ATATCACCCRGCGCAAGYGGGAGGCGGCCCGTC GGRCGGAGCARTTGATAGCCTACCTGGAGGGCA CGTGCGTGGAGTGGCTCCKCAGATACCTGGAAA ACGGGAAGGAGACGCTGCAGCGCACGGGTACC AGGGGCCACGGGGCGCCTCCCTGATCG
30	CGGGGAGCCCCGCTTCATCGCCGTGGGCTACGTG GACGACACGCAGTTCGTGCGGTTCGACAGCGACG CCGCGAGCCAKAGGATGGAGCCGCGGGCGCCGT GGATAGAGCAGGAGGGGCCGGAGTATTGGGACG AGGAGACAGGGAAAGTGAAGGCCCACTCACAGA CTGACCGAGAGAACCTGCGGATCGCGCTCCGCTA CTACAACCAGAGCGAGGCCGGTGAGTGACCCCG GCCCGGGGCGCAGGTCACGACCCCCT	TTTGGTAAAGYCCTGTAAGCCTTCCGTCTCCCTGT TCGCYTAAATTATT
31	GCGGGGAGCCCCGCTTCMTCKTAKTGGGCTACGT GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC GCCGCGAGCCGGAGGATGGAGCCGCGGGCGCCG TGGATAGAGCAGGAGGGTCCGGAGTATTGGGAC GGGGAGACACGGAAAGTGAAGGCCCACTCACAG ACTCACCGAGTGGACCTGGGGACCCTGCGCGGCT ACTACAACCARAGCGAGGCCGGTGAGTGACCCCG GCCCGGGGCGCAGGTCACGACCCCT	GAKTTRAACTCCTGYACGCTCCGCCTCCCTGATM GCMTAACKTTAGA
34	CGCRGGGAGCCCCGCTTCAYCGCYKTGGGCTACG TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA CGCCGCGAGCCAKAGGATGGAGCCGCGGGCGCC GTGGATAGAGCAGGAGGGGCCGGAGTATTGGGA CGAGGAGACAGGGAAAGTGAAGGCCCACTCACA GACTGACCGAGAGAACCTGCGGATCGCGCTCCGC TACTACAACCAGAGCGAGGCCGGTGAGTGACCCC GGCCCGGGGCGCAGGTCACGACCCC	GGGCGCTTCCTCCGCGGCTATSTTAGSACGCCTA CKACGGAAGGATTACRTCGCTYTGAMCSAGGAC CTGCGCTCYTGGACCGCGGSGGACATGGCRGAT CAKATCACCCAGCGCAAGTGGGAGGCGGCCCGT GKGGCGGARCARYTGATAKCCTACCTGGAGGGC RAGTKCGTGGAGTGGCTCCKCAGATACCTGGAA AACGGRAAGGAGACGCTGCAGCGCRCGGGTACC AGGGGCCACRGGGCGCCTCCCTGATCRC

37

GCGGGGAGCCCCGCTTCATCGCCGTGGGCTACGT
GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC
GCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCG
TGGATAGAGCAGGAGGGGCCGGAGTATTGGGAC
GAGGAGACAGGGAAAGTGAAGGCCCACTCACAG
ACTGACCGAGAGAACCTGCGGATCGCGCTCCGCT
ACTACAACCAGAGCGAGGCCGGTGAGTGACCCC
GGCCCGGGGCGCAGGTCACGACCCCC

TGGCGCTTCCTYGYCGGTTTGTTTGSAGTCCTACT
CGGCAAGGCTTACGTCGCTGTGACCCAGGACCT
GSGCTCATGGACCGCCRSGGACATGGCCWAACC
TCTCACAGGGMCWGWGYGGGAGGCTTCCCGT
GTGGSGKAACAAYTGATATCCTACCTGGAGGGS
GWGACGKGGARWGGCTCCTCCCTGCCTGCCTA
AACGGRAAGGAGACGCTGCAGSGCACGGGTACC
RGGGGCCACRGGGCGCCTCCCTGATCGCCT

38

CGCGGRGAGCCCCGCTTCATCGCCGTGGGCTACG
TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA
CGCCGCGAGCCAKAGGATGGAGCCGCGGGCGCC
GTGGATRGAGCAGGAGGRGCCGGAGTATTGGGA
CCGGAASACACRGAWTGYSAAGGCCCASKCACAG
ACTGACCGAGTGRACCTGGGGACCCTGCKCGGYT
ACTACAACCAGAGCGAGGMCGGWGAGTGACCC
CGGCCCGGGGCGCAGGTCACGACCCC

CCTTTCCGTTTCGTCTTTTTCCCCCCCTATTTTTTTT
YYAMMSK

39

AGTGGAGAGCCCCGCTTCATCGCAGTGGGCTACG
TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA
CGCCGCGAGCCAGARGATGGAGCCGCGGGCGCC
GTGGATRGAGCAGGAGRGGCCKGAGTATTGGGA
CCAGGAGACACGGAATRTGAAGGCCCASTCACAG
ACTGACCGAGYGRACCTGGGGACCCTGCGCGGCT
ACTACAACCAGAGCGAGGMCGGTGAGTGACCCC
GGCCSGGGGCGCAGGTCACGACCCC

GWKTSGGMSCCTTCCAGGCTGAGTGTCCATGGT
TGYTKGTGKGT

41

GCGGGGAGYCCCGCTTCATCGCMGTGGGCTACG
TGGACGACACGCAGTTCGTGCGGTTCGACAGCGA
CGCCGYGAGCCAGAGGATGGAGCCGCGGGCGCC
GTGGATAGAGCAGGAGGGKCCGGAGTATTGGGA
CGGGRAGACACGGAATGTGAAGGCCCACTCMCA
GACTGACCGAGTGGACCTGGGGACCCTGCGCGG
CTACTACAACCAGAGCGAGGCCGGTGAGTGACCC
CGGSCCGGRGCGGAGGTWACMACCTGY

CTTTCTGAGCTCCTTCATAAAACTCCWACCACGG
CARGGCTTACATCGCCCTGAMCCAGGACCTGSG
CTCTCGGACCGCCASGGACATGKYCRATCMTAM
CACMRGGMMCAAGYGGRACGCGKCCCCTGYGS
MGGAGAASAGSARASCCTACCAGGAGGGMRCG
GGSGWGGAGTGGCTCCCCCCATGCCTGCCTAAC
GGRAAGGAGACGCTGCAGCGCACGGGTACCAG
GGGCCACGGGGCGCCTCCCTGATCGCCTAA

43

GTGGAGAGCCCCGCTTCATCGCAGTGGGCTACGT
GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC
GCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCG
TGGATRGAGCAGGAGAGGCCTGAGTATTGGGAC
CAGGAGACACGGAWTGTGAAGGCCCASTCACAG
ACTGACCGAGWGRACCTGGGGACCCTGCGCGGC
TACTACAACCAGAGCGAGGCCGGTGAGTGACCCC
GGCCGGGGGCGCAGGTCACGACCCCT

GATRWGAGCATTGAACCTKTGCCTGTTGAAACK
TKKAWCTY

44 GCGGGGAGCCCCGCTTCATCKCCGTGGGCTACGT GGTCGGGCGCTTCCTCCGCGGCTATGTTAGSACG
GGACGACACGCAGTTCGTGCGGTTCGACAGCGAC CCAACGACGGCRGGATTACTCGCCCTGAACGAG
GCCGCGAGYCMGAGRATGGAGCCGCGGGCGCC GACCTGCGCTCWYGGACCGCGGSGGACATGGC
GTGGATRGAGCRGGAGGGGCCGGAGTATTGGGA GGMTCAKATCACCARGCGCAAGTGGGAGGCGG
CCRGGAGACACGGAWTGTGAAGGCCCAGKCACA CCCRTGAGGCGGARCARWTGAKAGCCTACCTGG
GACTGACCGAGWGRACCTGSGGACCCTGCKCGG AKGGCAMGTGCGTGGAGTGGCTCCGCASATACC
CTACTACAACCAGAGCGAGGCCGGTGAGTGACCC TGGAGAACGGRAAGGAGACGCTGCAGCGCACG
CGGCCGGGGGCGCAGGTCACGACCCCT GKTACCAGGGGCCACGGGGCGCCTCCCTGAT

52 GCGGRGAGCCCCGCTTCATCGCAGTGGGCTACGT NNNNN
GGACGACRCGCAGTTCGTGCGGTTCGACAGCGA
MGCCKMGAGCCGGAGGATGGAGCCGCGGGCGC
CGTGGATRGAGCAGGAGRGKCCKGAGTATTGGG
ACSGGGAGACACGGAAAGTGAAGGCCCACTCAC
AGACTSACCGAGWGRACCTGGGGACCCTGCGCG
GCTACTACAACCAGAGCGAGGCCGGTGAGTGACC
CCGSCCSGGGGCGCAGGTCACGACCCCC

53 MGYGGAGAGCCCCGCTTCATCKCMGTGGGCTAC ACGGGCGCTTCCTCCGCGSGTATSGGMAGGACG
GTGGACGACACGCAGTTCGTGCGGTTCGACAGCG CCTACGACGGCAGGATTACATCGCCCTGAACGA
ACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGC GGACCTGCGCTCTTGGACCGCGGCGGACATGGC
CGTGGATRGAGCAGGAGRGGCCTGAGTATTGGG GGMTCAGATCACCAAGCGCAAGTGGGAGGCGG
ACCAGGAGACACGGAATGTGAAGGCCCAGTCAC CCCRTGAGGCGGAGCARTTGAKAGCCTACCTGG
AGACTGACCGAGTGGACCTGGGGACCCTGCGCG ATGGCACGTGCGKGGAGTGGCTCCGCASATACC
GCTACTACAACCAGAGCGAGGCCGGTGAGTGACC TGGAGAACGGRAAGGAGACGCTGCAGSGCACG
CCGGCCGGGGGCGCAGGTCACGACCCC GGTACCAGGGGCCACGGGGCGCCTCCCTGAW

57 CMGYRGRGAGCCCCGCTTCATCKCMRTGGGCTAC ATTACAGCTCATTCTCTATTTARCTTTCTTTTCTGA
GTGGACGACACGCAGTTCGTGCGGTTCGACAGCG CTTTTTTT
ACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGC
CGTGGATAGAKCAGGAGRGGCCKGAGTATTGGG
ACSAGGAGACASGGAAWGTGAAGGCCCACTCAC
AGACTGACCGAGAGAACCTGSGGAYCSYGCKCSG
CTACTACAACCAGAGCGAGGCCGGTGAGTGACCC
CGSCCSGGGGCGCAGGTCACGACCC

58 GTGGAGAGCCCCGCTTCATCGCAGTGGGCTACGT GCGCTTCCTTGCTGCTAATTTAGMACTCCTACAC
GGACGACACGCAGTTCGTGCGGTCGACAGSGACS GGAGGGATTATATCGCCCTGTMCCAGGACCTGS
CCSCGAGCCAKASGATGGASCCGCGGSCGCCGTG GCTCTCGGACCGCCRSGGACATGGYCRATCCTAT
GATGGAGCASGASAGGSCTGASTATTGGSACCAG CCCCCGGCGCGWGTGGGAGGCGGCCCGTGGGG
GAGACWGGSAATGWGAAGGCCCACTCMCAKAC SGGGAAATTGATATCCTACCTGGATGGSACGGAY
TGACCGASTGKACCTGGSSACCCTGCGCGRCTACT GTGGTRWGGCTCCTCCCTGACTGSCTAAACGGG
ACAACCAGAKMGAKGMCGGAGAGTGACCCCGG AAGGAGACGCTGCAGCGYACGSGTACCAGGGGC
CCGSGAGCGCAGGTCACGACCTCTC CACGGGGCGCCTCCCTGATCGCCTAAC

APPENDIX 5

DNA sequences of HLA-B after chromatogram trimming. Trimming was done by taking out 50 bases from the 5' end and limiting the sequence to 330 bp.

	HLA-B	
#	**Exon 2**	**Exon 3**
1	CTCGCCCCCAGGCTCCCACTCCWTGAGGTATTTC TACACCGCCATGTCCCGGCCCGGCCGCGGGGAG CCCCGCTTCATCGCAGTGGGCTACGTGGACGAC ACCCAGTTCGTGAGGTTCGACAGCGACGCCGCG AGTCCGAGGATGGCGCCCCGGGCGCCATGGATA GAGCAGGAGGGGCCGGAGTATTGGGACGGGGA GACACGGAACATGAAGGCCTCCGCGCAGACTTA CCGAGAGAACCTGCGGATCGCGCTCCGCTACTA CAACCAGAGCGAGGCCGGTGAGKGACCCCGGCC CGGGGCGCAGGTCACGACTCCCCATCCCCCACG	RAAGGAGGTTTGGTTGCAACCGGGGGCCGGASG GSSGCCTCCTCCGGGGKWTTAACCRGTYCGCCTAC AACGGCAAGGTTTACTTCSCCCTGAACGAGGACCT GSGTTCCGGGACCGCCGSGRACACGGSGGTTCAA ATMMCCCASSGMARGTGGAAGGSGGCCCGGGG GGSGRASMASYGRAAASCCTACCTGAAGGGCWG GGGSGGGAAGTGSTTCCGMAAATACCTGAAAAA CGGAAAGGAAACGTTGMASSGCGSGGKWACCAG GGGMRGGGGGAASCCTTCCCCATTTCCWTTARAT CTCCCGGAATGSCCTCCCMCGARAAAAG
2	CCTCGCCCCCAGGCTCCCACTCCWTGAGGTATTT CCACACCKCCRTGTCCCGGCCCGGCCGCGGGGA GCCCCGCTTCATCACCGTGGGCTACGTGGACGAC ACGYTGTTCGTGAGGTTCGACAGCGACGCCRCG AGTCCGAGRRAGGAGCCGCGGGCGCCRTGGATA GAGCAGGAGGGGCCGGAGTATTGGGACCGGGA GACACAGATCTSCAAGRCCAASRCACAGACTKAC CGAGAGRRCCTGCGGAMCCTGCKCSGCTACTAC AACCAGAGCGAGGCCGGTGAGTGACCCCGGCCC GGGGCGCAGGTCACGACTCCCCATCCCCCAC	CCACTTTTTGGTAGTTCCTGYTSCCACGAGTKSAG GAGGAWGG
4	TCCTCGCCCCCAGGCTCCCACTCCWTGAGGTATT TCTACACCGCCATGTCCCGGCCCGGCCGCGGGG AGCCCCGCTTCATCKSAGTGGGCTACGTGGACGA CACCCAGTTCGTGAGGTTCGACAGCGACGCCGC GAGTCCGAGGAMGGAGCCCCGGGCGCCRTGGA TAGAGSAGGAGGGGCCGGAGTATTGGGACCGG AACACACAGATCTTCAAGACCAACACACAGACTT ACCGAGAGAACCTGCGGATCGCGCTCCGCTACT ACAACCAGAGCGAGGCCGGKGAGTGACCCCGG CCCGGGGCGCAGGTCACGACTCCCCATCCCCCA	NNNNN
6	CCTCRCCCCCAGGCTCCCACTCCWTGAGGTATTT CTACACCKCCGTGTCCCGGCCCGGCCGCGGGGA GCCCCGCTTCATCTCAGTGGGCTACGTGGACGAC ACSCAGTTCGTGAGGTTCGACAGCGACGCCGCG AGTCCGAGAGAGGAGCCGCGGGCGCCGTGGAT AGAGCAGGAGGGGCCGGARTATTGGGACCGGA	CCAGAGSATGTAYGGCTSCGACSTGGGGCCSGAC GGGCGCCTCCTCCGCGGGYATGACCAGTACGCCT ACGACGGCAAGGATTACATCGCCCTGAACGAGGA CCTGCGCTCCTGGACCGCCGCGGACACSGCGGCT CAGATCACCCAGCGCAAGTGGGAGGCGGCCCGT GAGGCGGAGCAGCKGAGAGCCTACCTGGAGGGC

	ACACACAGATCTRCAAGRCCMASRCACAGACTG	RMGTGCGTGGAGTGGCTCCGCAGATACCTGGAG
	ACCGAGAGAGCCTGCGGAACCTGCGCGGCTACT	AACGGGAAGGACAMGCTGSAGCGCGCTGGTACC
	ACAACCAGAGCGAGGCCGGTGAGTGACCCCGGC	AGGGGCAGTGGGGAGCCTTCCCCATCTCCTATAG
	CCGGGGCGCAGGTCACGACTCCCCATCCCCCCA	RTCKCCSGGGATGGCCTCCCACGAGAAG
11	CCTYGCCCCMRGGYYCCMMCTCMTTGAGGAAT	CATCATCCAGARGATGTATGGCTGCGAMCGGGG
	TTCWACMCCGCCWTGYCCGGGCCGGGCGGSGG	GCCCGACGGGCGCCTCCTCCGCGGGCATGACCM
	GGAGCCCGGTTTCTTCSMRKGGGSTTACGGGGC	GTCCGCCTACGACGGCAAGGATTACMTCGCCCTG
	CGAMCCCCRKTTCGGGAGGTTCGACGGGAACGC	AACGAGGACCTGAGCTCCTGGACCGCGGCGGAC
	CSSAAGCCCAAGGAGGGASCCCCGGSCGCCRTG	ACCGCGGCTCARAWCACCCAGCGCAAGTGGGAR
	GATAAAKRAGAAGGGGCGGAAKTATGGGAACC	GCGGCCCGTGTGGCGGAGCAGCTGARARCCTACC
	GGAACCCMCRAATTTCCARAACCAAMACACRAA	TGGAGGGCCTGTGCGTGGAGTGGCTCCGCAGAT
	TTAACGAAAAAAMCTGGGGAATCGGGSTCGGTT	ACCTGGAGAACGGGAAGGAGACGCTGCAGCGCG
	ACAACAACCRAASSGAGGCCGGGGAKGGACCCC	CGGGTACCAGGGGCAGKGGGGAGCCTTCCCCATC
	GGCCGGGGSGGAAGGYMMSAATTCCCCTTCCCC	TCCTATAGGTCGCCGGGGATGGCCTCCCAC
	MMC	
12	CTCCTCGCCCCCAGGCTCCCACTCCATGAGGTAT	GGATGKATGGSTGCSAACTGGGGSCCGAMGGGC
	TTCTACACCKCCRTGTCCCGGCCCGGCCGCGGGG	GCCTCCTCCGSGGGATGAMCMGKCCGSCTACSA
	AGCCCCGCTTCWTCKCAGTGGGCTACGTGGACG	MGGSAAGGATTACMTYSSCCTGAACGARGACCTG
	ACACSCAGTTCGTGAGGTTCGACAGCGACGCCG	AGCTCCTGGAMCGSSGSGGACACCGSGGSTYARA
	CGAGTCCGAGRRMGGAGCCSCGGGCGCCRTGG	TYACCCMRSGSAAGKGGGARGSGGSCCGKGKGGS
	ATAGAGCAGGAGGGGCCGGAGTATTGGGACCG	GGARCARCTGAAARCCTACCTGGARGGSCTGKGS
	GAACACACAGATCTWCAAGACCAACACACAGAC	GKGGARKGGCTCCSSARATACCTGGARAACGGGA
	TKACCGAGAGARCCTGCGGAWCSYGCKCSGCTA	ARGARAMGCTKCARCGSGCGGGTACCARGGGSA
	CTACAACCAGAGCGAGGCCGGTGAGTGACCCCG	GKGGGGARCCTTCCCCMTYTYCTATARGWCGCCS
	GCCCGGGGCGCAGGTCACGACTCCCCATCCCCC	GGGATGGCCTCCCMCSARAARAAGARG
13	TCCTCACCCCCAGGCTCCCACTCCATGAGGTATTT	NNNNN
	CTACACCTCCGTGTCCCGGCCCGGCCGCGGGGA	
	GCCCCGCTTCWTCTCAGTGGGCTACGTGGACGA	
	CACCCAGTTCGTGAGGTTCGACAGCGACGCCGC	
	GAGTCCGAGAGAGGAGCCGCGGGCGCCGTGGA	
	TAGAGCAGGAGGGGCCGGAGTATTGGGACCGG	
	AACACACAGATCTACAAGGCCCAGGCACAGACT	
	GACCGAGAGAGCCTGCGGAACCTGCGCGGCTAC	
	TACAACCAGAGCGAGGCCGGTGAGTGACCCCGG	
	CCCGGGGCGCAGGTCACGACTCCCCATCCCCCA	
14	CTCCTCGCCCCCAGGCTCCCACTYCWTGAGGTAT	ACTTGGCAGACGATGTATGGCTGCGAATTGGGGC
	TTCTACACCGCCATGTCCCGGCCCGGCCGCGGG	CGGACGGGCGCCTCCTCCGCGGGCATRACCAGTA
	GAGCCCCGCTTCATYGCAGTGGGCTACGTGGAC	CGCCTACGACGGCAARGATTACMTCGCCCTGAAC
	GACACCCAGTTCGTGAGGTTCGACAGCGACGCC	GAGGACCTGAGCTCCTGGACCGCGGCGGACACC
	GCGAGTCCGAGGACGGAGCCCCGGGCGCCATG	GCGGCTCAGATCACCCAGCGCAAGTGGGAGGCG
	GATAGAGCAGGAGGGGCCGGAGTATTGGGACS	GCCCGKGAGGCGGAGCAGCTGAGAGCCTACCTG
	GGGAGACACRGAWCWYSAAGRCCWMCRCGCA	GAGGGCCTGTGCGTGGAGTGGCTCCSCAGACACC
	GACTTACCGAGAGAACCTGCGGATCGCGCTCCG	TGGAGAACGGGAAGGARACSCTGCAGCGCGCGG
	CTACTACAACCAGAGCGAGGCCGGTGAGTGACC	

	CCGGCCCGGGGCGCAGGTCACGACTCCCCATCC CCC	GTACCAGGGGCAGTGGGGAGCCTTCCCCATCTCC TATAGGTCGCCSGGGATGGCCTCCCMCS
17	CCTYGCCCCMRGSTTCCMCTTCMTGGAGKAATT YCTACCCCGCCWTGYCCGGSCCGGSCGGGGGGG AGCCCCGTTTCTTTSMRKGGGSTTACGGGAAGG ACCCCCAKTTCGGGAGGTTCGACGGGAACGCCS GAAGYCCAAGAAGGGASCCCCGGSGGCCRGGAA TAAASMAGAAGGGSCCGAAKAATGGGAACCGA AACACMCRAATCTCCAAAACCAAAACACAAACTT ACGGAAAAAGCCTGGGAAACCTGSSGGGSTTACA ACAMCCRAASGGAGGCCGKGAAKGGACCCCGSC CGGGGSGGAAGKYMMSAATTCCCCACCCCCMM C	TCCARAGGATGTTTGGYTGCGACCTGGGGCCGGA CGGGCGCCTCCTCCGCGGGTATAACCAGTTCGCC TACRACGGCAAGGATTACWTCGCCCTGAACGAG GATCTGCGCTCCTGGACCGCCGCGGACACGGSGG YTCARATCACCCAGCGCAAGKGGRAGGSGGCCCG KGAGGSGGAGCAGCGGARAGCCTACCTGGAGGG CACGKGCGKGGAGKGGYTCCGCARATACCTGGAR AACGGRAAGGARACGYTGCAGCGCGCGGGTACC AGGGGCAGKGGGGAGCCTTCCCCATCTCCCGWA RATCTCCCGGRATGGCCTCCCACGARAA
19	GKAYWATAGRTCTCGWRGMWGSKCCWCTGTG CTCCCATCCYCCACRWAA	CCRAAGGAGGTTKGGTTGCAACCTGGGGCCGGAS GGSGGCCTCCTCCGGGGGWTTAACCRGTYCSCCT ACAASGGMAAGGTTTACTTCGCCCTGAACGAGGY TTTGSGTTCCGGGACCGCCGSGAACACGGSGGTT CAAATMACCCASSGMARGGGGAAGGGGGCCCG GGAGGSGRASMASSGAAAASCCTACCTGAAGGGC YGGGGSGGGAAGGGSTTCCGMAAATACCTGAAA AACGGAAAGGAAACGTTGMASSGGGGGGKAACC AGGGGMAGGGGGAASCCTTCCCCTTCTCCWGAA RATCTCCCGGAATGSCCTCCCMCAAAARR
20	CTCGCCCCCAGGCTCCCACTCCATGAGGTATTTC CACACCGCCATGTCCCGGCCCGGCCGCGGGGAG CCCCGCTTCWTCACCGTGGGCTACGTGGACGAC RCGYTGTTCGTGAGGTTCGACAGSGACGCCACG AGTCCGAGGAAGGAGCCGCGGGCGCCATGGAT AGAGCAGGAGGGGCCGGAGTATTGGGACCGGG AGACACAGATCTCCAAGACCAACACACAGACTTA CCGAGAGAGCCTGCGGAACCTGCGCGGYTACTA CMRCCMGAGCGAGGYCGSKGAGYGACCCCGGC CCGGGGCGCAGGTCACGACTCCCCATCCCCCWC G	CCAGTGGATGTATGGCTGCRACCTGGGGCCCGAC GGGCGCCTCCTCCGCGGGTATGACCAGTCCGCCT ACGACGGCAAGGATTACATCGCCCTGAACGAGGA CCTGCGYTCCTGGACCGCCGSGGACACGGCGGYT CARATCACCCAGCGCAAGTGGGAGGCGGCCCGT GCGGSGGAGCAGCARARAGCCTACCTGGAGGGC ACGTGCGTGGAGTGGYTCCGCAGATACCTGGAGA ACGGGAAGGAGACGCTGCAGCGCGCGGGTACCA GGGGCAGKGGGGAGCCTTCCCCATCTCCTGTAGA TCTCCCGGGATGGCCTCCCACGARAAR
21	CCTCGCCCCCAGGCTCCCACTYCWTGAGGTATTT CTACACCGCCATGTCCCGGCCCGGCCGCGGGGA GCCCCGCTTCATCGCAGTGGGCTACGTGGACGA CACSCAGTTCGTGAGGTTCGACAGCGACGCCGC GAGTCCGAGGAYGGAGCCCCGGGCGCCATGGAT AGAGCAGGAGGGGCCGGAGTATTGGGACCGGA ASACACRGATCWTSAAGACCWCCSASMRACTTA CCGAGAGAACCTGCGGATCGCGCTCCGCTACTA CAACCAGAGCGAGGCCGGTGAGTGACCCCGGCC CGGGGCGCAGGTCACGACTCCCCATCCCCCACG	TCCAGAGGATGTTTGGCTGCGACCTGGGGCCGGA CGGGCGCCTCCTCCGCGGGTATRACCAGTYCGCCT ACGACGGCAAGGATTACATCGCCCTGAACGAGGA TCTGCGCTCCTGGACCGCCGCGGACACGGCGGCT CARATCACCCAGCGMAAGTGGGAGGCGGCCCGT GAGGCGGAGCAGCGRARAGCCWACYKGGAGGG CACGTGCGTGRAGKGGYTCCGCAGATACCTGGAG AACGGRAAGGARACGYTGCAGSGSGCGGGTACCA GGGGCRGKGGGGAGCCTTCCCCATCTCCCGKAGA TCTCCCGGGATGGCCTCCCACGARAA

22 TCCTCGCCCCCAGGCTCCCACTCCATGAGGTATTT NNNNN
CCACACCGCCATGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCWTCWCCGTGGGCTACGTGGACG
ACRCGYTGTTCGTGAGGTTCGACAGSGACGCCAC
GAGTCCGAGGAAGGAGCCGCGGGCGCCATGGA
TAGAGCAGGAGGGGCCGGAGTATTGGGACCGG
GAGACACAGATCTCCAAGACCAACACACAGACT
TACCGAGAGARCCTGCGGAWCSYGCTCSGCTAC
TACMRCCMGAGCGAGGYCGGKGAGTGACCCCG
GCCCGGGGCGCAGGTCACGACTCCCCATCCCCC
W

25 CTCGCYCCCAGGCTCCCACTCCWTGAGGTATTTC CACTTGGCAGAGGATGTATGGYTGCGACCTGGGG
YACACCGCYATGTCCCGGCCCGGCCGCGGGGAG CCCGACGGGCGCCTCCTCCGCGGGTATAACCAGT
CCCCGCTTCATCWCCGTGGGCTACGTGGACGAC TAGCCTACGACGGMAAGGATTACMTCGCCCTGA
ACGCTGTTCGTGAGGTTCGACAGCGACGCCRCG ACGAGGACCTGAGCTCCTGGACCGCGGCGGACA
AGTCCGAGRRAGGAGCCGCGGGCGCCRTGGATA CCGCGGMTCAGATCACCCAGCGCAAGTGGGAGG
GAGCAGGAGGGGCCGGAGTATTGGGACCGGRA CGGCCCGTGTGGCGGAGCAGGACAGAGCCTACC
SACACAGATCTMCAAGRCCMACRCACAGACTKA TGGAGGGCCTGTGCGTGGAGTCGCTCCGCAGATA
CCGAGAGAGCCTGCGGAACCTGCGCGGYTACTA CCTGGAGAACGGRAAGGAGACGCTGCAGCGCGC
CAACCAGAGCGAGGCCGGTGAGTGACCCCGGCC GGGTACCAGGGGCAGWGGGGAGCCTTCCCCATC
CGGGGCGCAGGTCACGACTCCCCATCCCCCCAC TCCTATAGGTCGCCGGGGATGGCCTCCCAC

26 TAA NNNNN

27 TCCTCGCCCCCAGGCTCCCACTCCATGAGGTATTT KCYGCCAACCCGCGGGGATRAAAAGGGGGC
CTACACCGCCATGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCATCGSAGTGGGCTACGTGGACGA
CACCCAGTTCGTGAGGTTCGACAGCGACGCCGC
GAGTCCGAGGACGGAGCCCCGGGCGCCATGGA
TAGAKSAGGAGGGGCCGGAGTATTGGGACCGG
AACACACAGATCTTCAAGACCAACACACAGACTT
ACCGAGAGAACCTGCGGATCGCGCTCCGYTACT
ACMRCCMGASCGAGGYCGSKGAGTGACCCCGG
CCCGGGGCGCAGGYCACGACTCCCCMTCCCCCW

28 TCCTCGCCCCCAGGCTCCCACTCCWTGAGGTATT GAGGATGKATGGSTGCGACCTGGGGSCCGACGG
TCTACACCGCCATGTCCCGGCCCGGCCGCGGGG GCGCCTCCTCCGSGGGYATGACCAGGWCGCCTAC
AGCCCCGCTTCATCGSAGTGGGCTACGTGGACG SACGGSAGGATTACMTCGCCCTGAACGARGAMC
ACACCCAGTTCGTGAGGTTCGACAGCGACGCCG TGAGCTCCTGGACCGSSGSGGACACCGSGGCTCA
CGAGTCCGAGGACGGAGCCCCGGGCGCCATGG RATCACCCMGCGSAAGKGGGARGSGGSCCGKGK
ATARAKSAGGAGGGGCCGGAGTATTGGGACCG GGSGGARCARCTGARARCCTACCTGGAGGGCCTG
GAACACACAGATCTTCAAGACCAACACACAGACT KGSGKGGAGKGGCTCCSCARATACCTGGARAACG
TACCGAGAGAACCKGCGGATMGCGCTCCGYTAC GGAAGGARAMGCTGCAGCGSGCGGGTACCAGG
TACMRCCMGAGCGAGGYCGGKGAGTGACYCCG GGSAGKGGGGARCCTTCCCCATCTCCYATARGTCG
GCCCGGGGCGCAGGTCACGASTCCCCATCCCCC CCSGGGATGGCCTCCCACGAAAGAGGAA
W

30 CCTCGCCCCCAGGCTCCCACTCCATGAGGTATTT NNNNN
CTACACCGCCATGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCATYGSAGTGGGCTACGTGGACGAC
ACCCAGTTCGTGAGGTTCGACAGCGACGCCGCG
AGTCCGAGGACGGAGCCCCGGGCGCCATGGATA
GAGSAGGAGGGGCCGGAGTATTGGGACCGGRA
SACACAGATCTYCAAGACCAACACACAGACTTAC
CGAGAGARCCTGCGGAWCSYGCKCSGCTACTAC
MACCAGAGCGAGGCCGGTGAGTGACCCCGGCC
CGGGGCGCAGGTCACGACTCCCCATCCCCCWC

31 CTKWCKCMCGTCTCGWSGCWRCKCCWMTRTG NNNNN
CTCCCCATCCCYCAYGTAA

34 TCCTCGCCCCCAGGCTCCCACTYCATGAGGTATTT CCKAAGAAGGTGGYTTGGATCTGGGGTCCGAAG
CTACACCGCCATGTCCCGGCCCGGCCGCGGGGA GGCGGCTTCTTCGGGGGGATTAACTGACCGGCTA
GCCCCGCTTCATCGCAGTGGGCTACGTGGACGA ACAAGGGACGGATTACCTCSCCCTGSACTAAGAAC
CACSCAGTTCGTGAGGTTCGACAGCGACGCCGC TGACCTTCTGGACCGGSCCGGAACACGCRGGTCA
GAGTCCGAGGAYGGMGCCCCGGGCGCCATGGA RAATCCCCAGCCCAAGTGGGAAGGGGGCCGGGC
TAGAGCAGGAGGGGCCGGAGTATTGGGACCGG CGSGGAGCAASTGAAAAGCTACCTTGCAGGGCTG
AASACACRGAWCWTSAAGACCTCCGACMGACT GGCGGGGSAGGGSTTCCCTCAAACCTGGACAAAG
TACCGAGAGAACCTGCGGATCGCGCTCCGCTACT GAAAGGAAAAGSTTCSSTGCGGGGSGACCMRGG
ACAACCAGAGCGAGGCCGGTGAGTGACCCCGGC CSRGGGGGGAGGCTTCCCCATTTCCTATAAGATAS
CCGGGGCGCAGGTCACGACTCCCCATCCCCCAC CCCGSCAGGSCTCCCACCACAAGAG

37 TCCTCGCCCCCAGGCTCCCACTCCATGAGGTATTT NNNNN
CTACACCGCCATGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCWTCGSAGTGGGCTACGTGGACGA
CACCCAGKTCGTGAGGYTCGACAGSGACGCCGC
GAGTCCGASGACGKAGCCCCGGKCGCCATGGAT
ARATSAGGAGGGGCCGGAGTATTGGGACCGGA
ACACACAGATCTTCAAGACCAACACACAGACTTA
CCGAGAGAACCTGCGGATCGSGCTCCGYTACTAC
MRCCMGASCGAGGYCGSGGAGYGACCCCGGCC
CGGGGCGCAGGTCACGACTCCCCATCCCCCW

38 CCTCRCCCCCAGGCTCCCACTCCATGAGGTATTTC NNNNN
TACACCKCCRTGTCCCGGCCCGGCCGCGGGGAG
CCCCGCTTCWTCKCAGTGGGCTACGTGGACGAC
ACCCAGTTCGTGAGGTTCGACAGSGACGCCGCG
AGTCCGAGRAAGGAGCCGCGGGCGCCGTGGAT
AGAGCAGGAGGGGCCGGAGTATTGGGACCGGR
ACACACRGATCTTGAAGGCCTMCGMMCWSACT
GACCGAGAGASCCTGCGGAWCCYGCTCCGCTAC
TACAACCAGAGCGAGGCCGGKGAGTGACCCCGG
SCCGGKGCGCAGGYSACGACTCCCCATCCCCCAC

39	CRTTGCMSGTTCTCGACGYWGCKCCWMTGTGC TCCCATCCCCCAYGTAA	CCTMCRARGAGKATGGGTTTGAAGGGGGSCCGA CGGGCGGCTCCTCCCCGGGGATGACCAGSCCGGC TACSACGGGACGGATTACMTCSCCCTGAACGAGG ACCTGSGCTCCTGGACCGGCGCGGACMCCGCGG CTCARATCACCCMSCCCAAGTTGGAGGGGGSCCC KGCGGYGGAGCAGATGARAACCTAACTGGAGGG SAGGKGCGGGGAGTGGCTCCGCTGATACCTGGAR AACGGAAAGGAAARGCTGCAMCGCGCCGGKACC ARGGGMAGKGGGGAGSCTTCCCCATCCCCCTATA RATTTCCCGGGATGGCCTCCCACSARAA
41	TRKKAMMTGCSGAGCGCGCCGGRAGGAGGKTC GGTCGSSTATCAAMC	CTCCAGKGAAGGWTTGGTCGCAACCGGGGGCCC GAGGGGSGCCTCCTCCGGGGKATTAACCGGYCCG CTTACAAGGGMAAGGTTAACTTSSCCCGGAMGA AGAACCGGSGTTCCGGGACGGCCGGGAACCCGG GGGTTAAAATCCCCCRSSGAARGGGGAAGGGGG CCCGGGGGGSGGAASMASTRAAAASCCAACCTGAA GGSCWGGGGGGGGAAGGGSTTCCSAAAATACCT GAAAAACGGAAAGGAACCGTTGMRSSGCGCGGK ACCCAGGGSMAGGGGGAAGCCTTCCCATCTTCTT GAAAACCTCCCGGATGGSCCTCCCACAAGA
43	CCTCRCCCCCAGGCTCCCACTCCATGAGGTATTTC TACACCKCCGTGTCCCGGCCCGGCCGCGGGGAG CCCCGCTTCWTCTSAKTGGGCTACGTGGACGAC ACCCAGTTCGTGAGGTTCGACRGSGACGCCGCG AGTCCGAGRAAGGAGCCGCGGGCGCCGTGGAT AGAGCAGGAGGGGCCGGAGTATTGGGACCGGA ACACACAGATCTAGAAGGMCCASGMMCWSACT GACCGAGAGASCCTGCGGAACCTGCKCSGSTACT ACAACCAGAGCGAGGCCGGKGAGTGACCCCGGS CCGSKGCGCASGYSMCGACTCCCCATCCCCCMC	CAGAGSATGKAYGGCTGCSACCTGGGGCCSGACG GGCGCCTCCTCCGSGGGATGACCAGTWCGCCTAC GACGGCAAGGATTACATCGCCCTGAACGAGGACC TGCGCTCCTGGACCGSCGCGGACACSGCGGCTCA RAWCACCCAGCGCAAGTKGGARGCGGCCCGTGM GGCGGARCAGCKGARARCCTACCTGGAGGGCRM GKGCGTGGAGTGGCTCCGCAGATACCTGGAGAA CGGGAAGGARAMGCTGSAGCGCGCWGGTACCA GGGGCAGTGGGGAGCCTTCCCCATCTCCTATAGR TCKCCSGGGATGGCCTCCCACGARAAGAR
44	CCTYGCCCCMRGGTCCCMCCTCMTTGAGGAATT TCWCCCCCCCGTGTCCGGGCCCGGCCGCGGGGA GCCCCGTTTCTTTKRRTGGGSTTCCGGGGCGGAC CCCCRSTTCKGAAGGTTCGARGGGAAGCCGYSAA WCCCAAGGAGGAAGCCCGGGSGCCCWTGGAA AAAKGAGAAGGGGCGGAAKTTTGGGAACCGGA ACCRCAGAATTTCCARAACCAACCCCCRAATTAC CCAAAAAAGCTGGGGAAACTGGGGGGGTTAYTA CACCCGAACGAAGSCCGGGGAKGGACCCCGGCC GGGGGGGARGKYMMSAATTCCCMTTCCCCMM SK	AAGAAGGWTGGSTTGSAACTGGGGGCCGGAGG GSSGCCTCCTCCGGGGSWTTAMCCRGTCCSCCTA CAASGGMAAGGTTTACTTCSCCCTGAASGAGGAC TTGSGTTCCGGGACCGCGGSGRACACGGSGGTTM AAATMMCCCASSGMARGGGGAAGGGGGCCCGG GGGGSGAASMASTGAAAASCCTACCTGAAGGGCY GGGGSGGGAAGGGSTTCCGMAAATACTTGAAAA ACGGAAAGGAAACGTTGMRSGGGGGGGKAACCA GGGGMAGGGGGAAGCCTTCCCCTTTTCCYGAAG ATKSCCGGGAATGSCCTCCCACAAAAAAGGA
52	CCTCGCCCCCAGGCTCCCACTCCATGAGGTATTT CYACACCGCCATGTCCCGGCCCGGCCGCGGGGA GCCCCGCTTCWTCRCMGTGGGCTACGTGGACG	CACCCTCCAGAGGATGWATGGCTGCGACCTGGG GCCCGACGGGCGCCTCCTCCGCGGGKATGACCAG TCCGCCTACGACGGCAAGGATTACATCGCCCTGA

	ACACSCTGTTCGTGAGGTTCGACAGCGACGCCRC	ACGAGGACCTGAGCTCCTGGACCGCSGCGGACAC
	GAGTCCGAGGAMGGAGCCSCGGGCGCCATGGA	CGCGGCTCAGATCACCCAGCGCAAGTKGGAGGCG
	TAGAGCAGGAGGGGCCGGAGTATTGGGACCGG	GCCCGTGCGGCGGAGCAGCTGAGAKCCTACCTGR
	GAGACACRGATCTYSAAGACCTCCKCSCWGACTT	AGGGCMYGTGCGKGGAGTGGCTCCGCAGATACC
	ACCGAGAGAACCTGCGGATCGCGCTCCGCTACT	TGRASAACGGRAAGGAGACGCTGCAGYGCGCRG
	ACARCCAGAGCGAGGCCGGTGAGTGACCCCGGC	GTACCAGGGGMAGTGGGGAGCCTTCCCCATCTC
	CCGGGGCGCAGGTCACGACTCCCCATCCCCCAC	CTATAGATCRCCSGGGATGGCCTCCCAC
53	TCCTCRCCCCCAGGCTCCCACTCCATGAGGTATTT	TCCRAAGAAGGWTGGSTTCSAACTGGGGSCCGA
	CTACACCKCCGTGTCCCGGCCCGGSCGCGGGGA	MGGGSGGCTTCCCCCGGGGSATTAMCCGKYCGS
	GCCCCGCTTCWTCTSAGYGRGCTACGTGGACGA	CTACCAAGGSMAGGGTTACCTTCSCCTGGACGAA
	CRCCCAGTTCGTGAGGYTCGACRGSGACGTCGCG	GAACTGGSGTTCTGGAACGGGGSGGAAACCGGG
	AGTCCSASGAAGGAGCCGCGGKCGCCGTGGATA	GSTTAAAATMCCCCRSGSAAGKGGGAAGGGGSCC
	RAGGAGGAGGGGCCGGAGTATTGGSACCGGAA	GGGGGGGGGGAASARSTGAAAASCTACCTGGAAG
	CACACAGATCTWCAAGACCCASKCMCAGACTGA	GSCWGGGGGGGGAAKGGSTTCCSARAAACCTGG
	CCGAGAGAGTCTGCGGAACCTGCGCGGCTACTA	AAAAAGGGAAAGAAAASSTTGARSGGGGGGGKA
	CMACCAGASCGAGGYCGGTGAGTGACCCCGGCC	CCCRGGGSMRKGGGGAASCTTCCCCCTTCTCCSRT
	CGGKGCGCAGGYSACGACTCCCCATCCCCCA	AGKCSSCCGGGATGGSCCTCCCMSAAAAR
57	CTYGCCCCMRGSTYCCMMTTCMTGARGKATTTC	CCAGGRGATGATGGGTGAGAAAAGGGGCCGGAC
	CWCMCCTGCGWGGYCCGGSCCGGSCGGGGGG	GGKCGCCTCCTCCGCGGGCTRACCCGTTCGCCTAC
	AASCCCGGTTCMTTTSMRKGGGSTTAGKGGAAG	GACGGAAGGATTACWTCGCCCTGAACGAGGACC
	GAMCCCCRKTCGGGAGGGTCGAAAGGGAACCC	TGAGCTCCTGGACCGCSGCGGACACSGCGGCTCA
	CSGAATCCCAAGAAGGGGGCCCGGGSGCCCWG	GATCACCCAGCGCAAGTGGGAGGCGGCCCGTGW
	GATARAASAAGAAGGGSCGGAATATTGGGAACG	GGCGGAGCARCKRAGAGCMTACCTGGAGGGSM
	GAAAACCCCGAACTCCGARACCCTCCGCMRACT	YGTGCGTGGAGTGGCTCCGCAGATACCTGGAGAA
	TACCGAAAAGAMCTGGGGAATCTGGKCCGSTAC	CGGGAAGGAGACGCTGCAGCGCGCGGGTACCAG
	CACCACCCRAASGAGGSCGGKGAAKGGACCCCG	GGGCAGTGGGGAGCCTTCCCCATCTCCTRTAGRTC
	SCCGGGGSGGARGKYMMSAATTCCCCCCCCCCC	KCCSGGGATGGCCTCCCACGAGAAGAGG
	MSKA	
58	CCTCGCCCCCAGGCTCCCACTCCWTGAGGTATTT	NNNNN
	CCACACCTCCGTGTCCCGGCCCGGCCGCGGGGA	
	GCCCCGCTTCATCTCAGTGGGCTACGTGGACGAC	
	ACCCAGTTCGTGAGGTTCGACAGCGACGCCGCG	
	AGTCCGAGARAGGAGCCGCGGGCGCCGTGGAT	
	AGAGCAGGAGGGGCCGGAGTATTGGGACCGGA	
	ACACACAGATCTACAAGGCCCAGGCACAGACTG	
	ACCGAGAGAGCCTGCGGAACCTGCGCGGCTACT	
	ACAACCAGAGCGAGGCCGGTGAGTGACCCCGGC	
	CCGGGGCGCAGGTCACGACTCCCCATCCCCCAC	

APPENDIX 6

Top 5 BLAST hits showing significant alignments with the highest bit scores for HLA-A locus. Alignments that produced unique bit scores are shown in bold.

Sample	Allele Source	Length (bp)	HLA-A Score (bits)	Identities %	Positives %	E-value
1 (exon 2)	A*31:08	546	430.2	97.8	97.8	8.10E-121
	A*24:99	822	430.2	97.8	97.8	8.10E-121
	A*24:98	822	430.2	97.8	97.8	8.10E-121
	A*24:97	546	430.2	97.8	97.8	8.10E-121
	A*24:96	822	430.2	97.8	97.8	8.10E-121
2	**A*01:19**	**546**	**450.3**	**99.6**	**99.6**	**1.50E-126**
	A*01:173	**546**	**450.3**	**99.6**	**99.6**	**1.50E-126**
	A*36:05	822	448.2	99.6	99.6	6.10E-126
	A*36:04	546	448.2	99.6	99.6	6.10E-126
	A*36:03	1012	448.2	99.6	99.6	6.10E-126
4	**A*01:19**	**546**	**443.9**	**98.7**	**98.7**	**1.10E-124**
	A*01:173	**546**	**443.9**	**98.7**	**98.7**	**1.10E-124**
	A*36:05	822	441.9	98.7	98.7	4.50E-124
	A*36:04	546	441.9	98.7	98.7	4.50E-124
	A*36:03	1012	441.9	98.7	98.7	4.50E-124
6	A*69:03	546	425	95.5	95.5	6.10E-119
	A*69:01:02	546	425	95.5	95.5	6.10E-119
	A*69:01:01:02	1098	425	95.5	95.5	6.10E-119
	A*69:01:01:01	1098	425	95.5	95.5	6.10E-119
	A*68:99	546	425	95.5	95.5	6.10E-119
11 (exon 3)	**A*03:222**	**546**	**413.1**	**97**	**97**	**1.20E-115**
	A*32:04	546	405	96.5	96.5	3.00E-113
	A*03:99	546	405	96.5	96.5	3.00E-113
	A*03:96	546	405	96.5	96.5	3.00E-113
	A*03:95	546	405	96.5	96.5	3.00E-113
12 (exon 2)	A*02:69	895	285.8	87.4	87.4	2.30E-77
	A*02:589	546	285.8	87.4	87.4	2.30E-77
	A*02:584	546	285.8	87.4	87.4	2.30E-77
	A*02:579	546	285.8	87.4	87.4	2.30E-77
	A*02:503	546	285.8	87.4	87.4	2.30E-77
13	A*30:99	1098	406.3	93.7	93.7	2.50E-113
	A*30:84	546	406.3	93.7	93.7	2.50E-113
	A*30:80	822	406.3	93.7	93.7	2.50E-113

	A*30:77	822	406.3	93.7	93.7	2.50E-113
	A*30:69	546	406.3	93.7	93.7	2.50E-113
14	A*68:31	822	430.4	96.4	96.4	1.50E-120
	A*68:20	546	430.4	96.4	96.4	1.50E-120
	A*68:15	1098	430.4	96.4	96.4	1.50E-120
	A*68:05	1098	430.4	96.4	96.4	1.50E-120
	A*68:03:03	546	430.4	96.4	96.4	1.50E-120
17	A*80:02	822	424.7	96.4	96.4	7.60E-119
	A*68:66:02	1098	424.7	96.4	96.4	7.60E-119
	A*68:66:01	546	424.7	96.4	96.4	7.60E-119
	A*68:14	546	424.7	96.4	96.4	7.60E-119
	A*68:10	546	424.7	96.4	96.4	7.60E-119
19	A*03:89:01	546	457.9	100	100	7.60E-129
	A*02:99	546	457.9	100	100	7.60E-129
	A*02:97:02	822	457.9	100	100	7.60E-129
	A*02:97:01	822	457.9	100	100	7.60E-129
	A*02:95	1098	457.9	100	100	7.60E-129
20	A*24:89	822	423.4	96.4	96.4	1.80E-118
	A*11:18	546	423.4	96.4	96.4	1.80E-118
	A*03:89:02	1098	423.4	96.4	96.4	1.80E-118
	A*03:89:01	546	423.4	96.4	96.4	1.80E-118
	A*03:198	546	423.4	96.4	96.4	1.80E-118
21 (exon 2)	A*02:676	546	431.3	97.8	97.8	3.80E-121
	A*02:655	1098	431.3	97.8	97.8	3.80E-121
	A*02:646	1098	431.3	97.8	97.8	3.80E-121
	A*02:643N	822	431.3	97.8	97.8	3.80E-121
	A*02:631	546	431.3	97.8	97.8	3.80E-121
22 (exon 2)	A*31:08	546	400.3	93.3	93.3	8.10E-112
	A*24:99	822	400.3	93.3	93.3	8.10E-112
	A*24:98	822	400.3	93.3	93.3	8.10E-112
	A*24:97	546	400.3	93.3	93.3	8.10E-112
	A*24:96	822	400.3	93.3	93.3	8.10E-112
25	A*31:08	546	450.3	99.1	99.1	1.50E-126
	A*24:99	822	450.3	99.1	99.1	1.50E-126
	A*24:98	822	450.3	99.1	99.1	1.50E-126
	A*24:97	546	450.3	99.1	99.1	1.50E-126
	A*24:96	822	450.3	99.1	99.1	1.50E-126
26 (exon 2)	A*03:89:01	546	406.8	93.3	93.3	8.90E-114
	A*02:99	546	406.8	93.3	93.3	8.90E-114
	A*02:97:02	822	406.8	93.3	93.3	8.90E-114

	A*02:97:01	822	406.8	93.3	93.3	8.90E-114
	A*02:95	1098	406.8	93.3	93.3	8.90E-114
27 (exon 2)	A*69:03	546	404.4	94.1	94.1	4.60E-113
	A*69:01:02	546	404.4	94.1	94.1	4.60E-113
	A*69:01:01:02	1098	404.4	94.1	94.1	4.60E-113
	A*69:01:01:01	1098	404.4	94.1	94.1	4.60E-113
	A*68:99	546	404.4	94.1	94.1	4.60E-113
28	A*30:88	546	339.9	90.9	90.9	2.50E-93
	A*30:85	546	339.9	90.9	90.9	2.50E-93
	A*30:80	822	339.9	90.9	90.9	2.50E-93
	A*30:76N	546	339.9	90.9	90.9	2.50E-93
	A*30:70N	545	339.9	90.9	90.9	2.50E-93
30 (exon 2)	A*31:08	546	439.7	99.6	99.6	1.10E-123
	A*24:99	822	439.7	99.6	99.6	1.10E-123
	A*24:98	822	439.7	99.6	99.6	1.10E-123
	A*24:97	546	439.7	99.6	99.6	1.10E-123
	A*24:96	822	439.7	99.6	99.6	1.10E-123
31 (exon 2)	A*02:676	546	425.1	97.8	97.8	2.80E-119
	A*02:655	1098	425.1	97.8	97.8	2.80E-119
	A*02:646	1098	425.1	97.8	97.8	2.80E-119
	A*02:643N	822	425.1	97.8	97.8	2.80E-119
	A*02:631	546	425.1	97.8	97.8	2.80E-119
34	A*31:08	546	442.4	97.8	97.8	3.50E-124
	A*24:99	822	442.4	97.8	97.8	3.50E-124
	A*24:98	822	442.4	97.8	97.8	3.50E-124
	A*24:97	546	442.4	97.8	97.8	3.50E-124
	A*24:96	822	442.4	97.8	97.8	3.50E-124
37	A*31:08	546	457.9	100	100	7.60E-129
	A*24:99	822	457.9	100	100	7.60E-129
	A*24:98	822	457.9	100	100	7.60E-129
	A*24:97	546	457.9	100	100	7.60E-129
	A*24:96	822	457.9	100	100	7.60E-129
38 (exon 2)	A*69:03	546	401.4	93.3	93.3	3.90E-112
	A*69:02	546	401.4	93.3	93.3	3.90E-112
	A*69:01:02	546	401.4	93.3	93.3	3.90E-112
	A*69:01:01:02	1098	401.4	93.3	93.3	3.90E-112
	A*69:01:01:01	1098	401.4	93.3	93.3	3.90E-112
39 (exon 2)	A*30:98	546	421.5	96	96	3.30E-118
	A*30:95	822	421.5	96	96	3.30E-118
	A*30:93	1025	421.5	96	96	3.30E-118

	A*30:92	546	421.5	96	96	3.30E-118
	A*30:89	1098	421.5	96	96	3.30E-118
41	A*68:30	546	434.9	96.9	96.9	6.30E-122
	A*33:09	895	434.9	96.9	96.9	6.30E-122
	A*02:90	822	434.9	96.9	96.9	6.30E-122
	A*02:634	546	434.9	96.9	96.9	6.30E-122
	A*02:415	822	434.9	96.9	96.9	6.30E-122
43 (exon 2)	A*30:99	1098	430.2	97.8	97.8	8.10E-121
	A*30:98	546	430.2	97.8	97.8	8.10E-121
	A*30:95	822	430.2	97.8	97.8	8.10E-121
	A*30:93	1025	430.2	97.8	97.8	8.10E-121
	A*30:92	546	430.2	97.8	97.8	8.10E-121
44	A*68:66:02	1098	409.1	94.2	94.2	3.80E-114
	A*68:66:01	546	409.1	94.2	94.2	3.80E-114
	A*68:13:02	546	409.1	94.2	94.2	3.80E-114
	A*68:13:01	866	409.1	94.2	94.2	3.80E-114
	A*31:89	546	409.1	94.2	94.2	3.80E-114
52 (exon 2)	A*02:676	546	408.5	94.2	94.2	2.70E-114
	A*02:655	1098	408.5	94.2	94.2	2.70E-114
	A*02:646	1098	408.5	94.2	94.2	2.70E-114
	A*02:643N	822	408.5	94.2	94.2	2.70E-114
	A*02:631	546	408.5	94.2	94.2	2.70E-114
53	A*30:98	546	434.6	98.2	98.2	7.60E-122
	A*30:96	546	434.6	98.2	98.2	7.60E-122
	A*30:95	822	434.6	98.2	98.2	7.60E-122
	A*30:93	1025	434.6	98.2	98.2	7.60E-122
	A*30:92	546	434.6	98.2	98.2	7.60E-122
57 (exon 2)	A*31:10	822	396.2	93.2	93.2	1.30E-110
	A*31:08	546	396.2	93.2	93.2	1.30E-110
	A*31:07	546	396.2	93.2	93.2	1.30E-110
	A*30:99	1098	396.2	93.2	93.2	1.30E-110
	A*30:84	546	396.2	93.2	93.2	1.30E-110
58	A*30:98	546	317.1	86.5	86.5	1.80E-86
	A*30:95	822	317.1	86.5	86.5	1.80E-86
	A*30:93	1025	317.1	86.5	86.5	1.80E-86
	A*30:92	546	317.1	86.5	86.5	1.80E-86
	A*30:89	1098	317.1	86.5	86.5	1.80E-86

APPENDIX 7

Top 5 BLAST hits showing significant alignments with the highest bit scores for HLA-A locus. Alignments that produced unique bit scores are shown in bold.

			HLA-B			
Sample	**Allele Source**	**Length (bp)**	**Score (bits)**	**Identities %**	**Positives %**	**E-value**
1	B*57:83	1089	562.7	99.6	99.6	2.70E-160
	B*57:82	1089	562.7	99.6	99.6	2.70E-160
	B*57:79N	1088	562.7	99.6	99.6	2.70E-160
	B*57:29	1089	562.7	99.6	99.6	2.70E-160
	B*57:11	1089	562.7	99.6	99.6	2.70E-160
2 (exon 2)	**B*27:05:05**	**895**	**487.9**	**93.4**	**93.4**	**4.50E-138**
	B*50:51	1089	485.8	93.4	93.4	1.90E-137
	B*50:44	1089	485.8	93.4	93.4	1.90E-137
	B*50:37	1012	485.8	93.4	93.4	1.90E-137
	B*50:02	1089	485.8	93.4	93.4	1.90E-137
4 (exon 2)	B*53:01:01	1089	523.2	97.8	97.8	1.00E-148
	B*51:148	1089	523.2	97.8	97.8	1.00E-148
	B*51:05	1089	523.2	97.8	97.8	1.00E-148
	B*51:02:02	1089	523.2	97.8	97.8	1.00E-148
	B*51:01:02	1089	523.2	97.8	97.8	1.00E-148
6	B*81:02	932	516.9	96.7	96.7	1.60E-146
	B*81:01	1089	516.9	96.7	96.7	1.60E-146
	B*67:05	1012	516.9	96.7	96.7	1.60E-146
	B*67:03	895	516.9	96.7	96.7	1.60E-146
	B*67:01:02	1089	516.9	96.7	96.7	1.60E-146
11	B*58:90	546	519.7	96.3	96.3	2.40E-147
	B*58:89	546	519.7	96.3	96.3	2.40E-147
	B*58:85	546	519.7	96.3	96.3	2.40E-147
	B*58:83	546	519.7	96.3	96.3	2.40E-147
	B*58:81	546	519.7	96.3	96.3	2.40E-147
12	B*78:02:01	1089	515.1	93.4	93.4	6.00E-146
	B*59:07	1012	515.1	93.4	93.4	6.00E-146
	B*59:01:01:02	1089	515.1	93.4	93.4	6.00E-146
	B*59:01:01:01	1089	515.1	93.4	93.4	6.00E-146
	B*53:05:02	1089	515.1	93.4	93.4	6.00E-146
13 (exon 2)	B*81:02	932	536.8	99.6	99.6	8.40E-153
	B*81:01	1089	536.8	99.6	99.6	8.40E-153
	B*42:08	1089	536.8	99.6	99.6	8.40E-153

	B*42:05:02	619	536.8	99.6	99.6	8.40E-153
	B*42:01:01	1089	536.8	99.6	99.6	8.40E-153
14	**B*51:158:01**	**546**	**509.8**	**97**	**97**	**2.30E-144**
	B*78:09	1089	501.7	96.7	96.7	6.50E-142
	B*78:06	822	501.7	96.7	96.7	6.50E-142
	B*78:05	822	501.7	96.7	96.7	6.50E-142
	B*78:03	546	501.7	96.7	96.7	6.50E-142
17	*C*18:10*	*1101*	*473.1*	*91.7*	*91.7*	*2.70E-133*
	*C*18:09*	*546*	*473.1*	*91.7*	*91.7*	*2.70E-133*
	*C*18:06*	*546*	*473.1*	*91.7*	*91.7*	*2.70E-133*
	*C*18:05*	*546*	*473.1*	*91.7*	*91.7*	*2.70E-133*
	*C*18:04*	*546*	*473.1*	*91.7*	*91.7*	*2.70E-133*
19 (ex 3)	***C*04:18***	***546***	***134.6***	***76.8***	***76.8***	***9.80E-32***
	C*04:01:62	***1101***	***134.6***	***76.8***	***76.8***	***9.80E-32***
	C*04:01:17	***546***	***134.6***	***76.8***	***76.8***	***9.80E-32***
	*C*18:10*	*1101*	*126*	*76.3*	*76.3*	*4.00E-29*
	*C*18:09*	*546*	*126*	*76.3*	*76.3*	*4.00E-29*
20	B*50:51	1089	514.9	96.7	96.7	6.70E-146
	B*50:44	1089	514.9	96.7	96.7	6.70E-146
	B*50:37	1012	514.9	96.7	96.7	6.70E-146
	B*50:02	1089	514.9	96.7	96.7	6.70E-146
	B*50:01:08	1089	514.9	96.7	96.7	6.70E-146
21	B*53:01:01	1089	480.7	94.5	94.5	1.30E-135
	B*51:148	1089	480.7	94.5	94.5	1.30E-135
	B*51:05	1089	480.7	94.5	94.5	1.30E-135
	B*51:02:02	1089	480.7	94.5	94.5	1.30E-135
	B*51:01:02	1089	480.7	94.5	94.5	1.30E-135
22 (exon 2)	**B*49:07**	**1089**	**500.2**	**94.9**	**94.9**	**8.80E-142**
	B*49:01:08	**1089**	**500.2**	**94.9**	**94.9**	**8.80E-142**
	B*49:01:01	**1089**	**500.2**	**94.9**	**94.9**	**8.80E-142**
	B*44:18	**1089**	**500.2**	**94.9**	**94.9**	**8.80E-142**
	B*49:47	546	498.1	94.8	94.8	3.60E-141
25	B*45:19	1089	526	98.1	98.1	3.00E-149
	B*45:16	546	526	98.1	98.1	3.00E-149
	B*45:13	822	526	98.1	98.1	3.00E-149
	B*45:07	822	526	98.1	98.1	3.00E-149
	B*45:06	546	526	98.1	98.1	3.00E-149
26	N/A					

27 (exon 2)	B*53:01:01	1089	516.5	96.7	96.7	1.10E-146
	B*51:148	1089	516.5	96.7	96.7	1.10E-146
	B*51:05	1089	516.5	96.7	96.7	1.10E-146
	B*51:02:02	1089	516.5	96.7	96.7	1.10E-146
	B*51:01:02	1089	516.5	96.7	96.7	1.10E-146
28	B*53:01:01	1089	524	95.6	95.6	1.20E-148
	B*51:148	1089	524	95.6	95.6	1.20E-148
	B*51:05	1089	524	95.6	95.6	1.20E-148
	B*51:02:02	1089	524	95.6	95.6	1.20E-148
	B*51:01:02	1089	524	95.6	95.6	1.20E-148
30 (exon 2)	**B*15:01:01:02N**	**1208**	**508.3**	**94.7**	**94.7**	**3.10E-144**
	B*78:09	1089	502.2	95.2	95.2	2.20E-142
	B*78:02:02	1089	502.2	95.2	95.2	2.20E-142
	B*78:02:01	1089	502.2	95.2	95.2	2.20E-142
	B*53:05:02	1089	502.2	95.2	95.2	2.20E-142
31	N/A					
34	B*15:411	1089	475.6	94.9	94.9	4.50E-134
	B*15:408	1089	475.6	94.9	94.9	4.50E-134
	B*15:196	1012	475.6	94.9	94.9	4.50E-134
	B*15:17:03	1089	475.6	94.9	94.9	4.50E-134
	B*15:17:01:02	1089	475.6	94.9	94.9	4.50E-134
37 (exon 2)	B*53:01:01	1089	481.6	93.4	93.4	3.40E-136
	B*51:148	1089	481.6	93.4	93.4	3.40E-136
	B*51:05	1089	481.6	93.4	93.4	3.40E-136
	B*51:02:02	1089	481.6	93.4	93.4	3.40E-136
	B*51:01:02	1089	481.6	93.4	93.4	3.40E-136
38 (exon 2)	**B*07:273**	**546**	**450.5**	**92.3**	**92.3**	**8.20E-127**
	B*81:05	546	446.3	91.9	91.9	1.50E-125
	B*07:83	546	446.3	91.9	91.9	1.50E-125
	B*07:81	546	446.3	91.9	91.9	1.50E-125
	B*07:79	546	446.3	91.9	91.9	1.50E-125
39 (exon 3)	*C*07:58*	*546*	*264.6*	*84.8*	*84.8*	*7.10E-71*
	*C*07:573*	*546*	*264.6*	*84.8*	*84.8*	*7.10E-71*
	*C*07:537*	*546*	*264.6*	*84.8*	*84.8*	*7.10E-71*
	*C*07:412*	*1101*	*264.6*	*84.8*	*84.8*	*7.10E-71*
	*C*07:397*	*546*	*264.6*	*84.8*	*84.8*	*7.10E-71*

41 (exon 3)	***C*15:109***	***546***	***47.1***	***92.9***	***92.9***	***2.30E-05***
	C*08:46	***822***	***47.1***	***95.8***	***95.8***	***2.30E-05***
	*C*07:02:67*	*1015*	*44.9*	*92.6*	*92.6*	*9.80E-05*
	B*40:01:32	1012	36.5	88.9	88.9	0.035
	B*39:25N	820	34.4	88.5	88.5	0.15
43	**B*07:133**	**546**	**490.3**	**93.4**	**93.4**	**1.70E-138**
	B*81:02	932	484	93	93	1.40E-136
	B*81:01	1089	484	93	93	1.40E-136
	B*42:08	1089	484	93	93	1.40E-136
	B*42:05:02	619	484	93	93	1.40E-136
44	**B*57:85**	**546**	**86.7**	**78.2**	**78.2**	**5.30E-17**
	B*40:01:32	1012	82.4	77.3	77.3	1.00E-15
	*C*04:197*	*546*	*78.1*	*76.4*	*76.4*	*2.10E-14*
	B*07:190	546	73.8	76.4	76.4	4.10E-13
	*C*07:570*	*546*	*71.6*	*76.2*	*76.2*	*1.80E-12*
52	***C*07:450***	***1101***	***494***	***95.1***	***95.1***	***1.30E-139***
	*C*15:65*	*546*	*485.9*	*94.8*	*94.8*	*3.70E-137*
	*C*12:181*	*1101*	*485.9*	*94.8*	*94.8*	*3.70E-137*
	*C*07:96:02*	*546*	*485.9*	*94.8*	*94.8*	*3.70E-137*
	*C*07:96:01*	*546*	*485.9*	*94.8*	*94.8*	*3.70E-137*
53	**B*07:133**	**546**	**455.1**	**90.4**	**90.4**	**6.80E-128**
	B*07:39	546	446.5	90	90	2.70E-125
	B*07:125	546	446.5	90	90	2.70E-125
	B*81:02	932	440	89.7	89.7	2.40E-123
	B*81:01	1089	440	89.7	89.7	2.40E-123
57	B*59:03	822	413.2	93.3	93.3	2.80E-115
	B*57:76	546	413.2	93.3	93.3	2.80E-115
	B*56:22	546	413.2	93.3	93.3	2.80E-115
	B*55:13	546	413.2	93.3	93.3	2.80E-115
	B*39:34	1089	413.2	93.3	93.3	2.80E-115
58 (exon 2)	**B*42:02:01:02**	**1089**	**533.9**	**99.3**	**99.3**	**6.30E-152**
	B*42:02:01:01	**1089**	**533.9**	**99.3**	**99.3**	**6.30E-152**
	B*42:17	546	531.9	99.3	99.3	2.50E-151
	B*42:09	546	531.9	99.3	99.3	2.50E-151
	B*07:143	546	531.9	99.3	99.3	2.50E-151