6-2019

Genetic Variation of Human Leukocyte Antigen (Hla) Alleles among Rheumatiod Arthritis Patients in Fujairah

Amna Abdulla Harib Al-Dhmanie

United Arab Emirates University

College of Science

Department of Biology

GENETIC VARIATION OF HUMAN LEUKOCYTE ANTIGEN (HLA) ALLELES AMONG RHEUMATIOD ARTHRITIS PATIENTS IN FUJAIRAH

Amna Abdulla Harib Al-Dhmanie

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Science in Molecular Biology and Biotechnology

Under the Supervision of Dr. Ranjit Vijayan

June 2019

Declaration of Original Work

I, Amna Abdulla Harib Al-Dhmanie, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Genetic Variation of Human Leukocyte Antigen (HLA) Alleles among Rheumatoid Arthritis Patients in Fujairah*", hereby, solemnly declare that this thesis is my own original research work that has been done and prepared by me under the supervision of Dr. Ranjit Vijayan, in the College of Science at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: _____ Date: 21/7/2019

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Dr. Ranjit Vijayan
   Title: Associate Professor
   Department of Biology
   College of Science
   Signature _____ Date 11/6/2019

2) Member: Dr Mohammed Akli Ayoub
   Title: Assistant Professor
   Department of Biology
   College of Science
   Signature _____ Date 16/06/2019

3) Member (External Examiner): Prof. Syed Salman Ashraf
   Title: Professor
   Department of Chemistry
   Institution: Khalifa University, Abu Dhabi, UAE
   Signature Salman Ashraf Date 16/06/2019

This Master Thesis is accepted by:

Dean of the College of Science: Professor Ahmed Murad

Signature _____ Date 2/7/2019

Acting Dean of the College of Graduate Studies: Professor Ali Hassan Al Marzouqi

Signature _____ Date 18/11/2019

Copy 1 of 6

Abstract

Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation that affects the joints. It occurs when the immune system attacks the body's own tissues and organs. The precise cause of RA is not fully understood but evidence suggests that genes, hormones and environmental factors could be involved. It has been established that human leukocyte antigen (HLA) loci is a strong risk factor of RA. Specifically, the shared epitope (SE) region of HLA-DRB1 and variations at position 11 and 13 of HLA-DRB1 and position 77 of HLA-A have shown strong association with RA. The hypothesis of this study was that these variations could be associated with RA patients in the UAE. To study this, regions of the HLA-DRB1 and HLA-A genes from RA and control samples from Fujairah Hospital were sequenced and analyzed. Results indicated that there was minimal incidence of high risk HLA-DRB1 SE or variations at position 11 and 13 of HLA-DRB1 in these samples. An asparagine at position 77 (Asn77) appeared to be a much strong biomarker for RA in this population (Odds ratio = 20.52, p-value = 0.0001985). Additionally, missense mutations Arg56 and Glu76 were also observed with significantly higher incidence in RA samples. The relevance of these variations warrant further investigation. In summary, HLA-A appears to be a stronger indicator of RA than HLA-DRB1 in the samples analyzed.

Keywords: Rheumatoid arthritis, human leukocyte antigen, shared epitope, HLA-DRB1, HLA-A.

Title and Abstract (in Arabic)

التباين الوراثي لأليلات الكريات الدم البيض البشرية (HLA) بين مرضى التهاب المفاصل الروماتويدي في الفجيرة

الملخص

التهاب المفاصل الروماتويدي (RA) هو مرض مناعي ذاتي يتميز بالتهاب مزمن يصيب المفاصل. يحدث عندما يهاجم الجهاز المناعي أنسجة الجسم وأعضائه. السبب الأساسي للإصابة بإلتهاب المفاصل الروماتيدي ليس مفهوما بالكامل، لكن الأدلة تشير إلى أن الجينات والهرمونات والعوامل البيئية يمكن أن تكون سببا في ذلك. لقد ثبت أيضا أن مستضد كريات الدم البيضاء البشرية (HLA)، هو عامل خطر وقوي لالتهاب المفاصل الروماتويدي. على وجه التحديد، أظهرت منطقة الحلقة المشتركة (SE) من HLA-DRB1 والتغيرات في الموضع 11 و13 من HLA-DRB1 والموقع 77 من HLA-A ارتباطًا قويًا بالتهاب المفاصل الروماتيدي. وقد وضعنا فرضية أن هذه الاختلافات يمكن أن ترتبط بمرضى التهاب المفاصل الروماتويدي في دولة الإمارات العربية المتحدة. وعليه تم دراسة وتقييم تسلسل الحمض النووي HLA-DRB1 وHLA-A من مرضى إلتهاب المفاصل الروماتيدي من مستشفى الفجيرة ومقارنتها بعينات لأشخاص غير مصابيين بالمرض. أشارت النتائج إلى أن هناك نسبة ضئيلة من HLA-DRB1 SE عالية المخاطر وكذلك نسبة خطورة ضئيلة في الاختلافات في الموضع 11 و13 من HLA-DRB1 في هذه العينات. ظهر الحمض الأميني (اسبرجين) في الموضع 77 (Asn77) ليكون علامة حيوية قوية للمرض في هذه الفئة من السكان (نسبة الأرجحية = 20.52، القيمة p = 0.0001985). بالإضافة إلى ذلك؛ لوحظت طفرات خطأ في موقعيين آخريين Arg56 وGlu76 بنسبة أعلى بكثير في عينات في فئة المصابيين بالمرض. أهمية هذه

الاختلافات يستدعي المزيد من التحقيق. باختصار، يبدو أن HLA-A هو مؤشر أقوى على RA من HLA-DRB1 في العينات التي تم تحليلها.

مفاهيم البحث الرئيسية: التهاب المفاصل الروماتويدي، مستضد كريات الدم البيضاء، حلق مشترك، HLA-DRB1، HLA-A.

Acknowledgements

I would like to extend thanks to the many people, who so generously contributed to the work presented in this thesis.

Special mention goes to my enthusiastic supervisor Dr. Ranjit Vijayan for the useful comments, remarks and engagement through the learning process of this master thesis. You have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist.

I would like to express my gratitude to my co-supervisor Dr. Yusra Al Dhaheri for her academic support and for giving me so many wonderful opportunities. Similar, profound gratitude goes to Mr. Amanat Ali, who has been a truly dedicated mentor in my lab work. I am particularly indebted to Dr. Mohammed Enan for his support in this journey in UAE University's laboratories. I have very fond memories of my time there.

I would like to thank my committee for their guidance, support, and assistance during the preparation of this thesis. I would also like to thank the Chairman and all members of the Department of Biology at the United Arab Emirates University for assisting me throughout my studies and research.

Special thanks go to my mother, brothers, sisters, family, and friends who helped me along the way and for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. In addition, special thanks are extended to Mr. Al Halaky for his assistance and friendship.

Dedication

To my beloved parents, family, friends, and teachers

Table of Contents

Title i
Declaration of Original Work ii
Copyright iii
Advisory Committee iv
Approval of the Master Thesis v
Abstract vii
Title and Abstract (in Arabic) viii
Acknowledgements x
Dedication xi
Table of Contents xii
List of Tables xiv
List of Figures xv
List of Abbreviations xvii
Chapter 1: Introduction 1
1.1 Human Genome & Genetic Diversity 1
1.2 Overview of Rheumatoid Arthritis 2
1.3 History of Rheumatoid Arthritis 3
1.4 Common Types of Arthritis 5
1.5 Pathophysiology of RA 7
1.6 Autoimmune Response 7
1.7 Autoantibodies Associated with RA 9
1.8 Genetic Association of RA 10
1.9 Major Histocompatibility Complex (MHC) 11
1.10 Class I MHC/HLA Genes 12
1.11 Class II MHC/HLA Genes 14
1.12 Class III HLA Genes 15
1.13 MHC Functions and Pathways 16
1.14 Nomenclature and HLA Classifications 17
1.15 Association of HLA Genes with RA 20
1.16 RA in United Arab Emirates 24
1.17 Whatman FTA Technology 24
1.18 Hypothesis 26
Chapter 2: Methods 27
2.1 Ethical Approval 27

2.2 Patients and Controls 27
2.3 Sample Collection 27
2.4 Storage Condition 28
2.5 DNA Extraction 28
2.6 Primers 29
2.7 PCR amplification 31
2.8 Gel Electrophoresis 32
2.9 Purification of PCR products 32
2.10 DNA Sequencing 33
2.11 Analysis of Sequencing Data 33
2.12 HLA Allele Type Determination 33
2.13 Homology Modeling 34
2.14 Statistical Analysis 35

Chapter 3: Results 36
3.1 HLA Gene PCR Amplification 36
3.2 HLA-DRB1 Allele Typing 46
3.3 Sequence of Shared Epitope and Other Regions of HLA-DRB1 Associated with RA 50
3.4 HLA-A Allele Typing 54
3.5 Amino Acid Position in HLA-A Associated with RA 58
3.6 Other Locations in HLA-A with Higher Incidence in Patient 61
3.7 Odds Ratio for the Three Observed Positions in HLA-A 62
3.9 Principal Component Analysis 66
3.10 Fixation Index 67

Chapter 4: Discussion 68

Chapter 5: Conclusion 74

References 75

List of Tables

Table 1: The final aspect of HLA naming 20
Table 2: Distribution of HLA-DRB1 alleles and genotypes frequencies in RA among different population 23
Table 3: Primers set used in PCR to amplify different HLA genes 30
Table 4: HLA-DRB1 allele type in control (C) samples 46
Table 5: Summary of HLA-DRB1 allele types in control samples at 4-digit and 2-digit resolutions 47
Table 6: HLA-DRB1 allele types in patient (RA) samples 48
Table 7: Summary of HLA-DRB1 allele types in RA samples at 4-digit and 2-digit resolutions 49
Table 8: HLA-DRB1 Shared epitope and positions 11/13 in control samples 50
Table 9: Count of shared epitope sequence in control samples 52
Table 10: HLA-DRB1 Shared epitope and positions 11/13 in RA samples 52
Table 11: Count of shared epitope in RA samples 52
Table 12: HLA-A allele type in control samples 55
Table 13: Summary of HLA-A allele types in control samples at 4- and 2-digit resolutions 56
Table 14: HLA-A allele type in RA samples 56
Table 15: Summary of HLA-A allele types in RA samples at 4- and 2-digit resolutions 58
Table 16: HLA-A amino acid at position 77 in control samples 59
Table 17: HLA-A amino acid at position 77 in RA samples 60
Table 18: Summary of HLA-A amino acid at position 77 in all samples 61
Table 19: Variations in HLA-A with higher incidence in RA samples than control samples 63

List of Figures

Figure 1: Schematic view of a normal joint (a) and a joint affected by RA (b)..... 8
Figure 2: Antigens are the molecules recognized by the immune response, while epitopes are sites within antigens to which antigen receptors bind 10
Figure 3: An antibody binds an antigen directly, whereas a T-cell receptor binds a complex of antigen fragment and self-molecule 10
Figure 4: The structure of an MHC class I molecule 12
Figure 5: The relative position of the HLA genes on the short arm of chromosome 6 and the exons to be considered 12
Figure 6: The structure of an MHC class I molecule 13
Figure 7: Expression of MHC alleles is codominant. 14
Figure 8: The structure of an MHC class II molecule 15
Figure 9: The genetic organization of the major histocompatibility-complex (MHC) in human 16
Figure 10: The (a) MHC class I and (b) MHC class II antigen-presentation pathway 17
Figure 11: Interpretation of digits in HLA genes 19
Figure 12: Nomenclature of HLA Alleles 19
Figure 13: 3D models of HLA proteins with amino acid positions that are significantly associated with RA 21
Figure 14: 3D models of HLA proteins with additional amino acid positions that are known to be associated with RA 22
Figure 15: (A) Whatman FTA cards (B) FTA card that turns white when colorless samples are loaded 25
Figure 16: FTA card spotted with human whole blood 28
Figure 17: Harris® Uni-Core™ 1.2 mm puncher 29
Figure 18: BioRad T100™ Thermal Cycler 31
Figure 19: A custom BLAST server to search HLA sequences obtained from HLA Nomenclature 34

Figure 20: Gel electrophoresis image of PCR products of HLA-DRB1 exon 2 36
Figure 21: Gel electrophoresis image of PCR products of HLA-A exon 2 36
Figure 22: Gel electrophoresis image of PCR products of HLA-A exon 3 37
Figure 23: Chromatogram of an HLA-DRB1 sample visualized in FinchTV viewer 37
Figure 24: Sequence alignment of sample HLA-DRB1 exon 2 38
Figure 25: Sequence alignment of sample HLA-A exon 2 42
Figure 26: Sequence alignment of sample HLA-A exon 3 44
Figure 27: Three-dimensional models of wildtype and mutant HLA-A 62
Figure 28: Phylogenetic constructed from HLA-DRB1 exon 2 sequences 64
Figure 29: Phylogenetic constructed from all HLA-A exon 2 & 3 sequences 65
Figure 30: A plot of principal component 1 against component 2 obtained from PCA of HLA-DRB1 exon 2 sequences. 66
Figure 31: A plot of principal component 1 against component 2 obtained from PCA of HLA-A exon 2 and 3 sequences. 67

List of Abbreviations

ACPA	Anti–Citrullinated Protein Antibody
BLAST	Basic Local Alignment Search Tool
DNA	Deoxyribonucleic Acid
F_{ST}	Fixation Index
FTA	Flinders Technology Associates
HLA	Human Leukocyte Antigen
HLA-DRB1	Human Leukocyte Antigen – DR β chain 1
NCBI	National Center for Biotechnology Information
OA	Osteoarthritis
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RA	Rheumatoid Arthritis
RF	Rheumatoid Factor
SE	Shared Epitope

Chapter 1: Introduction

1.1 Human Genome & Genetic Diversity

Deoxyribonucleic acids (DNA) is the carrier of genetic information in most organisms. It is composed of four chemical bases: Adenine (A), Thymine (T), Cytosine (C) and Guanine (G). The Human Genome Project (HGP) was a ground-breaking undertaking that was initiated to determine the DNA sequence of the human genome and to map chromosomes to show the specific locations of genes (Holers, 2013). The first draft of the human genome sequence was completed in 2001. A natural extension of the HGP was to determine the genetic diversity among individuals and populations. Understanding genetic diversity in human population, and individual genetic variations, has several applications including the evaluation of the predisposition to genetic diseases, personalized medicine, pharmacology and even forensics (Smolen et al., 2016). Genetic variations or mutations result from changes in nucleotide sequence of the DNA; such as deletions, insertions, or even rearrangement of DNA sequences in the genome. DNA sequencing is an important technique used to define the exact order of nucleic acid bases to investigate gene function (Holers, 2013). The DNA of humans are 99.9% identical and the 0.1% difference accounts for several phenotypic variations and genetic disorders. While the genome plays a vital role encoding an individual's genetic makeup, the environment also critically impacts an individual. A comprehensive understanding of gene-environment interactions for high incidence diseases, especially in families and populations, such as type 2 diabetes, cardiovascular diseases, sickle cell anemia, cancer, and autoimmune disorders, could help in early

detection, treatment and management of these conditions (Cush et al., 2010; Entezami et al., 2011; Genetics Home Reference, 2013). One such high incidence autoimmune disease with genetic basis is rheumatoid arthritis (RA) (Smolen et al., 2016).

1.2 Overview of Rheumatoid Arthritis

RA is the second most predominant autoimmune disorder (Holers, 2013). RA is defined as a chronic inflammatory joint disease affecting, in particular, hands and feet leading to cartilage and bone erosion (Smolen et al., 2016). It occurs when the immune system attacks the body's own tissues and organs. RA is estimated to be found in up to 1% of the population (Choy, 2012; Oka et al., 2014; Saxena et al., 2017; Smolen et al., 2016). It is more common in women than in men, with a rate of two to three times more in women, which may be related to hormonal factors (Genetics Home Reference, 2013; Boekel et al., 2001). Lower prevalence rates of RA have been described in Asian countries and rural Africa. The highest prevalence rate was described in American and European populations (Carmona et al., 2002). Geographical factors also play a role in RA since environmental factors are not equally distributed. In this case, area of residency, nutritional factors, socioeconomic factors and other factors affect the prevalence of RA (Martín et al., 2011). The general public associates the term “arthritis” with pain and stiffness in joints. Arthritis in Greek means inflammation (itis) of the joints (arthron). This term describes a collection of more than hundred joints disorders without inflammation (Shlotzhauer et al., 2014). The term “*rheuma”* made its first appearance in the first century referring to a substance, derived from phlegm that flows in the body. It was believed to originate in the brain and flowed

to various parts of the body causing ailments (West, 2014). The symptoms of RA include pain, swelling, and stiffness of the joint, while in serious cases the inflammation affects the bone, cartilage, and other tissues within the joint (Cush et al., 2010).

RA is caused by a combination of genetic and environmental factors; however, in many cases the precise etiology is unknown. Variations in several genes have been studied as a risk factor for RA (Genetics Home Reference, 2013). Most of the genes implicated are known or thought to be involved in immune system function.

1.3 History of Rheumatoid Arthritis

RA was first described by Augustin Jacob Landré-Beauvais in 1800 at an asylum in France. He noticed the condition in poor patients, and suggested that the disease is gender-related as it affected women more than men. He thought that it was related to gout, which was then associated mostly with affluent patients. Thus, he called it Primary Asthenic Gout. However, his explanation of RA as gout, did not satisfy researchers in the field (Cush et al., 2010; Shlotzhauer et al., 2014). Sir Alfred Baring Garrod, an English physician during the mid to late 19th century, was the first person who differentiated gout from other arthritic conditions. His discovery was that patients with gout have uric acid in their blood, which was not the case in patients with arthritic conditions. In 1859, Sir Alfred published his classification of RA as a distinct disorder in a manuscript titled “Treatise on Nature of Gout and Rheumatic Gout” (Cush et al., 2010).

Fifty years later, Sir Archibald Edward Garrod, the fourth son of Sir Alfred Baring Garrod, wrote about his father's discovery in a book titled "A Treatise on Rheumatism and Rheumatoid Arthritis". He was the first to use the term "Rheumatoid Arthritis" when referring to this disease. He demonstrated that RA appears during a period that was marked by increase in pollution, dietary changes, and lifestyle transformations. In his book, he suggested that RA is a problem inherited from one's ancestors. He concluded this by studying skeletons found in some graves. Unfortunately, his book lacked profound evidence (Cush et al., 2010; Shlotzhauer et al., 2014). In Native American skeletons of the Mississippi River valley, RA had been identified as far back as 4500 BC. Some researchers demonstrated that RA originated as an infection in a small group of Native Americans; then it spread worldwide (Koehn et al., 2002).

In 1912, Frank Billings suggested that RA was a response to chronic infection. Scientists tried to find a link between different types of infections and the development of RA. By the end of the 1920s, the culprit was thought to be tuberculosis and some other bacteria like *Streptococcus* and *Mycoplasma*. Later, medical research focused on viruses, rubella, and parvoviruses.

In the 20th century, some researchers suggested that bacteria that inhabit our mouth and digestive tracts could cause an abnormal response of the immune system which may lead to the development of RA (Boekel et al., 2001). In the mid-20th century, investigators began using X-rays as a popular technique for comparing skeleta of living patients. Their observations were dependent on the nature of soft tissue found in a living patient. Therefore, it was suggested that the signs symptomatic of RA

damage seen on dry skeletons may not have been produced by RA (Cush et al., 2010). In 1940, Norwegian Eric Waaler discovered the Rheumatoid Factor (RF) autoantibodies (Brink et al., 2016; Cantagrel & Degboe, 2016). RFs are autoantibodies against the Fc domain of immunoglobulin G (IgG). For over half a century, RFs have been used as a key marker of RA. 1.4 Genetic History of Rheumatoid Arthritis

A large body of evidence suggests that there is an association between human leukocyte antigen (HLA) alleles and RA (Cantagrel & Degboe, 2016; Cush et al., 2010; Weisman, 2011). It is considered to be the strongest genetic risk factor for the disease (Drongelen & Holoshitz, 2017). In 1977, Andrew J. McMichael and colleagues reported that there was an increased frequency of occurrence of Human Leukocyte Antigen (HLA)-Cw3 and HLA-Dw4 proteins in RA patients (Choy, 2012). RJ. Winchester from the New York Sinai School of Medicine studied populations of patients with RA and families with multiple cases of RA, and provided definitive evidence that susceptibility to the disease is determined both by genetic factors associated with alleles of the major histocompatibility complex (MHC) as well as environmental factors. The development of seropositive, but not seronegative RA is associated with the presence in an individual of the HLA alloantigen HLA-DR4. A lot of genetic aspects of RA has been studied and is still under active investigation and research (Al-Mughales, 2015).

1.4 Common Types of Arthritis

There are over hundred types of arthritis that can cause weakness, pain, and life-difficulties. According to Arthritis Foundation, the most common types of arthritis

that affect patient's activity are osteoarthritis (OA), rheumatoid arthritis (RA), psoriatic arthritis (PsA), lupus, and gout. The difference between RA and OA is the nature of the disease (Freeman, 2018). RA is an autoimmune disease where joints and sometimes other organs are attacked by an individual's immune system, while OA is a degenerative joint disorder that results in wear and tear of joints. OA is the most common chronic condition of the joints that occurs when the cartilage between joints starts to collapse. There are several factors that causes OA including genetic variations, obesity, injuries, metabolic disorders as well as joint diseases such as RA. On the other hand, the causes of RA are not fully understood. The reason for the sudden abnormal response of the immune system to body-tissues is yet to be clearly elucidated. However, it has been reported that hormones, environmental factors and genes are the leading causes of RA. The symptoms of OA and RA are quite similar which confuses patients and physicians. Based on the Arthritis Foundation, symptoms include joint pain, swelling or stiffness, joint tenderness and fatigue particularly occurring in the morning, and limited range of motions. Gout is a type of arthritis that causes inflammation, usually in one joint, which begins suddenly. Gouty arthritis is caused by the deposition of crystals of uric acid in a joint (Driver, 2019). Accordance to the Lupus Foundation of America, lupus is considered as a type of arthritis since many symptoms of arthritis are also present. Lupus is an autoimmune disease in which the immune system is very active. A long-period of inflammation leads to uncontrolled lupus or "lupus arthritis". PsA is a form of arthritis which is developed by people with psoriasis (Mayo Clinic, 2018). PsA is believed to be caused by genes or infections and sometimes by other factors such as stress and injury. Symptoms of PsA include joint

pain, stiffness, skin problems, eyes problems, changes in nail color and shape and swelling in fingers and feet (Freeman, 2018).

1.5 Pathophysiology of RA

RA is one of the most prevalent forms of autoimmune arthritis that is characterized by inflammation of the joint lining the synovium. Synovial joints strengthen and support joints, including the cartilage, tendon, ligament, and muscle. Each of these structures are made up of different kind of tissues, and these tissues are made up of specialized cells (Figure 1) (Shlotzhauer et al., 2014). RA is a gradual inflammatory syndrome characterized by proliferation of the synovial membrane and constant uncontrolled inflammation resulting in chronic destructive polyarthritis (Hoxha, 2018). In RA the immune system does not recognize the body's tissue and attacks it as if it was an invader. Subsequently, a cascade of immune reactions results in inflammation, thickening of synovial lining, angiogenesis and cartilage destruction (Brink et al., 2016; Yamamoto et al., 2015).

1.6 Autoimmune Response

The immune system is a complicated mechanism that enables the body to defend itself against invaders. The production of autoantibodies against specific or several self–antigens in the body is the hallmark of autoimmune disorders. The cause of such antibodies production remains unclear. RA has been classified as an autoimmune disease for nearly a century since the identification of RFs. RA is an inflammation of the synovium, caused by an interaction between fibroblast-like

synoviocytes and cells of the innate immune system (Gierut et al., 2017). The function of the innate immune system is to recruit immune cells to an infection site by producing chemical factors like cytokines followed by an activation of a cascade of several components of the immune system (Janeway et al., 2008).

40-65% of RA patients are seropositive and around 20% are seronegative. In seropositive patients, blood tests show the presence of antibodies that attack the body and lead to joints inflammation. Antibodies elevated in RA patients include RFs, anti-citrullinated protein antibodies (ACPAs) or both (Nakken et al., 2017; Smolen et al., 2016; Verheul et al., 2015).

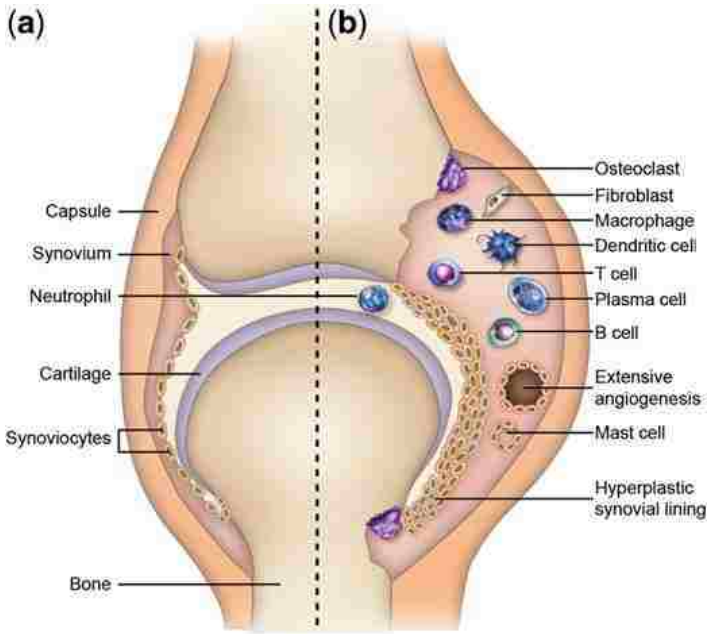

Figure 1: Schematic view of a normal joint (a) and a joint affected by RA (b). The joint affected by RA (b) shows increased inflammation and cellular activity (Choy, 2012).

1.7 Autoantibodies Associated with RA

Over the past 25 years, an abundance of autoantibodies has been associated with RA and its activity. The discovery of these autoantibodies continue without reaching a comprehensive idea about its association with the disease (Choy, 2012). RF was an early autoantibody linked with RA that is present in about 75% of RA seropositive patients. Anti-RA33 antibodies was found in 36% of RA sera while anti-calpastain was found in about 45% of RA sera (Brink et al., 2016; Nakken et al., 2017; Boekel et al., 2001; Verheul et al., 2015). Sa protein or citrullinated vimentin, was present in about 40% of the RA sera with high specificity (Brink et al., 2016; Nakken et al., 2017; Boekel et al., 2001; Verheul et al., 2015). Anti-citrullinated protein antibodies (ACPAs) are early markers for RA that are detectable before the onset of the clinical symptoms. ACPAs are the most specific autoantibodies used as RA markers. The role of ACPAs in RA, however, has not been clearly established. ACPAs are associated with the shared epitope (SE) (Figures 2 and 3) of HLA-DR1 and generated via T-helper cell and B-helper cell. Heavy chain binding protein (p68) and glucose-6-phosphate isomerase are also found in RA patients (Brink et al., 2016; Holers, 2013; Nakken et al., 2017; Boekel et al., 2001; Woude & Catrina, 2015; Verheul et al., 2015).

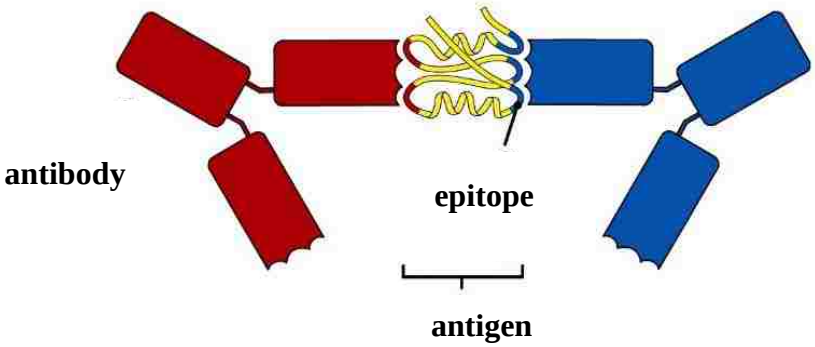

Figure 2: Antigens are the molecules recognized by the immune response, while epitopes are sites within antigens to which antigen receptors bind (Janeway et al., 2008).

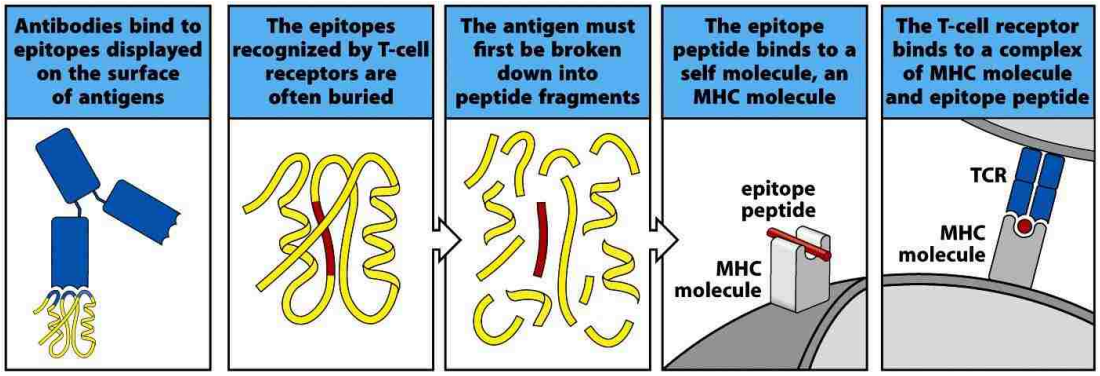

Figure 3: An antibody binds an antigen directly, whereas a T-cell receptor binds a complex of antigen fragment and self-molecule (Janeway et al., 2008).

1.8 Genetic Association of RA

Both genetic and environmental factors play a role in RA. Genetic factors play a strong role in RA progression and severity. Variations in several genes have been studied as risk factors for RA. Most of these genes are known or believed to be involved in immune system and forms part of the major histocompatibility complex

(MHC) or specifically human leukocyte antigen (HLA) in humans (Firestein & McInnes, 2017). It has been established that the most common components of HLA class II such as HLA-DRB1 is involved in several autoimmune disease including type 1 diabetes, Graves' disease and RA (Gough et al., 2007). Previous studies have demonstrated that RA progression and severity is associated with HLA-DRB1 polymorphisms (Weyand et al., 2000; Barnetche et al., 2002).

1.9 Major Histocompatibility Complex (MHC)

MHC is a gene family that exists in many species**.** It consists of more than 200 genes. In humans, the MHC genes cluster on chromosome 6. The cytogenetic location of MHC in human is on the short arm within 6p21.3 and spans around 4,000 kilobases (Mizuki & Kimura, 1996). HLA is the human form of MHC. MHC encodes a set of membrane glycoproteins called the MHC molecules. Genes in this complex are categorized into three main groups: class I, class II, and class III (Genetics Home Reference, 2009). The genes encoding the α chains of MHC class I molecules and the α & β chains of MHC class II molecules are linked within the complex (Figure 4) (Janeway et al., 1996). MHC class I proteins are found in all cells while MHC class II proteins are found in phagocytes. T helper cells recognize antigen in association with MHC II and T cytotoxic cells recognize antigen in association with MHC I (Janeway et al., 2008). The figure below shows the relative position of MHC class I and class II genes on the short arm of chromosome 6 at position 6p21.3 along with specific exons within these gene with known variations associated with RA (Figure 5) (Warren et al., 2012).

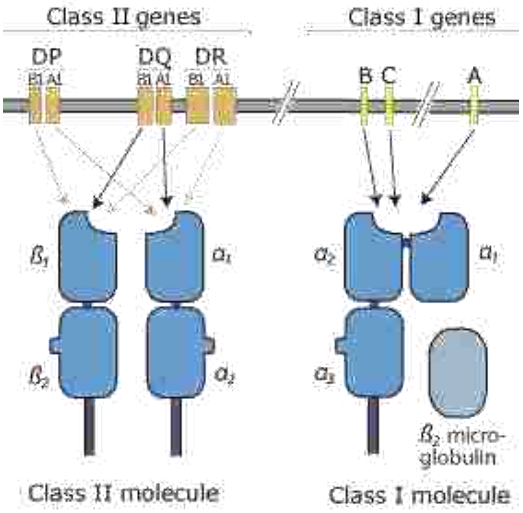

Figure 4: The structure of an MHC class I molecule (Transplantation Immunology, 2015).

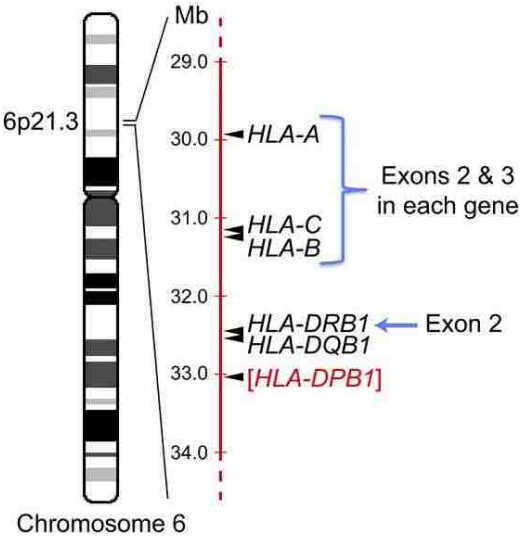

Figure 5: The relative position of the HLA genes on the short arm of chromosome 6 and the exons to be considered (Warren et al., 2012).

1.10 Class I MHC/HLA Genes

MHC I is made of an α chain (43 kDa) and small peptide called β2-

macroglobulin (12 kDa). The extracellular domains show variability in their amino acid sequences, resulting in grooves with different shapes (Figure 6). MHC I binds to short peptides of 8-10 amino acid and T cytotoxic cells recognize antigen associated with MHC I (Janeway et al., 2008).

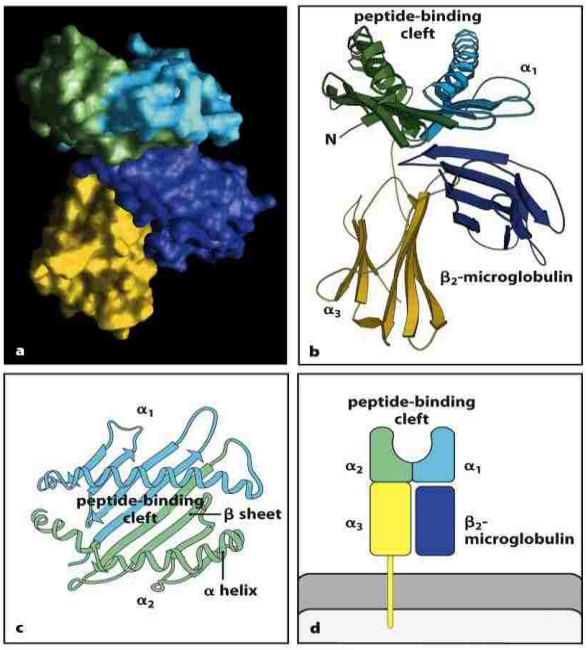

Figure 6: The structure of an MHC class I molecule (Janeway et al., 2008)

In humans, there are three major loci that code for MHC I molecules (HLA-A, HLA-B and HLA-C). Within a locus, there are different allelic forms of the molecules. For example, HLA-A2 and HLA-Aw68 class I molecules differ from one another at 13 amino acid residues. These distinctions in allelic forms of HLA molecules results in dramatic differences in the shape of the groove of the molecule, where foreign peptides are accommodated. Each individual can express up to two types of HLA-A alleles, one each inherited from a set of parents. Some individuals will inherit the same

HLA allele from both parents that thereby decreasing the individual's HLA diversity. The MHC is polymorphic and exhibits codominant expression. Thus, most individuals are likely to be heterozygous at each locus. Because the human population is extensively outbred, the MHC genetics in humans is extremely complex (Figure 7) (Janeway et al., 2008).

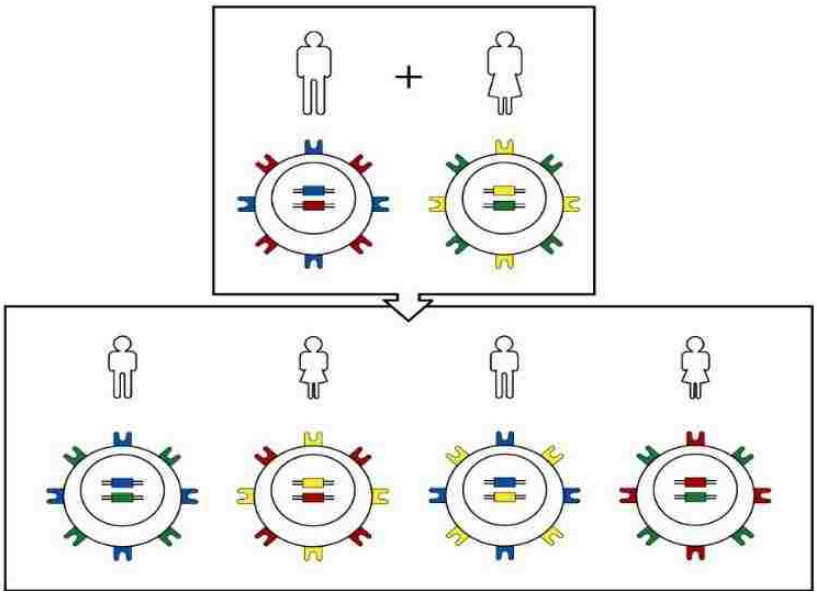

Figure 7: Expression of MHC alleles is codominant (Janeway et al., 2008).

1.11 Class II MHC/HLA Genes

The class II genes of MHC provides instructions for making proteins present on the surface of certain immune system cells (Janeway et al., 1996). The genetic locus encoding class II molecules is known as the D region in humans. The D region is further subdivided into several groups. There are seven super types of class II HLA-D genes (main DR, DR4, DRB3, main DQ, DQ7, main DP and DP2) (Greenbaum et al., 2011). Like the HLA class I genes, HLA-D class II genes also exhibit polymorphism

and codominant expression. MHC class II is always expressed by B lymphocytes and dendritic cells. It is composed of an α chain (30-34 kDa) and a β chain (26-29 kDa). In MHC class II, the extracellular domains show variability in their amino acid sequences resulting in grooves with different shapes (Figure 8). The length of the peptides bound is not constrained. It could be at least 13 amino acid long and can be much longer. T helper cells recognize antigen in association with MHC II (Janeway et al., 2008; Janeway et al., 1996). In plentiful numbers of individuals, HLA-DR cluster consist of an additional β chain gene and its product can pair with the DRα chain (Janeway et al., 1996; Scrivo et al., 2007).

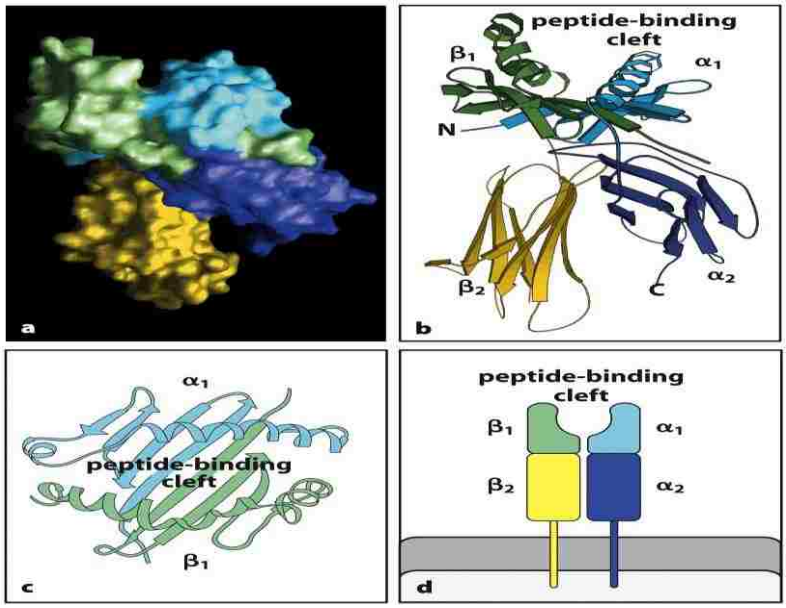

Figure 8: The structure of an MHC class II molecule (Janeway et al., 2008).

1.12 Class III HLA Genes

Class III HLA genes encode complement components that show no structural similarity to ether class I or class II molecules. MHC class III region resides between

the MHC class I and class II regions (Wennerstrom et al., 2013). These genes, along with genes encoding tumor necrosis factor (TNF), separate HLA class II and HLA class I genes on chromosome 6 in humans (Figure 9) (Matzaraki et al., 2017).

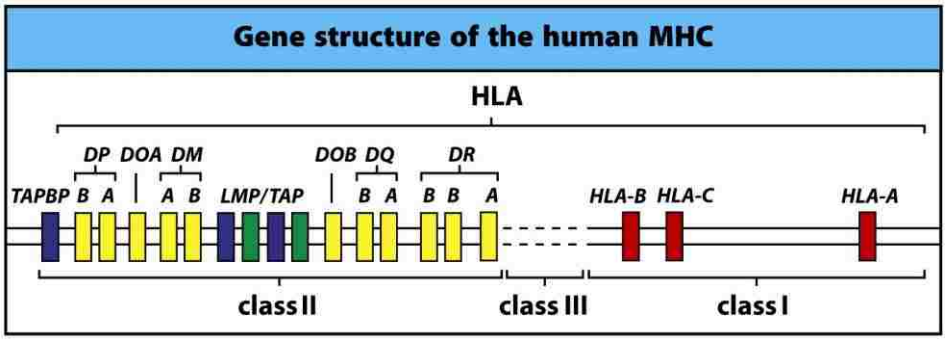

Figure 9: The genetic organization of the major histocompatibility complex (MHC) in human (Janeway et al., 2008).

1.13 MHC Functions and Pathways

The function of MHC is to bind to peptide fragments from infectious agent and present them on the cell surface. Appropriate T cells can then recognize the infected cell and destroy it. This mechanism obstructs pathogens from dodging the immune system. Class I and class II of MHC molecules binds peptides to T cells, however, each type has its own pathway for binding peptides (Figure 10) (Janeway et al., 2008; Kobayashi & Elsen, 2012).

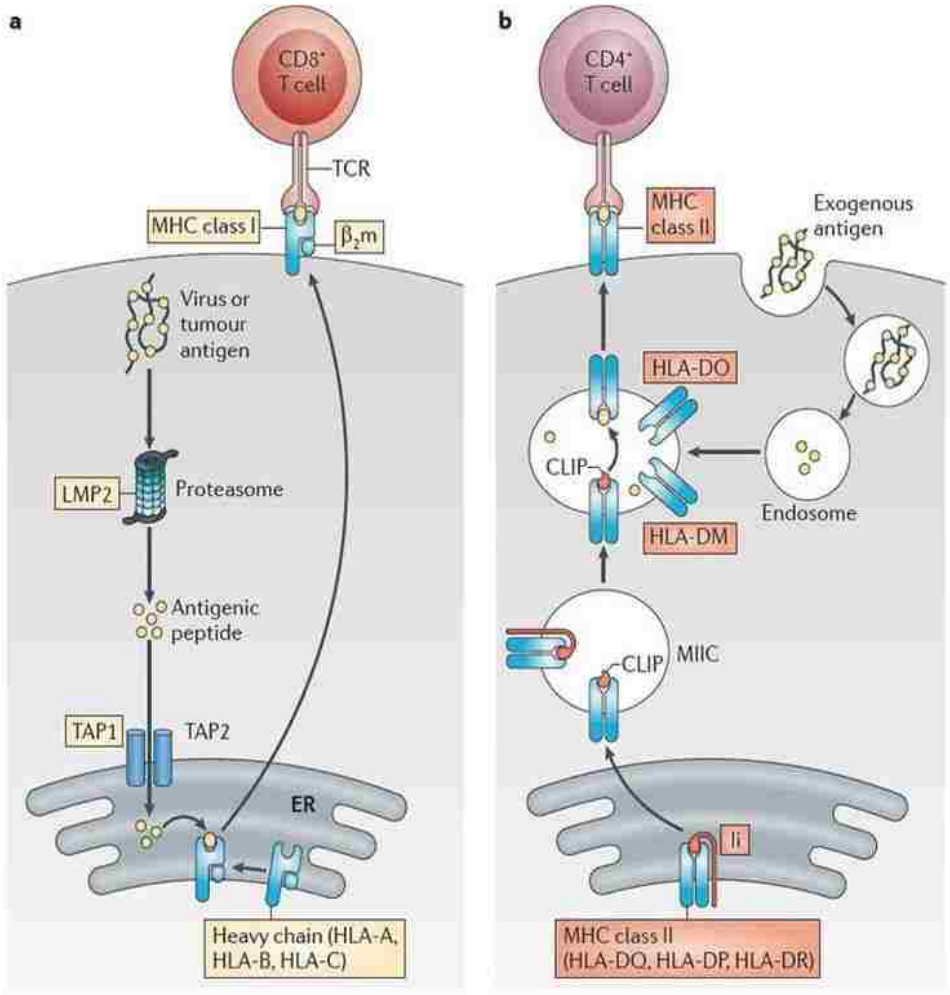

Figure 10: The (a) MHC class I and (b) MHC class II antigen-presentation pathway (Kobayashi & Elsen, 2012).

1.14 Nomenclature and HLA Classifications

HLA started as a list of antigens identified because of transplant rejections (Ka et al., 2017; Thorsby, 2009). The antigens were initially identified by massive statistical analysis of the interactions between blood types. This operation is in accordance with the principle of serotypes. HLA genes of the host consists of different antigens from those found on the surface of infectious agents. HLAs are alloantigen,

which are responsible for binding to non-self-antigens from member of the same species. In another word, the alloantigen is present only in some, but not all, members of a species and it stimulates the production of alloantibody by individuals lacking that antigen in grafting and organ transplants or even in blood group. This determines the acceptances or rejection of the organs and blood between individuals (Janeway et al., 2008). HLA vary between individuals due to genetic differences. One of the most complex regions of the human genome is the HLA locus. In spite of the fact that it represents only 0.3% of human genome, it constitutes 1.5% of genes in OMIM, and 6.4% of SNPs in the genome are based in this region (Ka et al., 2017; Thorsby, 2009). Many different versions or alleles of HLA genes have been identified by researchers. Each allele is assigned a specific notation. The most recent HLA naming system was developed in 2010 by the WHO Committee for Factors of the HLA System. They developed a systematic naming convention for identifying the variations in HLA alleles. There are two types of MHCs, Class I and Class II MHCs are named with the same system. Alleles that started with "HLA" represent the human MHC genes. The next part (HLA-A or HLA-B) identifies which gene the allele is a modification of. Each HLA allele has a unique number that consists of four set of digits separated by colons. The digits before the colon represents the type that is associated with the serological antigen supported by an allotype, the allele of the antibody chains found in the individual. The digits after the colon is used to list the subtypes of the DNA sequences that have been identified. For example, in HLA-DRB1*04, the HLA is a prefix, DRB1 is the gene followed by a separator and 04 is the allele group. Sometimes four digits follow, for example HLA-DRB1*04:01, in which case 01 is a specific HLA

protein. The third group of digits, for example 02 in HLA-DRB1*04:01:02, refer to allele variants which produce same protein. The fourth set of digits are used for non-coding region of the gene (Figure 11,12). The final aspect of HLA naming is a letter used to denote changes in expression. There are six letters, each with a different meaning (Table 1) (Marsh & System, 2011; Scally et al., 2013).

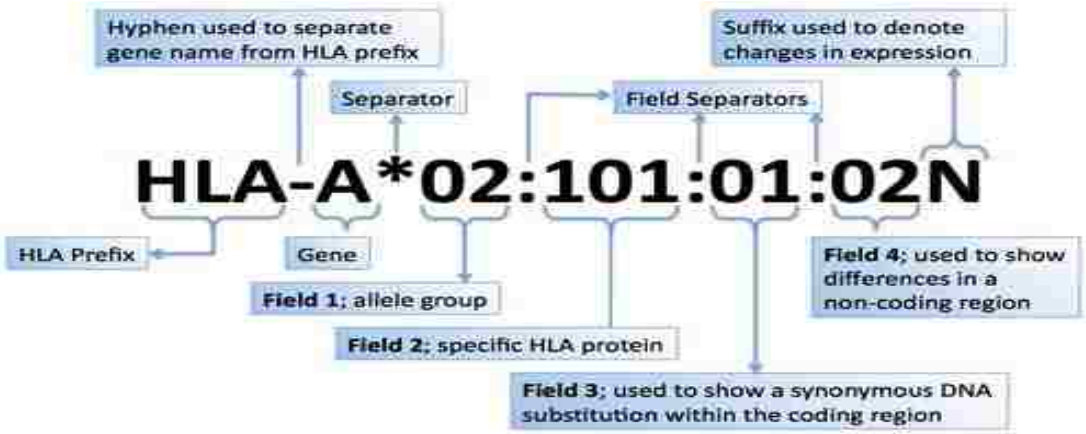

Figure 11: Interpretation of digits in HLA genes (Marsh & System, 2011).

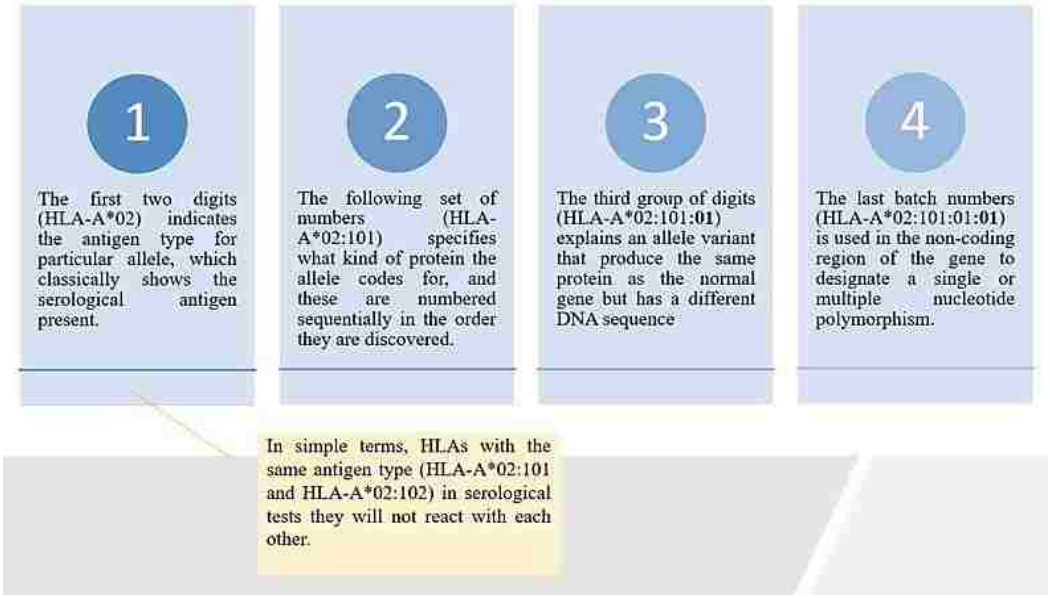

Figure 12: Nomenclature of HLA Alleles.

Table 1: The final aspect of HLA naming

Letter	Significance
N	Null Allele (Presents a non-functional protein)
L	Lower than normal cell surface expression
S	Soluble protein absent from the cell surface
Q	Questionable (allele may affect the expression)
C	Protein is found in cytoplasm and absent on cell surface
A	Aberrant or irregular expression

1.15 Association of HLA Genes with RA

The inheritance pattern of RA is unclear due to the involvement of numerous genetic and environmental factors. The HLA loci have been identified as the genes most associated with RA (Kochi et al., 2014; Viatte et al., 2015; Weyand et al., 2000). Several allelic variants of HLA gene have been associated with RA. Studies have indicated that RA is associated primarily with HLA-DRB1 alleles, especially 'HLA-DRB1*04' allele. HLA-DR is a cell surface receptor that provides information to make protein functions in immune system response. It provides instructions to present peptide antigens to the immune system to elicit or suppress T-helper cell responses (Janeway et al., 2008; Ruyssen-Witrand et al., 2015; Scally et al., 2013; Viatte et al., 2015; Wennerström et al., 2013). The association of the HLA-DRB1 with RA is strongly associated with conserved sequences at amino acid positions 70-74 in the protein referred to as the shared epitope (SE) (Ruyssen-Witrand et al., 2015; Scally et al., 2013). This conserved region with the sequence QKRAA/QRRAA/RRRAA,

which involves alleles HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*04:04, HLA-DRB1*10:01, have been strongly associated with RA (Ruyssen-Witrand et al., 2015). Additional classifications of HLA-DRB1 associated with RA has expanded this to the presence of an aspartic acid 'D' at position 70 and the DERAA SE motif between positions 70-74. Further, it has been demonstrated that the amino acid at position 11 and 13 also increases the risk of RA (Ruyssen-Witrand et al., 2015). The presence of Val11 in HLA-DRB1 has been reported to be strongly associated with seropositive RA, while the amino acids most associated with seronegative RA is HLA-DRB1 were Ser11 and Leu11 and the allele type HLA-DRB1*03. It has also been shown that HLA-A Asn77 is also significantly associated with seropositive RA (Figure 13) (Han et al., 2014).

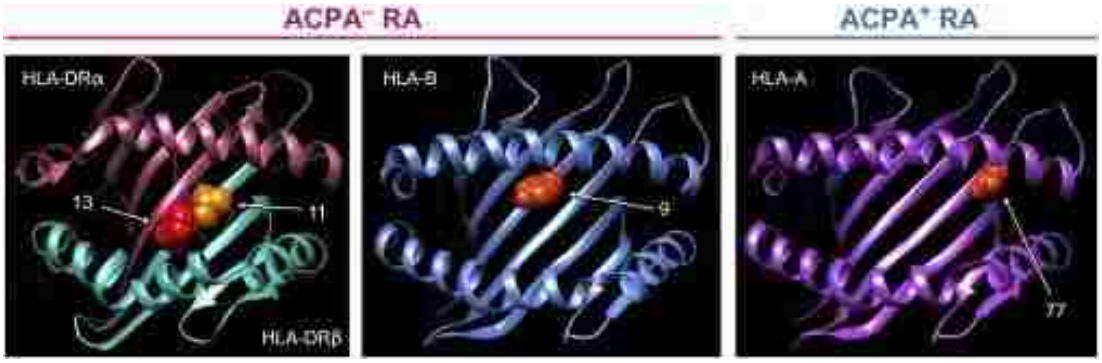

Figure 13: 3D models of HLA proteins with amino acid positions that are significantly associated with RA (Han et al., 2014).

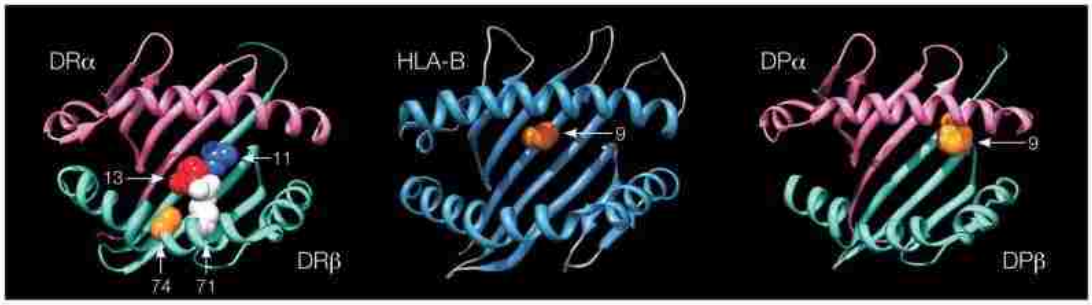

Figure 14: 3D models of HLA proteins with additional amino acid positions that are known to be associated with RA (Raychaudhuri et al., 2012).

Furthermore, three additional positions in HLA-DRB1 - His13, Lys71, and Ala74 - and single amino acid polymorphism in HLA-B - Asp9 - and HLA-DP1 - Phe9 – were associated with an increased risks of seropositive RA (Figure14) (Greenbaum et al., 2011; Raychaudhuri et al., 2012).

It was reported that in Black Sea Turkish population, the frequency of HLA-DRB1*01, *04, and *09 alleles are higher in RA patients and the conserved SE at position 70-74 is commonly QKRRA (Uçar et al., 2012). In African Americans population, a Val11 variation, corresponding to allele *09:01, was observed to impart high RA risk, which was also the same allele for RA in the Korean population. A recent study established that in European ancestry, HLA-DRB1*04:01, *04:04, and *01:01 are the most frequent RA risk alleles, whereas in East Asian populations, HLA-DRB1*04:05 and *01:01 were the most frequent RA risk alleles (Yamamoto et al., 2015). In addition, several subtypes of the DR4 allele, such as *04:02 and *04:03 appear to confer protection against the disease (Yamamoto et al., 2015). The protective effects of DRB1 allele, DRB1*13:02, in Japanese RA has also been reported (Oka et

al., 2014). A study involving the typing of HLA-DRB1 in Saudi Arabian RA patients found that HLA-DRB1*04 followed by HLA-DRB1*08 and HLA-DRB1*10 are alleles that could be used as predictors for RA. Additionally, the frequency of HLA-DRB1*06 was observed to be lower in RA compared to control (Al-Swailem et al., 2006). Table 2 provides a summary of HLA-DRB1 alleles that have been reported to be associated with RA in various populations.

Table 2: Distribution of HLA-DRB1 alleles and genotypes frequencies in RA among different population.

Population	HLA-DRB1 allele			Reference
Japanese	*04	*09:01	*15	(Al-Swailem et al., 2006; Oka et al., 2014; Shimane et al., 2013)
African American	*09:01			(Reynolds et al., 2014)
Korean	*09:01	*04:05		(Al-Swailem et al., 2006; Reynolds et al., 2014)
European	*01:01	*04:01	*04:04	(Reynolds et al., 2014)
Turkish	*01	*04	*09	(Uçar et al., 2012)
Saudi	*01	*04	*08	(Al-Swailem et al., 2006)
French	*01			(Uçar et al., 2012)
Latin Americans	*01			(Uçar et al., 2012)
Asians	*01	*04:05		(Uçar et al., 2012; Yamamoto et al., 2015)
Caucasians	*01	*09		(Uçar et al., 2012)
Chileans	*09			(Uçar et al., 2012)
Syrian	*01:01	*04:04	*04:05	(Saxena et al., 2017)

1.16 RA in United Arab Emirates

Over 20% of UAE's population suffer from at least one type of arthritis (Carmona et al., 2002). RA is the most common inflammatory arthritis (Drongelen & Holoshitz, 2017). In UAE, RA is widespread with delayed diagnosis. Dubai Bone and Joint Center classifies patients with RA by the symptoms and their ages. Only individuals who are 16 and above and have RA symptoms are categorized as RA patients (Badsha et al., 2008). Additionally, more females were diagnosed with RA than males (Badsha et al., 2008; Saxena et al., 2017). A recent study evaluated the factors causing RA in Arab populations. RA patients from Jordan, the Kingdom of Saudi Arabia, Lebanon, Qatar, and the United Arab Emiratis were examined to check the genetics and clinical features of the disease. The study concluded that HLA-DRB1 amino acid position 11 was strongly associated with seropositive RA in Arabs (Saxena et al., 2017).

1.17 Whatman FTA Technology

Flinders Technology Associates (FTA), or now Whatman FTA Card, is a filter paper based system manufactured by GE Health Care that provides a safe and reliable technique for storage of DNA at the room temperature for several years (Ndunguru et al., 2005). It has been reported by the manufacturers that DNA remains stable for at least 17 years or more. As proof, researchers were able to generate short tandem repeat (STR) data with good signal strength from DNA from 22-years-old blood (Mullen et al., 2009; Whatman FTA Technology). The FTA card is a special type of filter paper (Figure 15A) which consist of two types of reagents – one that regulate cell lysis and

protein denaturation and the other that protect DNA from denaturation (Entezami et al., 2011; Santos, 2018). Different kind of samples (blood, bacteria, buccal cells, cultured cells, plant material, and more) can be applied either directly or with a swab to FTA cards (Whatman). The cards are white, pink, or purple color. Pink and purple include a dye which turns white when colorless samples are applied (Figure 15B) (Whatman FTA Technology). FTA technology is widely for human DNA collection, archiving of nucleic acids for forensic applications and PCR-based genotyping (Ndunguru et al., 2005). After drying the sample, a small disc from the card is removed using a puncher to recover nucleic acids. The discs are washed with FTA reagents, and can be directly used for amplification by PCR. Since FTA cards require only a few drops of blood, which could significantly reduce the trauma associated with blood collection especially in children, as well as reduce the chance of infection, this study decided to investigate FTA cards as a means of collecting and extracting DNA for routine testing.

A **B**

Figure 15: (A) Whatman FTA cards (B) FTA card that turns white when colorless samples are loaded (Avantor; Cole-Parmer Canada Company).

1.18 Hypothesis

The HLA loci have been established as a strong risk factor associated with RA. Specifically, the SE region of HLA-DRB1 has been documented as the strongest indicator. Additionally, other variations at position 11 and 13 of HLA-DRB1 and position 77 of HLA-A are also very good risk factors. The hypothesis of this study was that these variations could be observed in RA patients in the UAE.

1.20 Objective of the Study

The main aim of this study was to understand the variation of HLA genes in RA patients from Fujairah Hospital in UAE. Identification of variations in HLA alleles could help in understanding the possible genetic cause of RA in this region which could further be expanded to the population of UAE at a later stage. Specifically, this would involve identifying the variations in the genes HLA-A from class I of MHC and HLA-DRB1 from class II of MHC, which are two genes that harbor variations associated with RA. Blood samples were collected and loaded on Whatman FTA card and exon 2 and 3 of HLA-A and exon 2 of HLA-DRB1 would then be amplified using polymerase chain reaction (PCR) and sequenced to compute the incidence of the known variants associated with RA.

Chapter 2: Methods

2.1 Ethical Approval

Ethical approval was obtained from the Center of Research and Statistics, Ministry of Health and Prevention (MOHAP) to collect samples from Fujairah Hospital. Ethical approval no: MOHAP /DXB-REC-SUB-30/2017, was granted on 23 August 2017 and renewed on 10 December 2018 up to 9 December 2019.

2.2 Patients and Controls

To be eligible for the study, patients had to be Emirati, should have been clinically diagnosed with RA, be at least ≥ 18 years of age, and be able to sign an informed consent. Healthy controls included individuals who had not shown any clinical symptoms of RA. A total of 54 RA patient samples and 44 healthy control samples were analyzed in this study. Subjects were enrolled from Fujairah Hospital during the period 25 August 2017 to March 2018.

2.3 Sample Collection

2 ml of whole blood was collected from patients and controls by venipuncture into vacuum collection tubes containing dipotassium EDTA (Ethylene Diamine Acetate). One drop (around 50 µl) was applied to a Whatman™ FTA Micro Card (Fig 16). Cards were dried at room temperature, transported to Department of Biology, United Arab Emirates University, for further analysis.

Figure 16: FTA card spotted with human whole blood.

2.4 Storage Condition

Whatman™ FTA cards that contained blood samples were stored at room temperature in a safe cabinet at UAE University with coded numbers. Original blood samples collected in EDTA tubes were stored at -50°C in a safe room at Fujairah Hospital.

2.5 DNA Extraction

FTA cards that contained blood samples were punched twice using Harris® Uni-Core™ 1.2 mm puncher (Figure 17) and placed in a PCR tube for washing. To avoid cross contamination between samples, an equivalent number of discs were taken from a blank FTA cards after each sample. The two small punched discs were washed

for 45 minutes in 200 µl of Whatman FTA purification reagent. The discs were then washed for 10 minutes in 200 µl of 1 mM TE buffer. The washed discs were left to dry at room temperature for up to 30 minutes. The two washed discs were then used for PCR. Using FTA card for PCR does not require DNA extraction or quantification since the amount of DNA in the disc is sufficient for PCR .

Figure 17: Harris® Uni-Core™ 1.2 mm puncher.

2.6 Primers

The sequence of HLA-DRB1 gene and HLA-A gene were retrieved from Ensemble genome browser. The sequences were further cross checked with the sequence obtained from National Center for Biotechnology Information (NCBI) database. The coding exons 2 and 3 were marked and primers were designed that covered this region. Melting temperature and GC content of each primer were also calculated. Designed primers were purchased from Macrogen (Korea). The optimum conditions for the primers were then determined experimentally using multiple PCR test runs. The list of primers and their optimal conditions are given in Table 3.

Table 3: Primers set used in PCR to amplify different HLA genes.

Gene	**Forward Primer**	**Reverse Primer**	**Annealing Temperature (**℃)	**Annealing Time (sec)**
HLA-DRB1 (exon 2)	5’- GATGGTGGCGTCGCTGTC-3’	5’-GCTCTCAGAACTGCTTGCT-3’	59	35
	5’-GAAGACGGAGGATGAGCTCC-3’	5’-GAACTGCTTGCTCCGGACTG-3’	59	35
HLA-A (exon 2)	5’-TTGGGTGTCGGGTTTCCAGAGA-3’	5’-CTCTCCCGGGACAAGGGTCTC-3’	66	40
	5’-TCGTCGCGGTCGCTGTTCTA-3’	5’-CTTCGGGGTGGATCTCGGACC-3’	66	40
HLA-A (exon 3)	5'-GAGACCCTTGTCCCGGGAGAG-3'	5'-GTCAGAGAGCAGGGCGGAACC-3'	67	40
	5'-CAGGAGACACGGAATGTGAAG-3'	5'-CAATTGTCTCCCCTCCTTGTG-3'	61	35

2.7 PCR amplification

The Go Taq® Green Master Mix (Promega, USA), which is a premixed ready-to-use solution, was used in this study. Go Taq® Green Master Mix contains Taq DNA polymerase, dNTPs, $MgCl_2$ and reaction buffers at optimal concentrations. A reaction volume of 50 µl was used for each PCR run. Two FTA discs were added to each PCR tube. The reaction conditions used in the Thermal Cycler (Figure18) were as follows: 1 cycle at 95ºC for 3 minutes followed by 34 cycles of denaturation at 95ºC for 30 seconds, annealing at temperature and time provided in Table 3, and finally extension at 72ºC for one minute. Amplified products were visualized on 1.5% (w/v) agarose gel using electrophoresis.

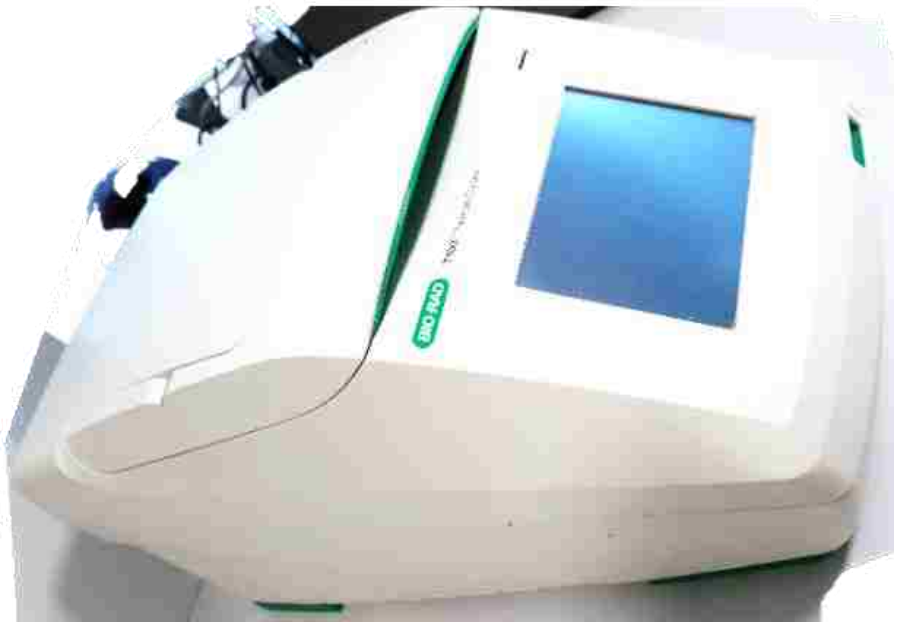

Figure 18: BioRad T100™ Thermal Cycler.

2.8 Gel Electrophoresis

1.5% (w/v) agarose CSL-AG500 (Cleaver Scientific) gel was prepared using 0.75 grams of agarose in 50 ml of 1 X Tris-Borate EDTA (TBE) buffer (0.89 M Tris-borate, 0.032 M EDTA, pH 8.3) and heated in a microwave for 1-2 minutes until it dissolved completely. The gel mixture was cooled to room temperature and 2 µl of ethidium bromide was added. When the gel had cooled adequately, it was poured into a gel tray. Once the gel solidified, 5 µl of 100 bp DNA ladder (N3231S-BioLab-ThermoFisher) was added followed by PCR products which were added immediately into wells without using a loading dye. The Go Taq® Green Master Mix used includes two dyes that allowed the monitoring of the progress of electrophoresis. Bio Rad PowerPac™ Basic machine was used for electrophoresis. The voltage and current were set at 120 V and 400 A, respectively, and the gel was allowed to run for 20 minutes. It was subsequently visualized and recorded using a Compact Digimage System, UVDI series gel documentation system.

2.9 Purification of PCR products

The amplified PCR products were purified using the Norgen PCR purification kit (Norgen Biotek, Canada). For this purpose, the binding buffer provided in the kit was added in 5:1 ratio to the PCR product. This was vortexed and pulse-spun briefly. The mixture was then transferred into a column provided in the kit and centrifuged for 1 minute at 8000 rpm. The flow-through was discarded and 500 µl of washing buffer (96-100% ethanol) was added to the column and centrifuged for 1 minute at 10,000 rpm. The flow-through was discarded again and the spun for 2 minutes at 14,000 rpm in order to dry the column. The column was then assembled with the provided 1.7 mL

elution tube. To increase DNA concentration 10 µl of elution buffer was added to the center of the column bed and kept standing at room temperature for 1 minute. The column was then centrifuged for 2 minute at 14,000 rpm. The last step was repeated once more. Quantity and quality of the purified DNA was evaluated by loading 1µl of purified DNA in NanoDrop 2000/c (Thermo Fisher Scientific, USA) which is a microvolume spectrophotometers.

2.10 DNA Sequencing

Samples that satisfied the required quality and quantity were sent to Macrogen, (Korea) for sequencing. The sequences and chromatograms were provided by Macrogen which were then analyzed further.

2.11 Analysis of Sequencing Data

FinchTV chromatogram viewer version 1.4.0 (Geospiza) was used to visualize the chromatograms. BioEdit Sequence Alignment Editor Version v7.0.5 (Tom Hall Ibis Therapeutics) and JalView (Waterhouse et al., 2009) were used to perform sequence alignment. Reference sequences for HLA-DRB1 (NCBI Accession: NG_029217) and HLA-A (NCBI Accession: NG_029921) were obtained from NCBI RefSeq database. Phylogenetic trees were generated using MEGA 6 (Kumar et al., 2016) using the Maximum Likelihood algorithm (Kumar et al., 2016).

2.12 HLA Allele Type Determination

A custom BLAST server containing all known HLA sequences obtained from HLA Nomenclature (http://hla.alleles.org) (Figure 19) was used for HLA allele

determination. Following the convention, for the class I gene, HLA-A, sequences of exons 2 and 3 were combined for allele type determination. For the class II gene, HLA DRB1, sequence of exon 2 was used for allele type determination using BLAST. HLA allele type frequencies were then computed.

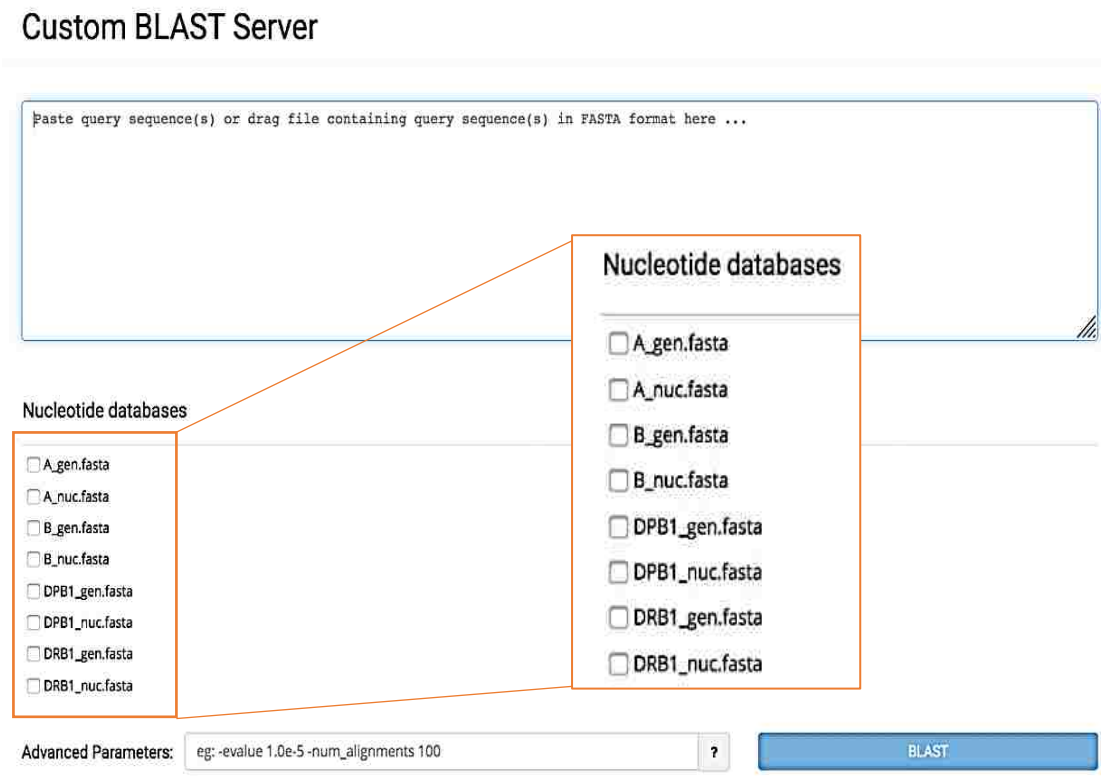

Figure 19: A custom BLAST server (http://compbio.uaeu.ac.ae/blast-hla) to search HLA sequences obtained from HLA Nomenclature (http://hla.alleles.org)

2.13 Homology Modeling

To locate the spatial orientation of the variations with higher incidence in patients, homology models of the mutant HLA proteins were generated using the online SWISSMODEL server (https://swissmodel.expasy.org/). The obtained models

were visualized in VMD 1.9.2 (Humphrey et al., 1996) and the Schrodinger Suite 2018-1 (Schrödinger, 2018).

2.14 Statistical Analysis

For computation of odds ratio for specific variations, Fisher's exact test was performed in R after generating contingency tables. To identify if there are genetic differences between patient and control groups, Fixation Index (F_{ST}) was computed using Arlequin 3.5.2 (Excoffier & Lischer, 2010). To identify if some or all of the patient and control samples cluster separately, Principal Component Analysis (PCA) was performed using aligned DNA sequences. The *adegenet* library was employed in R version 3.6.0 for performing PCA and plotting the principal components.

Chapter 3: Results

3.1 HLA Gene PCR Amplification

PCR was used to amplify genomic region encompassing exon 2 of HLA-DRB1 and exons 2 and exon 3 of HLA-A. These are regions essential for HLA typing and also known to harbor variations strongly associated with RA. Figures 20, 21 and 22 are sample gel electrophoresis images that shows bands of the expected size. HLA-DRB1 exon 2 amplicon has size 529 bp, HLA-A exon 2 is 705 bp, and HLA-A exon 3 is 598 bp. These samples were then purified and sent for DNA sequencing using the same primers used for PCR.

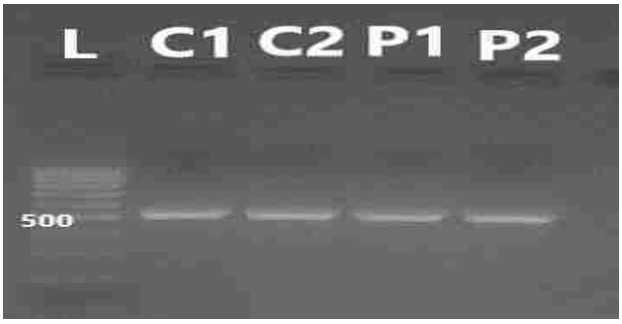

Figure 20: Gel electrophoresis image of PCR products of HLA-DRB1 exon 2.

Figure 21: Gel electrophoresis image of PCR products of HLA-A exon 2.

Figure 22: Gel electrophoresis image of PCR products of HLA-A exon 3.

Chromatograms obtained from sequencing were inspected using Finch TV Viewer (Figure 23) and the sequences were aligned using BioEdit Sequence Alignmen Editor and JalView (Waterhouse et al., 2009) to identify locations that harbor variations (Figures 24-26).

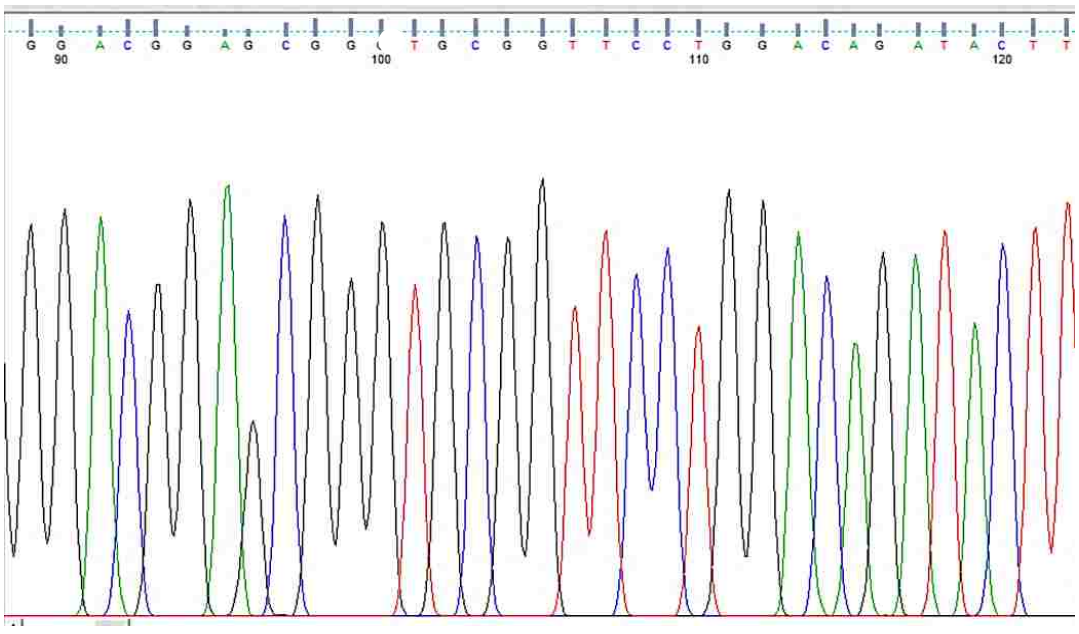

Figure 23: Chromatogram of an HLA-DRB1 sample visualized in FinchTV viewer.

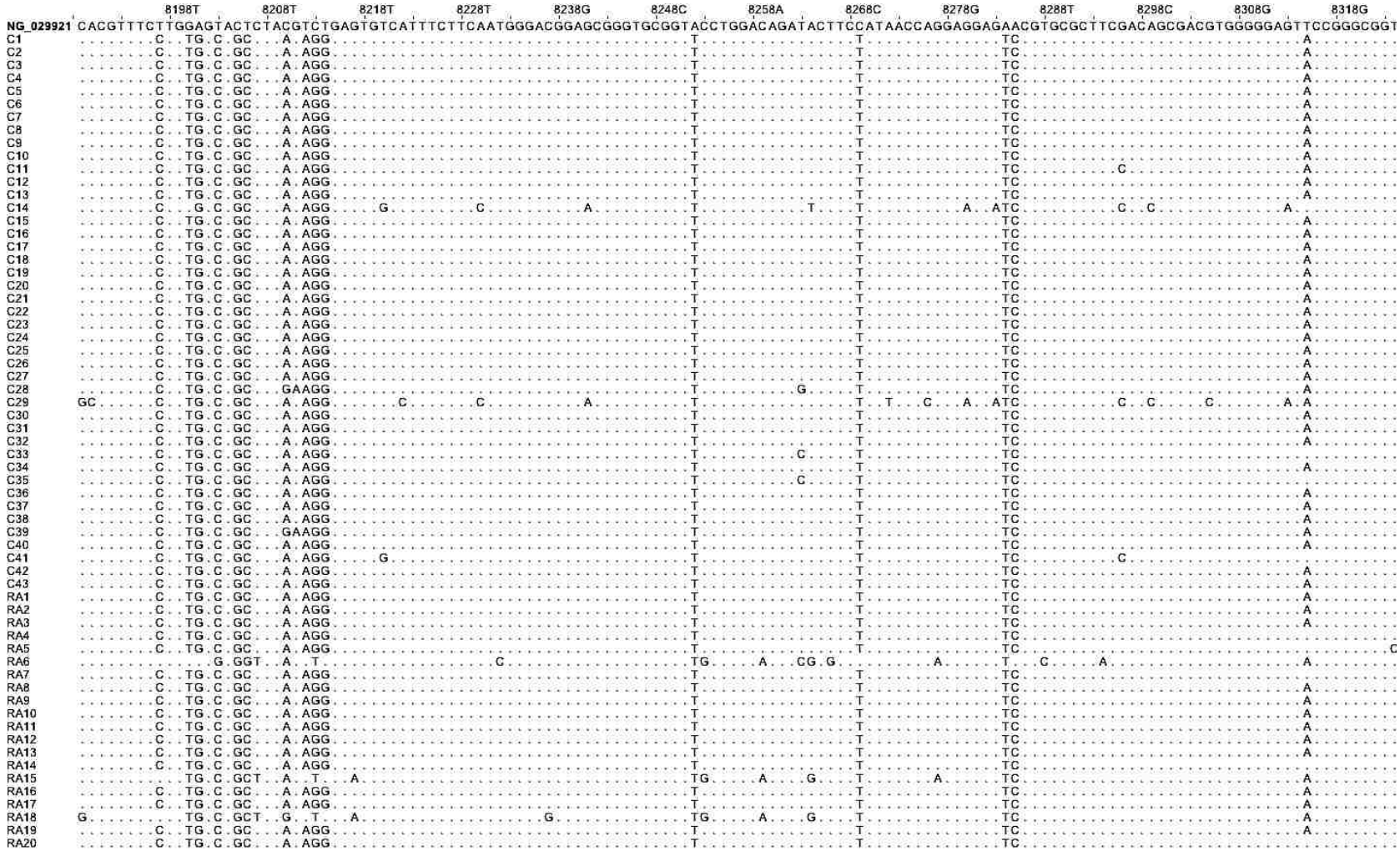

Figure 24: Sequence alignment of sample HLA-DRB1 exon 2 sequence with the reference HLA-DRB1 sequence obtained from NCBI (NG_029921). Sequence alignment was performed in JalView (Waterhouse et al., 2009).

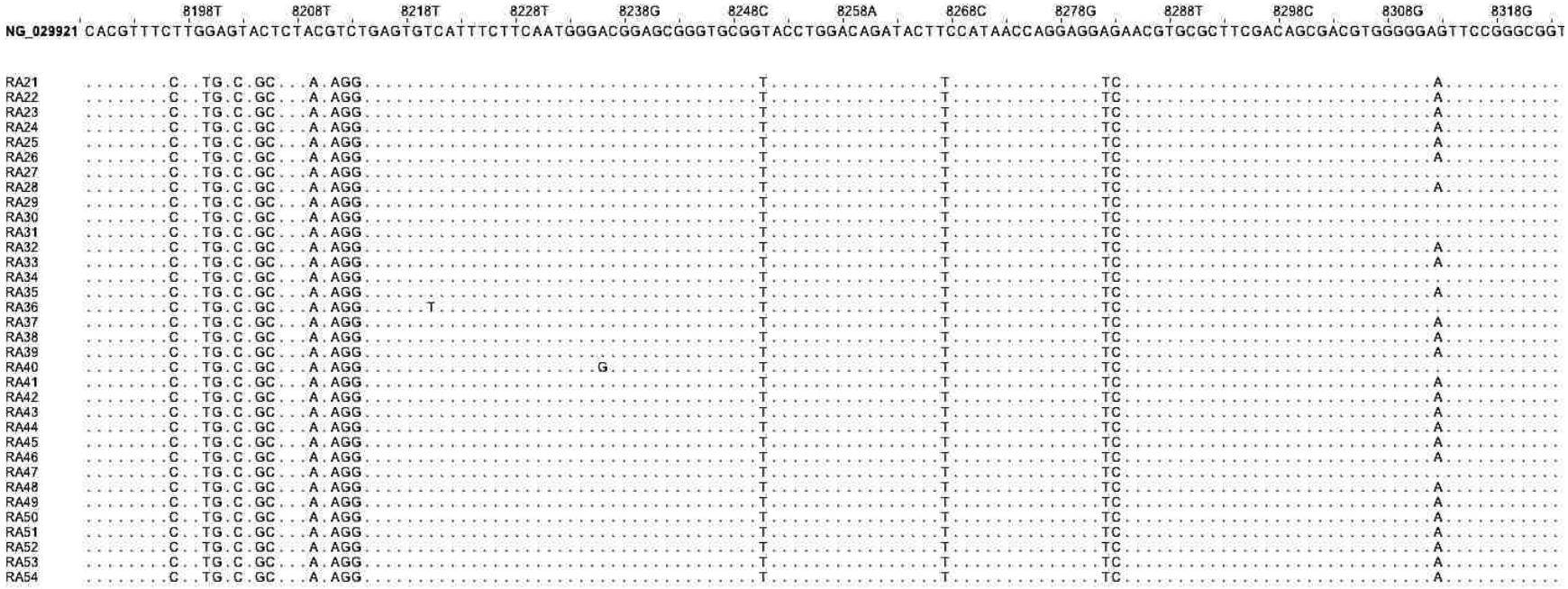

Figure 24: Sequence alignment of sample HLA-DRB1 exon 2 sequence with the reference HLA-DRB1 sequence obtained from NCBI (NG_029921). Sequence alignment was performed in JalView (Waterhouse et al., 2009) (Continued).

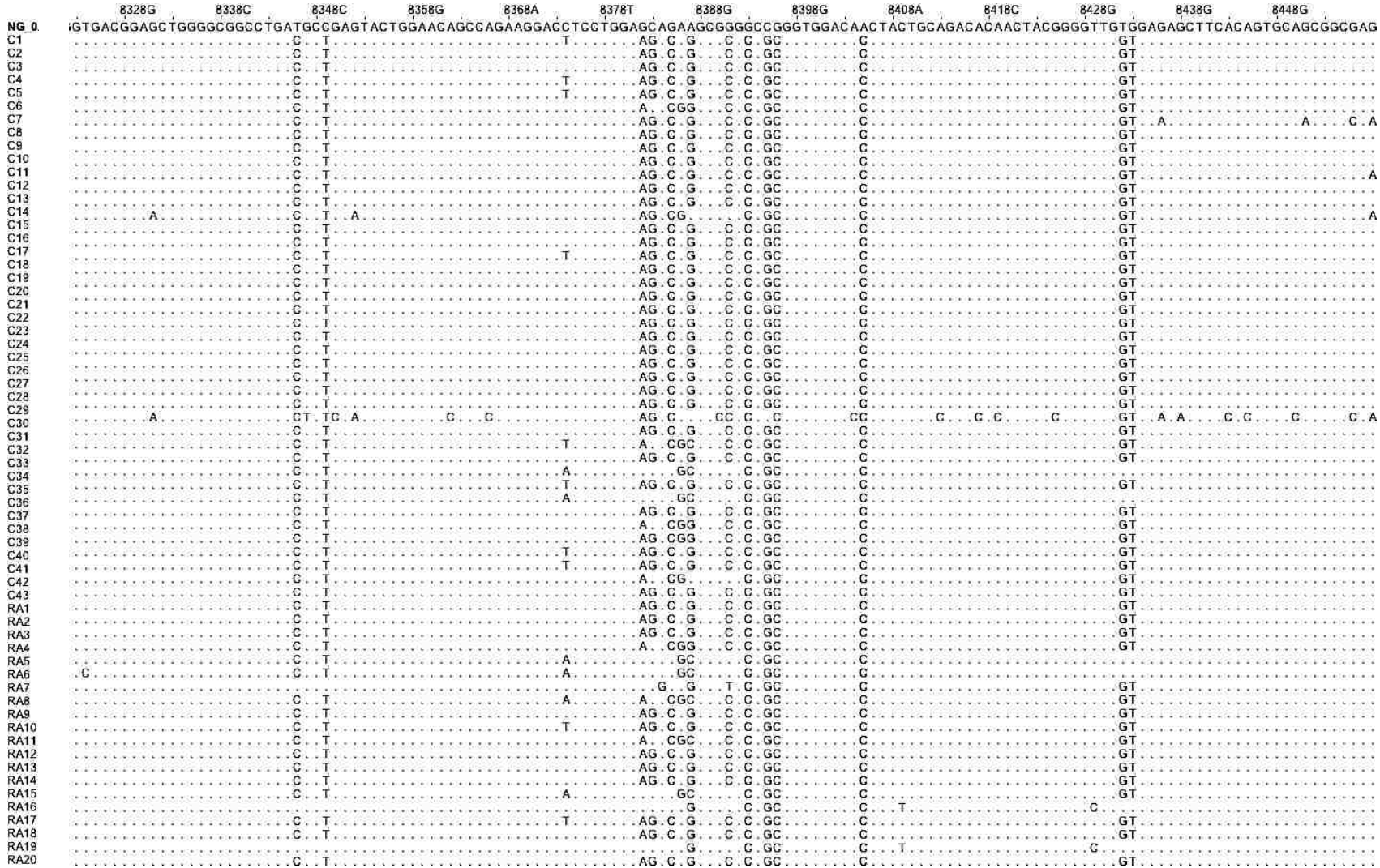

Figure 24: Sequence alignment of sample HLA-DRB1 exon 2 sequence with the reference HLA-DRB1 sequence obtained from NCBI (NG_029921). Sequence alignment was performed in JalView (Waterhouse et al., 2009) (Continued).

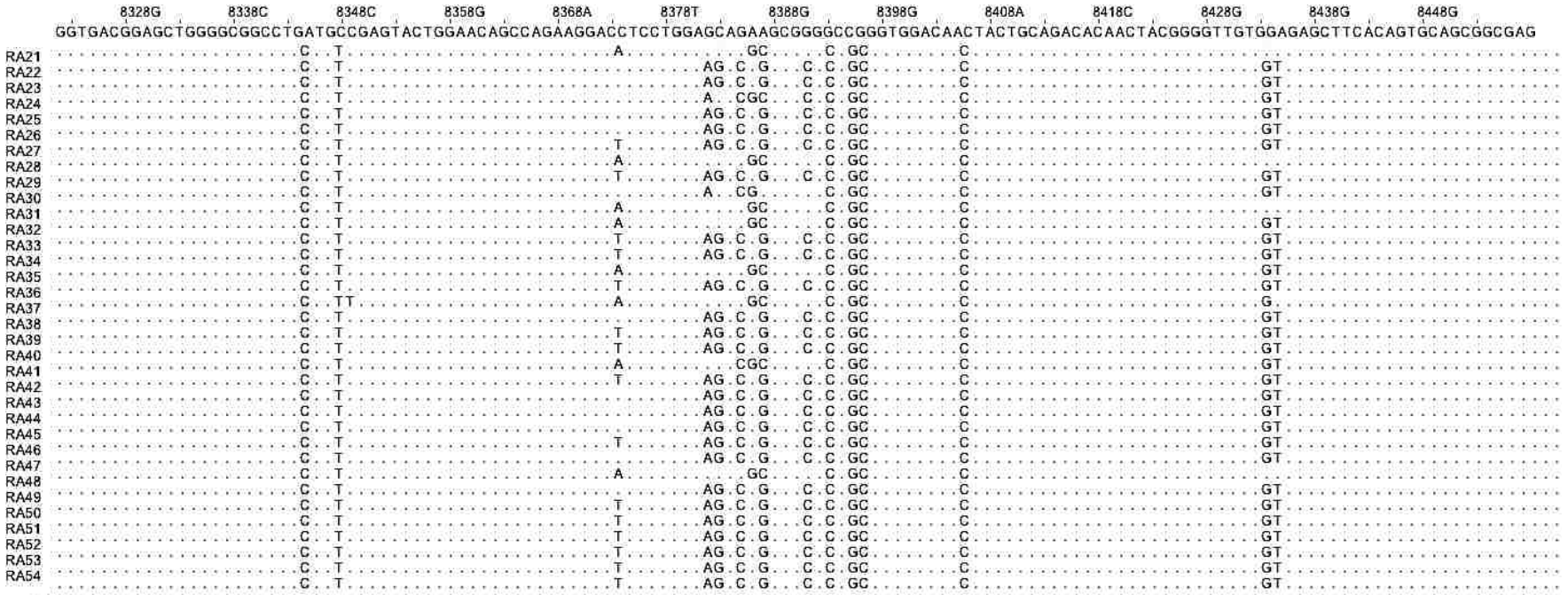

Figure 24: Sequence alignment of sample HLA-DRB1 exon 2 sequence with the reference HLA-DRB1 sequence obtained from NCBI (NG_029921). Sequence alignment was performed in JalView (Waterhouse et al., 2009) (Continued).

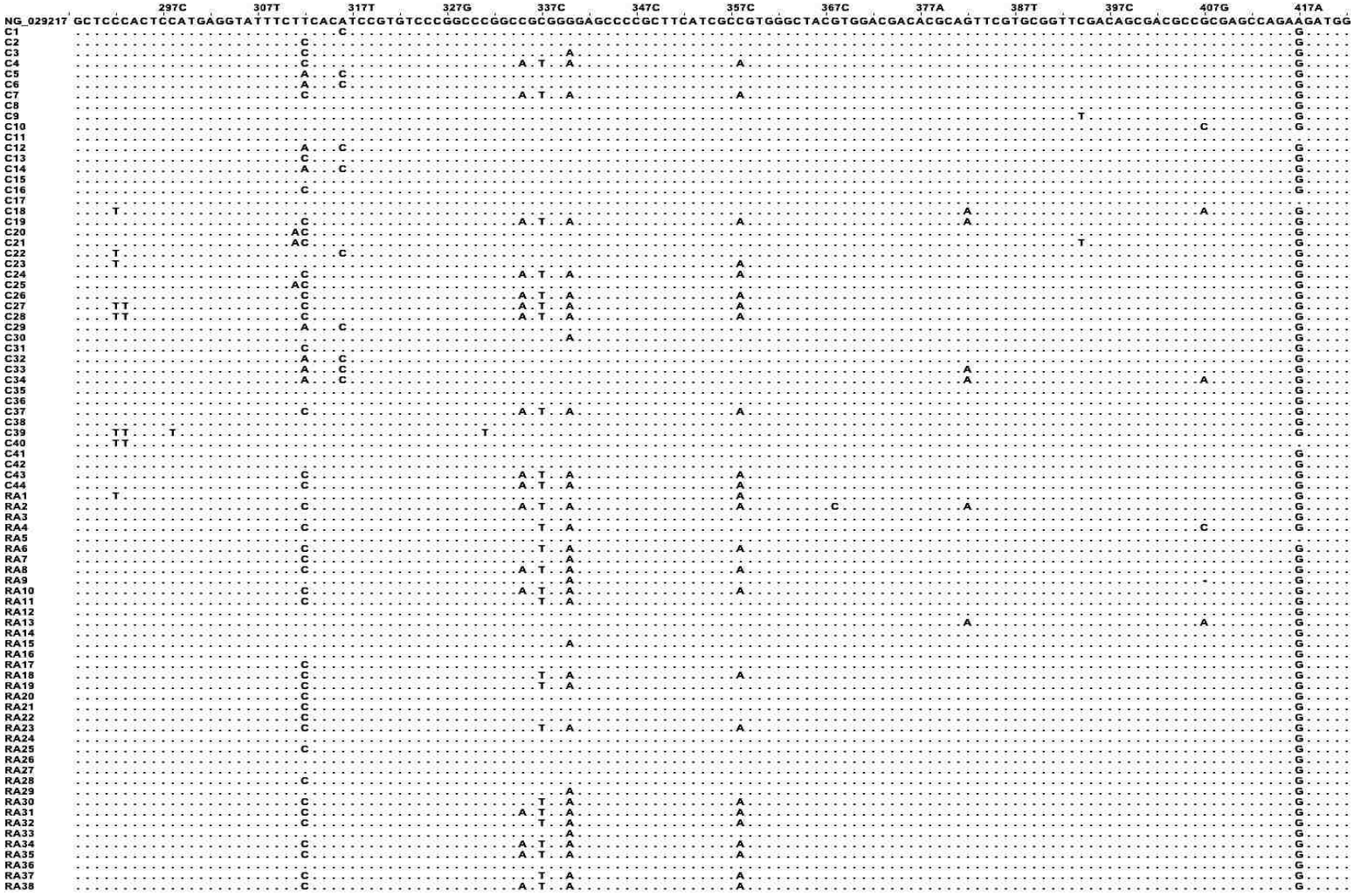

Figure 25: Sequence alignment of sample HLA-A exon 2 sequences with the reference HLA-A sequence obtained from NCBI (NG_029217). Sequence alignment was performed in JalView (Waterhouse et al., 2009).

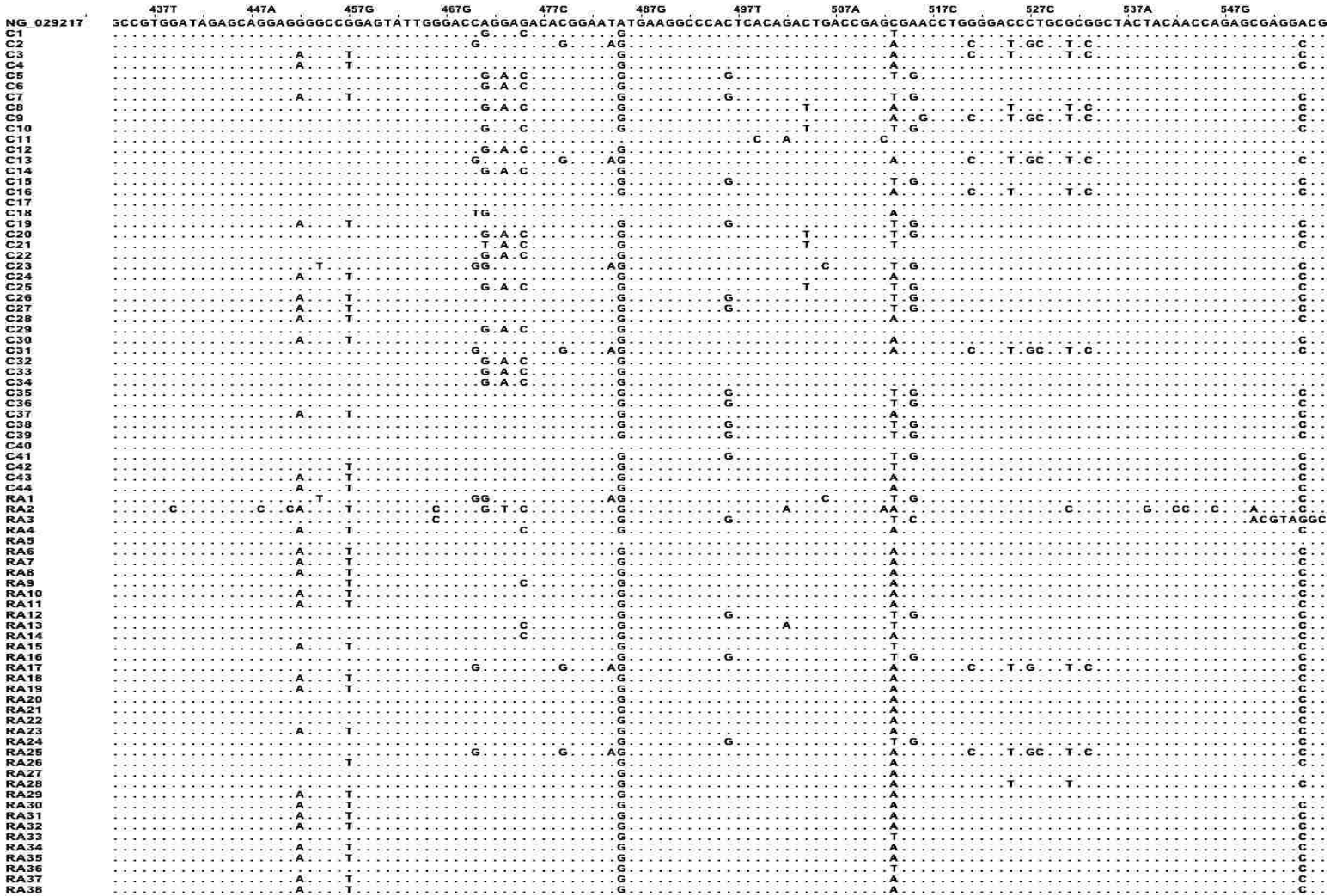

Figure 25: Sequence alignment of sample HLA-A exon 2 sequences with the reference HLA-A sequence obtained from NCBI (NG_029217). Sequence alignment was performed in JalView (Waterhouse et al., 2009) (Continued).

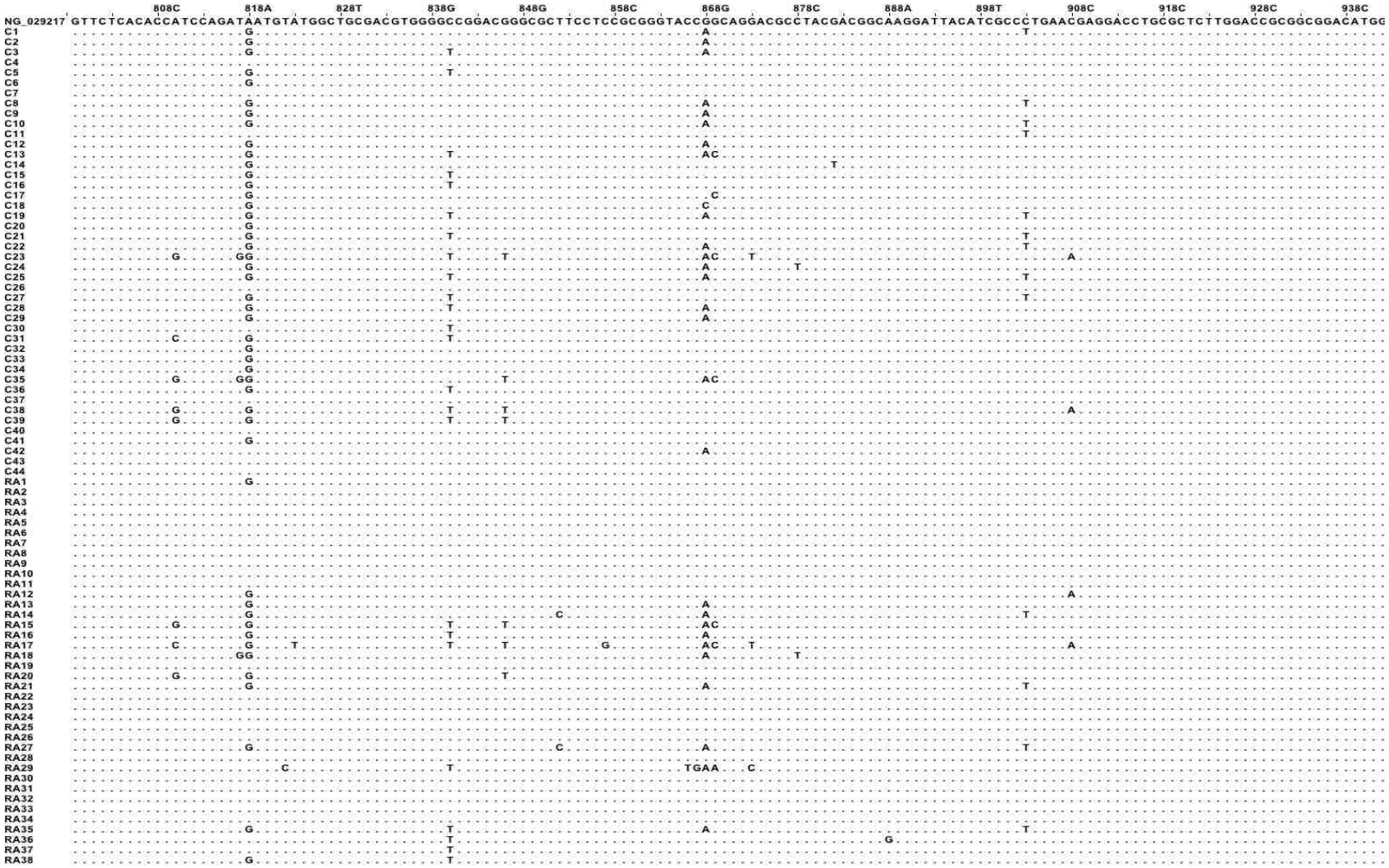

Figure 26: Sequence alignment of sample HLA-A exon 3 sequences with the reference HLA-A sequence obtained from NCBI (NG_029217). Sequence alignment was performed in JalView (Waterhouse et al., 2009).

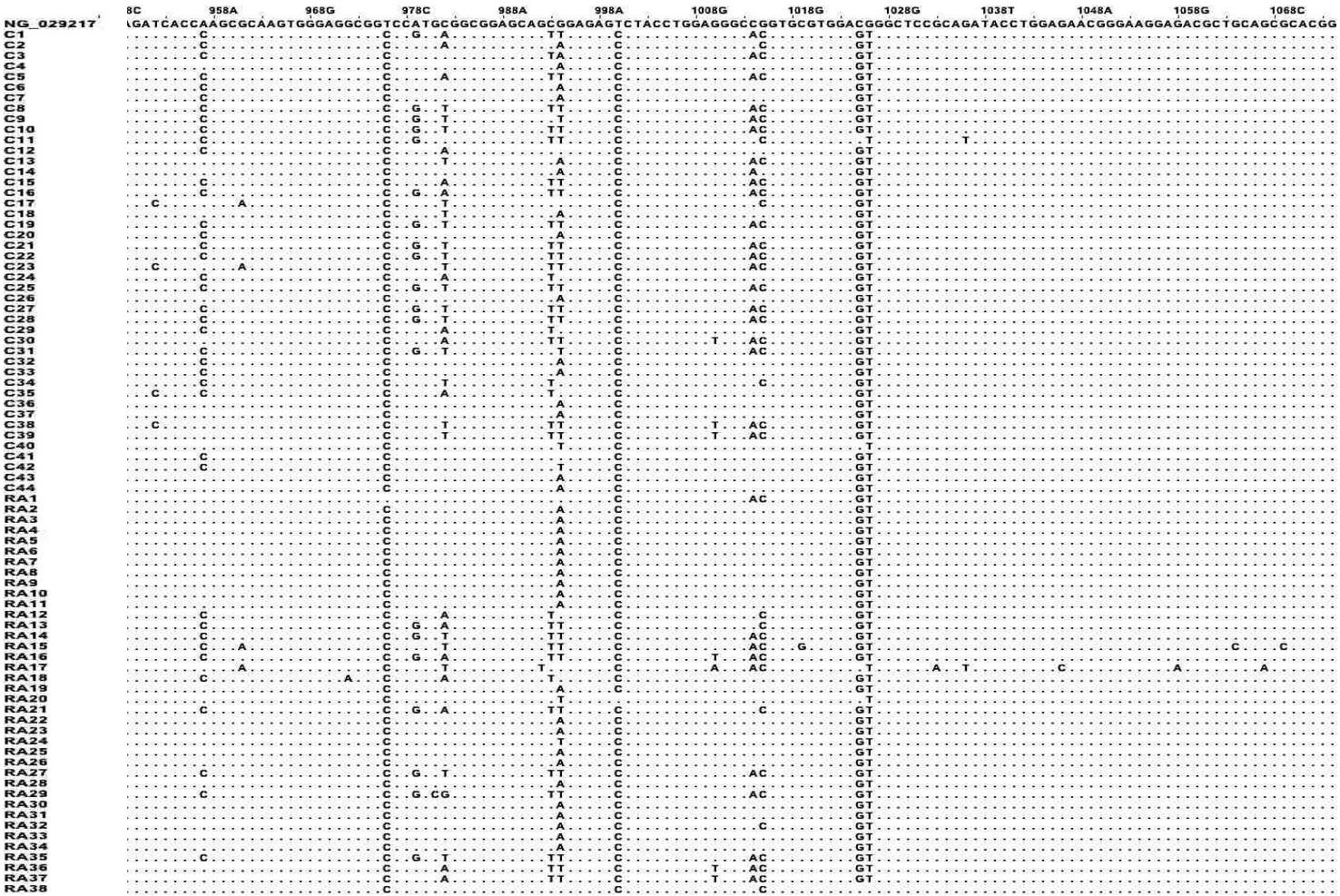

Figure 26: Sequence alignment of sample HLA-A exon 3 sequences with the reference HLA-A sequence obtained from NCBI (NG_029217). Sequence alignment was performed in JalView (Waterhouse et al., 2009) (Continued).

3.2 HLA-DRB1 Allele Typing

Allele typing of a HLA class II gene like HLA-DRB1 is normally performed with the sequence of exon 2. The exon 2 sequence of HLA-DRB1 was used as the input to perform a BLAST search against the HLA allele database deployed in a custom BLAST server in the lab (http://copbio.uae.ac.ae/blast-hla/). BLAST results were interpreted to identify the HLA-DRB1 allele type up to 4-digit resolution for control and patient samples. Allele types determined for controls are listed in Table 4 and a summary of the incidence of each allele type is provided in Table 5, while the allele types determined for patients are listed in Table 6 and the corresponding summary in Table 7.

Table 4: HLA-DRB1 allele type in control (C) samples.

Sample	Allele
C1	DRB1*16:01
C2	DRB1*16:02
C3	DRB1*16:02
C4	DRB1*16:01
C5	DRB1*16:01
C6	DRB1*16:02
C7	DRB1*16:02
C8	DRB1*16:02
C9	DRB1*16:02
C10	DRB1*16:02
C11	DRB1*16:02
C12	DRB1*16:02
C13	DRB1*16:02
C14	DRB1*15:02
C15	DRB1*16:02
C16	DRB1*16:02
C17	DRB1*16:01
C18	DRB1*16:02
C19	DRB1*16:02
C20	DRB1*16:02

Table 4: HLA-DRB1 allele type in control (C) samples (continued).

Sample	Allele
C21	DRB1*16:02
C22	DRB1*16:02
C23	DRB1*16:02
C24	DRB1*16:02
C25	DRB1*16:02
C26	DRB1*16:02
C27	DRB1*16:02
C28	DRB1*16:02
C29	DRB1*16:02
C30	DRB1*16:02
C31	DRB1*16:01
C32	DRB1*16:02
C33	DRB1*15:03
C34	DRB1*16:01
C35	DRB1*15:03
C36	DRB1*16:02
C37	DRB1*16:02
C38	DRB1*16:02
C39	DRB1*16:01
C40	DRB1*16:01
C41	DRB1*15:02
C42	DRB1*16:02
C43	DRB1*16:02

Table 5: Summary of HLA-DRB1 allele types in control samples at 4-digit and 2-digit resolutions.

Allele	Count (4-digits)	Count (2-digits)
DRB1*15:02	2	4
DRB1*15:03	2	
DRB1*16:01	8	39
DRB1*16:02	31	
Total	**43**	**43**

Tables 4 and 5 show that the most frequent HLA-DRB1 allele in control samples is DRB1*16. It was found in 39 samples out of 43 while the DRB1*15 allele comes next with 4 samples out of 43.

Table 6: HLA-DRB1 allele types in patient (RA) samples.

Sample	Allele
RA1	DRB1*16:02
RA2	DRB1*16:02
RA3	DRB1*16:02
RA4	DRB1*15:01
RA5	DRB1*10:01
RA6	DRB1*15:02
RA7	DRB1*16:02
RA8	DRB1*16:02
RA9	DRB1*16:02
RA10	DRB1*16:02
RA11	DRB1*16:02
RA12	DRB1*15:02
RA13	DRB1*01:02
RA14	DRB1*16:01
RA15	DRB1*16:02
RA16	DRB1*16:02
RA17	DRB1*16:02
RA18	DRB1*16:02
RA19	DRB1*16:02
RA20	DRB1*16:02
RA21	DRB1*16:01
RA22	DRB1*15:01
RA23	DRB1*16:01
RA24	DRB1*15:02
RA25	DRB1*15:01
RA26	DRB1*15:02
RA27	DRB1*15:01
RA28	DRB1*16:01
RA29	DRB1*15:02
RA30	DRB1*15:01
RA31	DRB1*15:02

Table 6: HLA-DRB1 allele types in patient (RA) samples (continued).

Sample	Allele
RA32	DRB1*16:01
RA33	DRB1*16:01
RA34	DRB1*15:02
RA35	DRB1*16:01
RA36	DRB1*15:114
RA37	DRB1*16:02
RA38	DRB1*16:01
RA39	DRB1*16:01
RA40	DRB1*15:02
RA41	DRB1*16:01
RA42	DRB1*16:02
RA43	DRB1*16:02
RA44	DRB1*16:02
RA45	DRB1*16:01
RA46	DRB1*16:02
RA47	DRB1*15:01
RA48	DRB1*16:02
RA49	DRB1*16:01
RA50	DRB1*16:01
RA51	DRB1*16:01
RA52	DRB1*16:01
RA53	DRB1*16:01
RA54	DRB1*16:01

Table 7: Summary of HLA-DRB1 allele types in RA samples at 4-digit and 2-digit resolutions.

Allele	Count (4-digits)	Count (2-digits)
DRB1*01:02	1	1
DRB1*10:01	1	1
DRB1*15:01	6	15
DRB1*15:02	8	
DRB1*15:114	1	
DRB1*16:01	17	37
DRB1*16:02	20	
Total	**54**	**54**

Interestingly, the DRB1*16 allele was the most abundant allele in patients. The high risk alleles type such as DRB1*01 and DRB1*10 were found only is 3 RA samples.

3.3 Sequence of Shared Epitope and Other Regions of HLA-DRB1 Associated with RA

The shared epitope region, which forms amino acids 70-74 of the mature protein, is by far regarded as the strongest indicator of an association between HLA genes and RA (Ruyssen-Witrand et al., 2015). From the sequence alignment and translation from this, the SE region and the corresponding amino acids representing this region are provided in Table 8 for controls and a summary of unique SE sequences observed is provided in Table 9. SE sequences for patients are provided in Table 10 and the corresponding summary in Table 11. Outside the SE regions, variations at positions 11 and 13 of the mature HLA-DRB1 protein are also known to be associated with RA. The corresponding sequences observed in these samples are also listed in the tables mentioned above. The sequence DRRAA was the most common SE in controls (34/43) as well as patients (35/54). The QARAA was also present in a number of patient samples (10/54), while the other observed SE sequences had low incidence in these samples.

Table 8: HLA-DRB1 Shared epitope and positions 11/13 in control samples.

Sample	Shared Epitope (DNA)	SE(70-74) Amino acids	Codons 11-13 (DNA)	Amino acids 11-13
C1	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C2	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C3	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C4	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C5	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R

Table 8: HLA-DRB1 Shared epitope and positions 11/13 in control samples. (Continued)

Sample	Shared Epitope (DNA)	SE(70-74) Amino acids	Codons 11-13 (DNA)	Amino acids 11-13
C6	CACGGGCGCGCCGCG	HGRAA	CCTAAGAGG	P-R
C7	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C8	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C9	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C10	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C11	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C12	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C13	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C14	GACGAGCGGGCCGCG	DERAA	NOT AVAILABLE	
C15	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C16	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C17	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C18	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C19	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C20	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C21	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C22	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C23	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C24	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C25	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C26	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C27	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C28	GACAGGCGCGCCGCG	DRRAA	CCTAAGAAG	P-K
C29	GACAAGCCCGCCCCG	DKPAP	CCTAAGAGG	P-R
C30	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C31	CACGCGCGCGCCGCG	HARAA	CCTAAGAGG	P-R
C32	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C33	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
C34	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C35	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
C36	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C37	CACGGGCGCGCCGCG	HGRAA	CCTAAGAAG	P-R
C38	GACGGGCGCGCCGCG	DGRAA	CCTAAGAGG	P-R
C39	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C40	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C41	CACGAGCGGGCCGCG	HERAA	CCTAAGAGG	P-R
C42	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
C43	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R

Table 9: Count of shared epitope sequence in control samples.

Shared epitope sequence	**Count**
DRRAA	34
HGRAA	2
DERAA	1
DKPAP	1
HARAA	1
QARAA	2
DGRAA	1
HERAA	1
Total	43

The low risk SE sequence DRRAA is more frequent in control samples which is related to HLA-DRB1*16 allele type. The amino acid at position 11 and 13 did not the high risk variant in the control samples.

Table 10: HLA-DRB1 Shared epitope and positions 11/13 in RA samples.

Sample	Shared Epitope (DNA)	SE(70-74) Amino acids	Codons 11-13 (DNA)	Amino acids 11-13
RA1	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA2	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA3	CACGGGCGCGCCGCG	HGRAA	CCTAAGAGG	P-R
RA4	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA5	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA6	CGGAGGCGTGCCGCG	RRRAA	GTTAAGTTT	V-F
RA7	CACGCGCGCGCCGCG	HARAA	CCTAAGAGG	P-R
RA8	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA9	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA10	CACGCGCGCGCCGCG	HARAA	CCTAAGAGG	P-R
RA11	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA12	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA13	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA14	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA15	CAGAGGCGGGCCGCG	QRRAA	CTTAAGTTT	L-F
RA16	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA17	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA18	CAGAGGCGGGCCGCG	QRRAA	CTTAGGTTT	L-F

Table 10: HLA-DRB1 Shared epitope and positions 11/13 in RA samples. (Continued)

Sample	Shared Epitope (DNA)	SE(70-74) Amino acids	Codons 11-13 (DNA)	Amino acids 11-13
RA19	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA20	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA21	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA22	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA23	CACGCGCGCGCCGCG	HARAA	CCTAAGAGG	P-R
RA24	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA25	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA26	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA27	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA28	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA29	CACGAGCGGGCCGCG	HERAA	CCTAAGAGG	P-R
RA30	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA31	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA32	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA33	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA34	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA35	GACAGGCGCGCCGC	DRRAA	CCTAAGAGG	P-R
RA36	CAGGCGCGGGCCGC	QARAA	CCTAAGAGG	P-R
RA37	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA38	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA39	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA40	CACGCGCGGGCCGCG	HARAA	CCTAAGAGG	P-R
RA41	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA42	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA43	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA44	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA45	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA46	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA47	CAGGCGCGGGCCGCG	QARAA	CCTAAGAGG	P-R
RA48	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA49	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA50	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA51	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA52	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA53	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R
RA54	GACAGGCGCGCCGCG	DRRAA	CCTAAGAGG	P-R

Table 11: Count of shared epitope in RA samples.

Shared epitope sequence	Count
DRRAA	35
HGRAA	1
QARAA	10
RRRAA	1
HARAA	4
QRRAA	2
HERAA	1
Total	**54**

In RA patients, the high risk SE sequences (RRRAA and QRRAA) were observed only in the 3 sample out of 54 which belong to the high risk allele type of HLA-DRB1 mentioned in Tables 6 and 7.

3.4 HLA-A Allele Typing

Unlike HLA-DRB1, a class II gene, allele typing of a HLA class I gene like HLA-A is normally performed with the sequence of exon 2 and 3. For this, the combined sequence of exon 2 and 3 of HLA-A was used as the input to perform a BLAST search against the HLA allele database. BLAST results were interpreted to identify the HLA-A allele type for control and patient samples. Allele types determined for controls are listed in Table 12 and a summary of the incidence of each allele type is provided in Table 13, while the allele types determined for patients are listed in Table 14 and the corresponding summary in Table 15. At 2-digit resolution the allele HLA-A*11 was the most common allele type observed in the controls (14/44) and patients (21/38). Specifically, the HLA-A*11:88 had the highest incidence of 15/38 among patients.

Table 12: HLA-A allele type in control samples.

Sample	Allele
C1	A*31:110
C2	A*31:08
C3	A*31:10
C4	A*11:88
C5	A*34:03
C6	A*11:40
C7	A*11:88
C8	A*32:15
C9	A*32:01
C10	A*33:119
C11	A*01:244
C12	A*26:70
C13	A*03:10/A*33:150/A*68:129
C14	A*11:40
C15	A*03:22
C16	A*29:02/A*33:18
C17	A*01:200
C18	A*11:17
C19	A*31:111/A*33:9
C20	A*33:59
C21	A*33:18
C22	A*33:13
C23	A*02:97
C24	A*30:117
C25	A*33:85
C26	A*11:88
C27	A*30:26
C28	A*30:09
C29	A*26:70
C30	A*03:57
C31	A*23:70
C32	A*11:40
C33	A*11:40
C34	A*26:172
C35	A*03:78
C36	A*11:26
C37	A*11:99
C38	A*03:343
C39	A*03:343
C40	A*01:136
C41	A*11:33
C42	A*11:199
C43	A*11:99
C44	A*11:99

Table 13: Summary of HLA-A allele types in control samples at 4- and 2-digit resolutions.

Allele	Count (4 digits)	Count (2 digits)
A*01:136	1	3
A*01:200	1	
A*01:244	1	
A*02:97	1	1
A*03:22	1	5
A*03:343	2	
A*03:57	1	
A*03:78	1	
A*11:17	1	14
A*11:26	1	
A*11:33	1	
A*11:40	4	
A*11:88	3	
A*11:99	3	
A*11:199	1	
A*23:70	1	1
A*26:70	2	3
A*26:172	1	
A*30:09	1	3
A*30:26	1	
A*30:117	1	
A*31:08	1	3
A*31:10	1	
A*31:110	1	
A*32:01	1	2
A*32:15	1	
A*33:13	1	5
A*33:18	1	
A*33:59	1	
A*33:85	1	
A*33:119	1	1
A*34:03	1	1
A*03:10/A*33:150/A*68:129	1	1
A*29:02/A*33:18	1	1
A*31:111/A*33:9	1	1
Total	44	44

Table 14: HLA-A allele type in RA samples.

Sample	Allele
RA1	A*03:89
RA2	A*11:88
RA3	A*11:88
RA4	A*11:88
RA5	A*01:136
RA6	A*11:88
RA7	A*11:88
RA8	A*11:88
RA9	A*11:40
RA10	A*11:88
RA11	A*11:88
RA12	A*03:56
RA13	A*74:34
RA14	A*74:34
RA15	A*31:21
RA16	A*03:280
RA17	A*24:87
RA18	A*26:19
RA19	A*11:88
RA20	A*01:192
RA21	A*74:34
RA22	A*11:88
RA23	A*11:88
RA24	A*11:199
RA25	A*11:178
RA26	A*11:40
RA27	A*74:10
RA28	A*11:88
RA29	A*30:12
RA30	A*11:99
RA31	A*11:88
RA32	A*11:88
RA33	A*11:199
RA34	A*11:88
RA35	A*31:79
RA36	A*03:267
RA37	A*03:72
RA38	A*01:192

Table 15: Summary of HLA-A allele types in RA samples at 4- and 2-digit resolutions.

Allele	Count (4 digits)	Count (2 digits)
A*01:136	1	3
A*01:192	2	
A*03:267	1	5
A*03:280	1	
A*03:56	1	
A*03:72	1	
A*03:89	1	
A*11:178	1	21
A*11:199	2	
A*11:40	2	
A*11:88	15	
A*11:99	1	
A*24:87	1	1
A*26:19	1	1
A*30:12	1	1
A*31:21	1	2
A*31:79	1	
A*74:10	1	4
A*74:34	3	
Total	**38**	**38**

3.5 Amino Acid Position in HLA-A Associated with RA

In HLA-A, an asparagine (N) at position 77 in the binding groove of the mature protein is known to carry a high RA risk while the wild type is an aspartic acid (D) (Han et al., 2014). Hence, codon 77 and the translated amino acid at this position were determined for all samples. The amino acid at position 77 for control samples are listed in Table 16 and patient samples in Table 17. A summary of the observed amino acids at this position in patients and controls are provided in Table 18.

Table 16: HLA-A amino acid at position 77 in control samples.

Sample	Amino acid
C1	N/D
C2	N
C3	N
C4	N
C5	D
C6	N
C7	D
C8	N/D
C9	S
C10	D
C11	N
C12	N
C13	E
C14	N/D
C15	D
C16	N
C17	N
C18	N/D
C19	D
C20	D
C21	N/D
C22	N/D
C23	D
C24	N
C25	D
C26	D
C27	D
C28	N
C29	N/D
C30	N
C31	N
C32	N
C33	N/D
C34	N/D
C35	D

Table 16: HLA-A amino acid at position 77 in control samples (Continued).

Sample	Amino acid
C36	D
C37	N
C38	D
C39	D
C40	N
C41	N/D
C42	N/D
C43	N
C44	N

Table17: HLA-A amino acid at position 77 in RA samples.

Sample	Amino acid
RA1	D
RA2	N
RA3	N
RA4	N
RA5	N
RA6	N
RA7	N
RA8	N
RA9	N
RA10	N
RA11	N
RA12	N/D
RA13	N/D
RA14	N
RA15	N
RA16	N/D
RA17	N
RA18	N
RA19	N
RA20	N/D
RA21	N/D
RA22	N
RA23	N
RA24	N/D
RA25	N
RA26	N
RA27	N
RA28	N

Table17: HLA-A amino acid at position 77 in RA samples (Continued).

Sample	Amino acid
RA29	N
RA30	N
RA31	N
RA32	N
RA33	N/D
RA34	N
RA35	N
RA36	N/D
RA37	N
RA38	N

Table 18: Summary of HLA-A amino acid at position 77 in all samples.

Amino acid	Count (Control)	Count (RA)
N	17	29
D	14	1
N/D	11	8
Non N/D	2	0
Total	**44**	**38**

Tables 16-18 shows the presence of the missense mutation at position 77 in HLA-A samples where aspartic acid was changed to asparagine. This was observed to be more frequent in patients than controls.

3.6 Other Locations in HLA-A with Higher Incidence in Patient

Based on the alignment obtained, frequency of variations in controls and patients were calculated at all locations in exon 2 and exon 3 of HLA-A. A higher incidence of variations was observed in patients at positions 56, where a arginine (R) replaces a glycine (G) in the reference, and 76, where a glutamic acid (E) replaces a

valine (V). To understand the significance of these, structural models of the wildtype (Figure 27A) and mutants Gly56Arg (Figure 27B), Val76Glu (Figure 27C) and Asp77Asn (Figure 27D) were generated using molecular modeling.

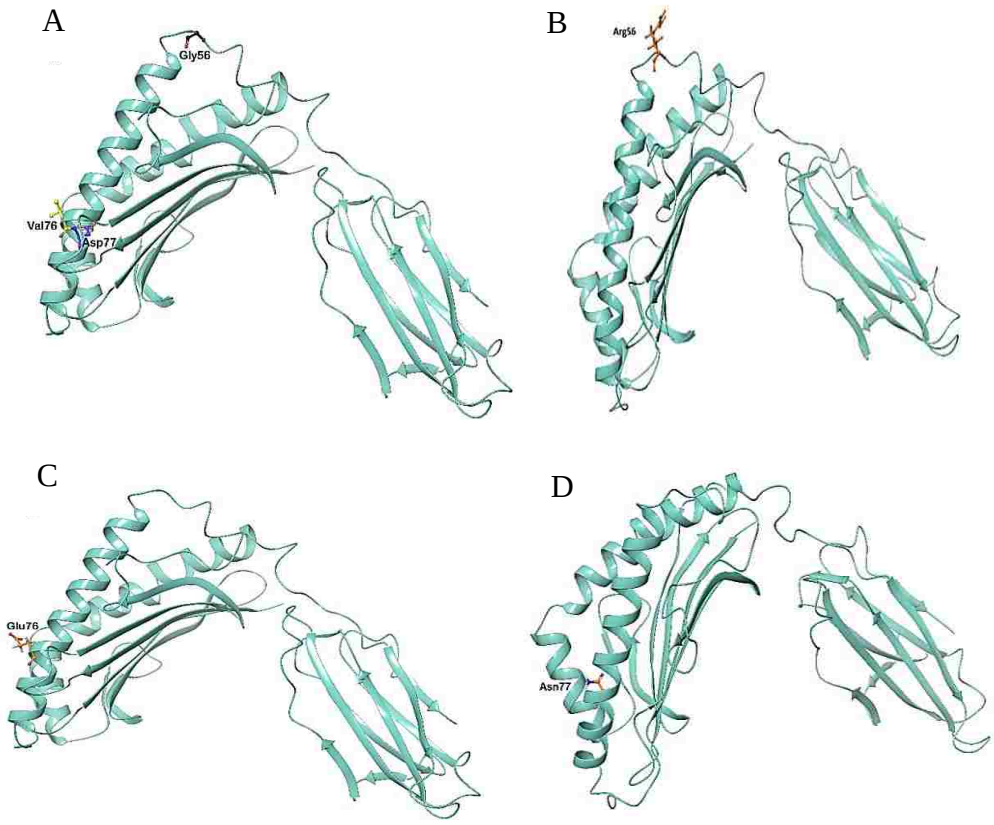

Figure 27: Three-dimensional models of wildtype and mutant HLA-A: (A) wildtype; (B) Gly56Arg; (C) Val76Glu and (D) Asp77Asn.

3.7 Odds Ratio for the Three Observed Positions in HLA-A

To estimate the significance of the observed variations, Fisher's exact test was performed to compute the odds ratio and significance using R. The computed values are shown in Table 19.

Table 19: Variations in HLA-A with higher incidence in RA samples than control samples.

Position	Variant	RA (n=38)	Control (n=44)	Odds Ratio (OR)	p-value (Fisher's test)
77	N	37	28	20.52	0.0001985
76	E	29	15	6.075	0.0001589
56	R	20	10	3.712	0.006273

3.8 Phylogenetic Analysis of Sequences

To understand the evolutionary relationship between samples, phylogenetic trees were constructed using all HLA-DRB1 exon 2 sequences (Figure 28) and HLA-A exon 2+3 sequences (Figure 29) using MEGA 6.

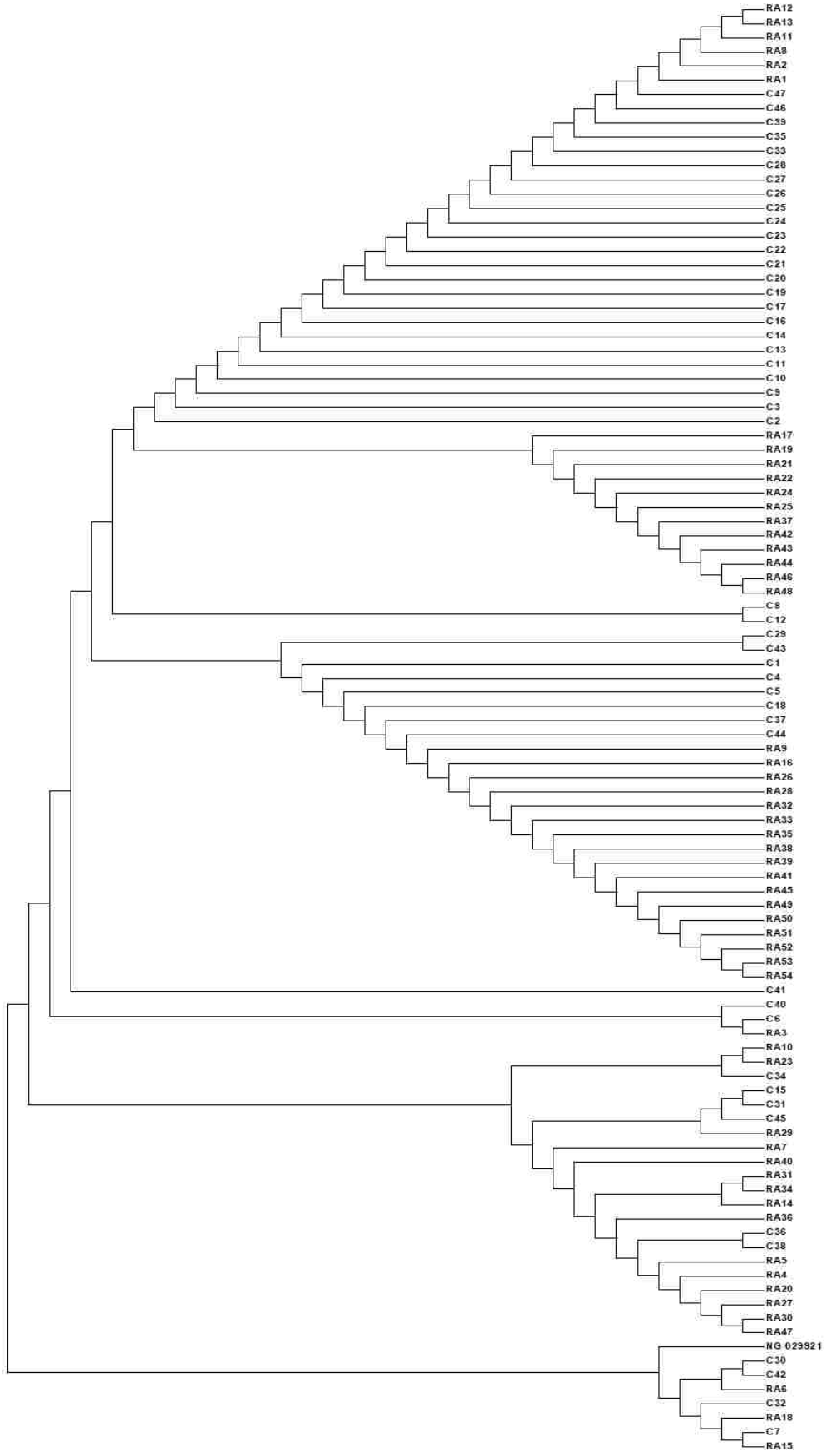

Figure 28: Phylogenetic constructed from HLA-DRB1 exon 2 sequences.

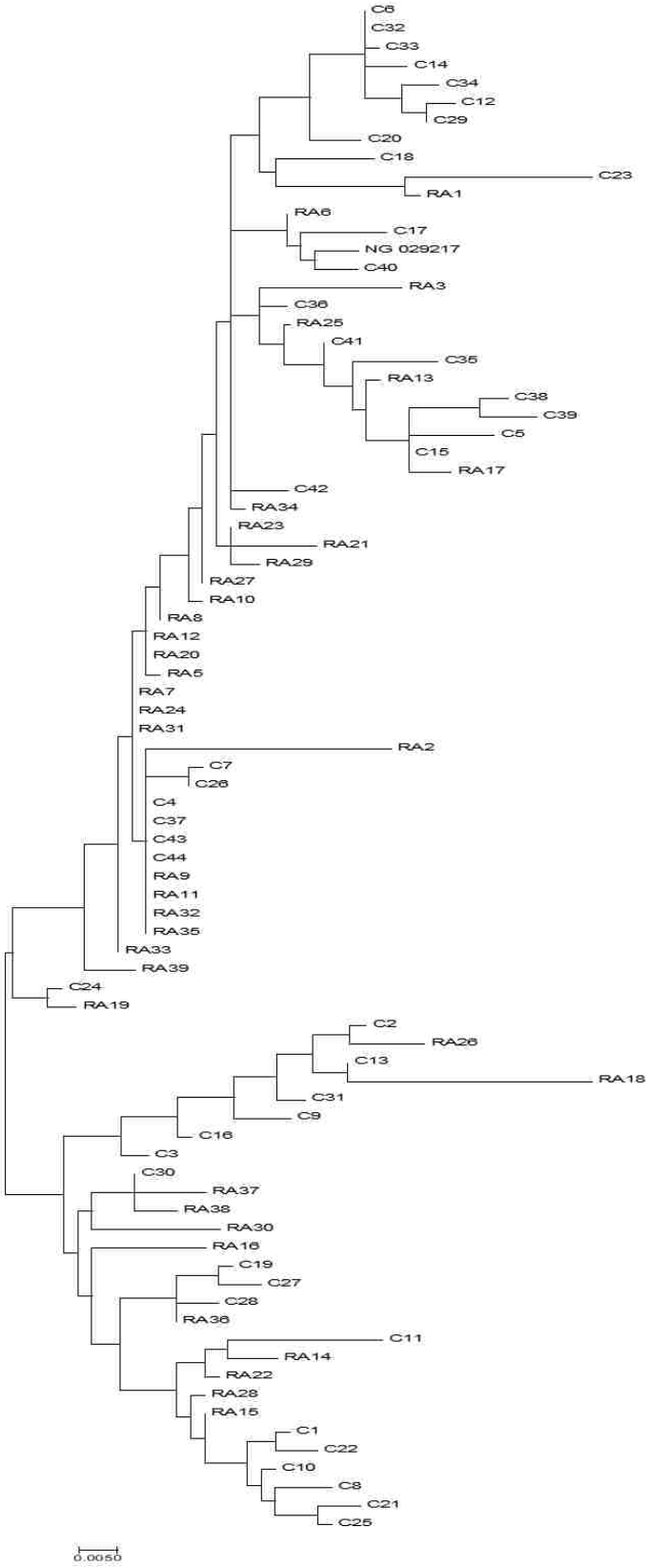

Figure29: Phylogenetic constructed from all HLA-A exon 2 & 3 sequences.

3.9 Principal Component Analysis

Principal Component Analysis (PCA) of the DNA sequences was performed in R using the *adegenet* package to reduce the dimensionality of the dataset and to determine if the patient and control samples would distinctly cluster. PCA plots obtained for principal component 1 versus principal component 2 for HLA-DRB1 is shown in Figure 30 and HLA-A in Figure 31.

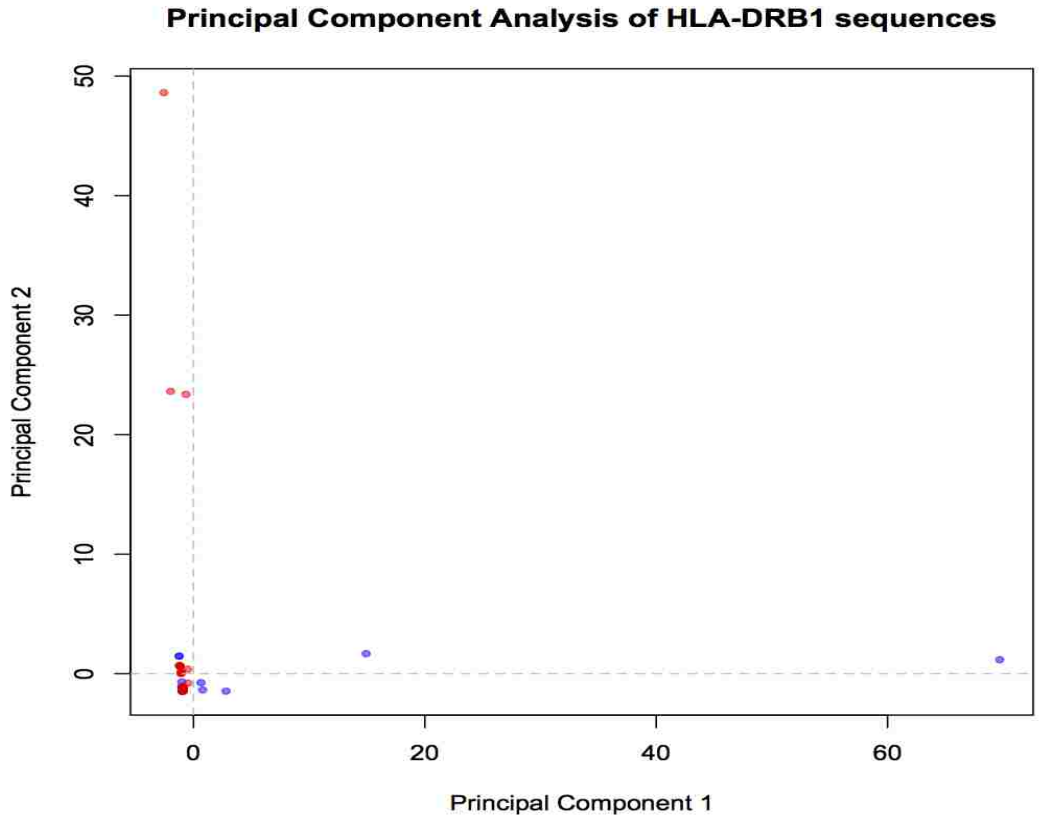

Figure 30: A plot of principal component 1 against component 2 obtained from PCA of HLA-DRB1 exon 2 sequences. Red dots represent patient samples and blue dot represents control samples.

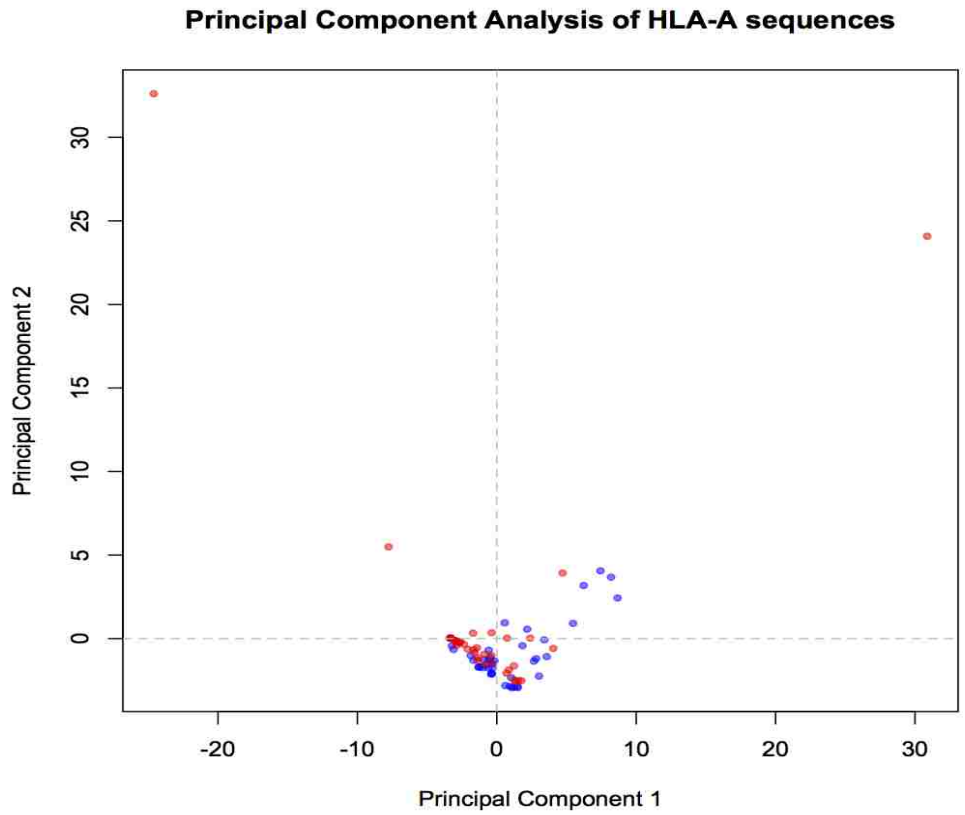

Figure 31: A plot of principal component 1 against component 2 obtained from PCA of HLA-A exon 2 and 3 sequences. Red dots represent patient samples and blue dot represents control samples.

3.10 Fixation Index

Fixation Index (F_{ST}) was computed for HLA-DRB1 and HLA-A sequences using Arlequin. The obtained F_{ST} values were 0.02263 (p-value 0.04985) for HLA-DRB1 and 0.06811 (p-value 0.00098) for HLA-A.

Chapter 4: Discussion

RA is a chronic heterogeneous autoimmune disorder of unknown etiology resulting in inflammation in the cartilage, synovium, and bone (O'Rielly & Rahman, 2010). It is the second most predominant autoimmune disorder. Early diagnosis and management is important for the notable improvement and favorable outcomes. The absence of timely intervention lead to negative impact to patient health together with a large societal burden (Kurko et al., 2013). RA usually appears after 30-40 years of life and is more common in women which may be related to hormonal factors. RA is reported to be present in 0.5-1% of the worldwide (Sangha, 2000; Trier et al., 2018).

Genetic factors often play a defining role in the pathogenesis of RA. Studies have reported that the heritability of RA is from 50-60% (MacGregor et al., 2000). There are multiple known loci that are associated with the risk of developing RA (Bowes & Barton, 2008). Consequently, various genes have been studied as risk factors for RA. Most of these genes are believed to be involved in immune system and forms part of the major histocompatibility complex (MHC), known in the humans as the Human Leukocyte Antigens (HLA). Importantly, this region is responsible for approximately 30% of genetic susceptibility to RA (Bowes & Barton, 2008). Specifically, among the HLA genes, HLA-DRB1 alleles are strongly associated with RA (Jun et al., 2007; Lee et al., 2008). However, multiple studies have indicated that other HLA genes are also linked with RA including HLA-A, HLA-B, and HLA-DPB1 (Coenen & Gregersen, 2009; Imboden, 2009). Interestingly, various alleles of HLA-DRB1 which are known to be associated with RA associate with what is known as the SE region (Huizinga et al., 2005). This region is encoded by exon 2 of the HLA-DRB1 gene and corresponds to amino acids 70-74 in the mature protein. It has been

established that the acid-base properties of SE amino acid could have impact on the interactions of HLA-DRB1SE motif (du Montcel et al., 2005). Alleles that have the conserved SE amino acids sequences such as QKRAA, QRRAA, and RRRAA in the third hypervariable region of the DRB1 chain between positions 70 to 74 are very strongly associated with RA and is often used as a genetic biomarker for RA risk association (Gregersen et al., 1987; Morgan et al., 2008). DRB1*04:01,*04:04, *04:05, *04:08, *13:03, *01:01, *01:02, *14:02, *14:06, *15:01, *15:02, and 15:03 are the most studied alleles linked with RA but not all alleles display the same magnitude of association (Gonzalez-Gay et al., 2002; Gorman et al., 2004; Imboden, 2009). SE allele's frequency varies between populations. It has been reported that DRB1*04:01 and 04:04 were significantly associated with RA in Caucasian populations, whereas Asian and Jewish RA patients were predominantly linked with DRB1*04:05 and *01:01 alleles (Barnetche et al., 2008; Okada et al., 2014; Ollier & Thomson, 1992). Based on this, the basic hypothesis of this study was that RA in Emirati patients could also be associated with the high risk SE sequences. This is the first study to our knowledge, to investigate the association of shared epitope with RA in Emirati population. For this, 44 control and 54 RA Emirati samples were obtained and evaluated from Fujairah Hospital after obtaining informed consent. HLA-DRB1 exon 2 of each subject was amplified and sequenced by using Sanger sequencing method.

For the HLA-DRB1 gene, the most prevalent allele in the patient group and control group was DRB1*16:02 which is considered as a low risk allele in the context of RA. The high risk alleles DRB1*10:01 was found in only one patient and DRB1*01:02 was found in 2 patients. Out of 43 control subjects 34 had the amino acid sequence DRRAA as the SE, while this SE was observed in 35 of 54 patient samples

(Tables 9 & 11). Studies have suggested that the DRRAA sequence which usually corresponds to HLA-DRB1*16, *11, and *12 alleles, is a low risk allele and is not associated with RA (du Montcel et al., 2005). Moreover, it was reported that amino acid at position 70 influences the susceptibility to RA and the presence of aspartic acid (D) at this position is believed to provide a protective effect as opposed to glutamine (Q) or arginine (R) (Trier et al., 2018). Interestingly, RRRAA is a high risk allele and mostly prevalent in Asian population (Mohan et al., 2017; Muazzam et al., 2013). This SE was observed in only 1 out of 54 RA patients and was absent in control subjects.

Together, the high risk SEs QRRAA/RRRAA were observed in only 3 patients. Surprisingly, the low risk SEs DRRRA/QARAA/DERAA were observed in 37 of 43 controls and 45 of 54 patients (Table 9 & 11). The shared epitope hypothesis was formulated on the basis of a large body of evidence suggesting that the SE is the most significant cause of RA that can be attributed to variations in the HLA genes (Holoshitz, 2010). Putting together the data in this study, it appears that the low risk SEs predominate both the patient and control samples and high risk SEs have low incidence indicating that the HLA-DRB1 SE region is not the major cause of RA in this sample of the population.

Looking outside the SE region of HLA-DRB1, it has been established that specific amino acids at positions 11 and 13 could also impart high RA risk. These amino acids line the base of the peptide binding groove (Scally et al., 2013). A valine or leucine at position 11 (neutral) and/or a histidine at position 13 (basic) have been associated with an increased risk of RA (du Montcel et al., 2005; Raychaudhuri et al., 2012; Reynolds et al., 2014; Scally et al., 2013). Therefore, this region was examined in both patient and control samples. However, in the vast majority of patient and control samples, a proline was found at position 11 and an arginine at position 13 as

shown in Tables 8 & 10, both of which have not been reported to increase the risk of RA. However, a leucine or valine was observed in 3 out of 54 patients placing them at a high risk of having RA. Nonetheless, very few of the RA cases could be attributed to variations in the shared epitope and positions 11 or 13 indicating that HLA-DRB1 is likely not the major cause for RA in this population.

Setting HLA-DRB1 aside, among the HLA genes, variations in HLA-A has also been shown to impart significant RA risk. A study of European seropositive RA patients demonstrated an additional independent RA-risk in HLA-A amino acid at position 77. A high risk was associated with asparagine at position 77 of the mature protein (Han et al., 2014). Interestingly, this variation was observed to impart protection in HIV controls (Morrison et al., 2007). To assess the association of HLA-A gene with RA in this population, exon 2 and exon 3 of HLA-A was sequenced to identify the allele types and genetic variations in this region. Out of 38 Emirati RA patients, 37 had missense mutation at amino acid 77 where aspartic acid was mutated into asparagine (D>N) whereas, this variation was present in only 28 healthy controls (Table. 18). This placed it at a significant risk for RA in this population with an OR of 20.52 (p-value 0.0001985) (Table 19). Interestingly, two missense variations - Val76Glu and Gly56Arg - with an OR of 6.075 (p-value 0.0001589) and 3.712 (p-value = 0.006273), respectively were also observed in this population (Table. 19). The Val76Glu mutation was observed in 29 out 38 RA patients whereas in controls it was present in only 15 individuals out of 44. Similarly, variant Gly56Arg was observed in 20 patients while only 10 controls had this variation. Amino acid 77 is critical to the function of HLA-A since it is present in binding groove. The amino acid aspartic acid has overall negative charge, while asparagine is a polar amino acid and mostly involve hydrogen bonding with other residues. Therefore, this variation could disrupt the

binding ability of HLA-A. Position 76 is adjacent to this in the binding groove. Valine is a non-polar amino acid while glutamic acid is negatively charged. Like position 77, a hydrophobic to charged amino acid at location 76 is likely to affect the binding characteristics of the binding groove. Structural models illustrate the location and proximity of this amino acid in the binding groove (Figure 27). Position 56 is located near the peptide binding groove of HLA-A (van Deutekom & Keşmir, 2015). A Gly>Arg results in a change from a small achiral amino acid to a large charged amino acid. This is likely to affect protein flexibility and interactions due to the change in size and charge of the amino acid change. While these variations may have a role in RA it is also likely that this may increase an individual's risk towards other comorbid conditions. These variations have not been reported in the literature. Hence, further studies are required to clearly elucidate the role of these variations.

Finally, to evaluate if the genetic sequences of patients and controls cluster separately into one or more groups, several kinds of analyses were performed. Phylogenetic analysis, PCA and fixation index all indicated that patients and controls form a homogeneous group and there is no clear clustering at the gene level (Figures 38-31). Therefore, it appears that variations at specific locations are key to a higher risk of RA.

While this study evaluated two HLA genes that account for the vast majority of RA cases, isolated variations in HLA-B and HLA-DPB1, both at position 9, have also been linked to RA (Han et al., 2014; S. Raychaudhuri et al., 2012). However, due to time and funding limitations, these could not be investigated in this study. Apart from the HLA region, a few other genes have also been associated with RA in this region (Yamamoto et al., 2015). A detailed evaluation of these regions with either conventional Sanger sequencing or Next Generation Sequencing (Chen et al., 2018;

Hosomichi et al., 2015) could be employed to get a comprehensive understanding of how genetic variations could play a role in RA in this population. Genome wide association studies (GWAS) could also be considered using a large sample size to elucidate genetic variations associated with RA in this population.

Chapter 5: Conclusion

RA is widely attributed to an autoimmune response. Variations in several HLA genes are associated with an increased risk for RA. Two specific HLA genes HLA-DRB1 and HLA-A that account for the vast majority of RA cases associated with the HLA region were investigated in this study. The SE hypothesis suggests that specific sequences between amino acids 70-74 of the HLA-DRB1 protein is the leading risk factor. Furthermore, positions 11 and 13 of this protein were also associated with RA. However, this study found extremely low incidence of these high risk variations in the samples collected from Fujairah Hospital in the UAE. Asparagine at position 77 of HLA-A is another leading RA risk factor. In the samples studied, 37 of 38 patients had the asparagine at position 77 suggesting that this variation could be a stronger biomarker for RA in this population. To confirm this finding, and to extend it further, evaluation of a larger sample of both patient and controls is recommended. Furthermore, other HLA and non-HLA genes could also be evaluated to find their association with RA in the Emirati population. Genome-wide approaches and the impact of the environment may also be necessary to gain a holistic understanding of genes, pathways and interactions that impart significant risk of RA in the population of UAE.

References

Al-Mughales, J. A. (2015). Immunodiagnostic significance of anti-RA33 autoantibodies in Saudi patients with rheumatoid arthritis. *Journal of immunology research, 2015*. doi:10.1155/2015/604305

Al-Swailem, R., Al-Rayes, H., Sobki, S., Arfin, M., & Tariq, M. (2006). HLA-DRB1 association in Saudi rheumatoid arthritis patients. *Rheumatology International, 26*(11), 1019.

Arthritis.org. (2019). *Sources of Arthritis Pain | Pain Management & Understanding | Arthritis.org*. [online] Available at: https://www.arthritis.org/living-with-arthritis/pain-management/understanding/types-of-pain.php [Accessed 22 Jun. 2018].

VWR. (2019). *FTA® Elute Cards, Whatman™*. [online] Available at: https://uk.vwr.com/store/product/7997552/fta-elute-cards-whatmantm [Accessed 10 Jan. 2019].

Badsha, H., Kong, K. O., & Tak, P. P. (2008). Rheumatoid arthritis in the United Arab Emirates. *Clinical Rheumatology, 27*(6), 739-742. doi:10.1007/s10067-007-0782-z

Barnetche, T., Constantin, A., Cantagrel, A., Cambon-Thomsen, A., & Gourraud, P. A. (2008). New classification of HLA-DRB1 alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. *Arthritis Research & Therapy, 10*(1), R 26. doi:10.1186/ar2379

Bowes, J., & Barton, A. (2008). Recent advances in the genetics of RA susceptibility. *Rheumatology (Oxford, England), 47*(4), 399-402. doi:10.1093/rheumatology/ken005

Brink, M., Hansson, M., Mathsson-Alm, L., Wijayatunga, P., Verheul, M. K., Trouw, L. A., Holmdahl, R., Rönnelid, J., Klareskog, L., Rantapää-Dahlqvist, S. (2016). Rheumatoid factor isotypes in relation to antibodies against citrullinated peptides and carbamylated proteins before the onset of rheumatoid arthritis. *Arthritis Research & Therapy, 18*(1), 43.

Cantagrel, A., & Degboe, Y. (2016). New autoantibodies associated with rheumatoid arthritis recognize posttranslationally modified self-proteins. *Joint, Bone, Spine: Revue du Rhumatisme, 83*(1), 11-17.

Carmona, L., Villaverde, V., Hernandez-Garcia, C., Ballina, J., Gabriel, R., Laffon, A., & Group, E. S. (2002). The prevalence of rheumatoid arthritis in the general

population of Spain. *Rheumatology (Oxford, England), 41*(1), 88-95. doi:10.1093/rheumatology/41.1.88

Driver, C. (2019). *Gout Attack Symptoms, Causes, Treatment and Diet*. [online] MedicineNet. Available at: https://www.medicinenet.com/gout_pictures_slideshow/article.htm [Accessed 8 Feb. 2019].

Chen, Y. J., Chang, W. A., Wu, L. Y., Hsu, Y. L., Chen, C. H., & Kuo, P. L. (2018). Systematic Analysis of Differential Expression Profile in Rheumatoid Arthritis Chondrocytes Using Next-Generation Sequencing and Bioinformatics Approaches. *International Journal of Medical Sciences, 15*(11), 1129-1142. doi:10.7150/ijms.27056

Choy, E. (2012). Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. *Rheumatology (Oxford, England), 51 Suppl 5*, V 3-11. doi:10.1093/rheumatology/kes113

Coenen, M. J., & Gregersen, P. K. (2009). Rheumatoid arthritis: a view of the current genetic landscape. *Genes and Immunity, 10*(2), 101-111. doi:10.1038/gene.2008.77

Coleparmer.ca. (2015). *Whatman FTA and CloneSaver Cards from Cole-Parmer Canada*. [online] Available at: https://www.coleparmer.ca/p/whatman-fta-and-clonesaver-cards/40496 [Accessed 13 Sep. 2018].

Cush, J. J., Kavanaugh, A., & Weinblatt, M. E. (2010). *Rheumatoid arthritis: early diagnosis and treatment*: Professional Communications.

du Montcel, S.T., Michou, L., Petit-Teixeira, E., Osorio, J., Lemaire, I., Lasbleiz, S., Pierlot, C., Quillet, P., Bardin, T., Prum, B. and Cornelis, F. (2005). New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis and Rheumatism, 52*(4), 1063-1068. doi:10.1002/art.20989

Entezami, P., Fox, D. A., Clapham, P. J., & Chung, K. C. (2011). Historical perspective on the etiology of rheumatoid arthritis. *Hand Clinics, 27*(1), 1-10.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Freeman, J. (2018). *RA vs. OA: Which is Worse – Rheumatoid Arthritis or Osteoarthritis? - RheumatoidArthritis.org*. [online] RheumatoidArthritis.org.

Available at: https://www.rheumatoidarthritis.org/ra/ra-vs-oa/ [Accessed 22 Apr. 2018].

Reference, G. (2009). *Histocompatibility complex*. [online] Genetics Home Reference. Available at: https://ghr.nlm.nih.gov/primer/genefamily/hla [Accessed 10 Mar. 2017].

Reference, G. (2013). *Rheumatoid arthritis*. [online] Genetics Home Reference. Available at: https://ghr.nlm.nih.gov/condition/rheumatoid-arthritis [Accessed 13 May 2017].

Firestein, G. S., & McInnes, I. B. (2017). Immunopathogenesis of rheumatoid arthritis. *Immunity, 46*(2), 183-196.

Funeraldnaid.com. (n.d.). [online] Available at: https://funeraldnaid.com/wp-content/uploads/2016/04/FTA-Technology-Overview.pdf [Accessed 22 Jun. 2019].

Gierut, A., Perlman, H., & Pope, R. M. (2010). Innate immunity and rheumatoid arthritis. *Rheumatic diseases clinics of North America, 36*(2), 271–296. doi:10.1016/j.rdc.2010.03.004

Gonzalez-Gay, M. A., Garcia-Porrua, C., & Hajeer, A. H. (2002). Influence of human leukocyte antigen-DRB1 on the susceptibility and severity of rheumatoid arthritis. *Seminars in Arthritis and Rheumatism, 31*(6), 355-360.

Gorman, J. D., Lum, R. F., Chen, J. J., Suarez-Almazor, M. E., Thomson, G., & Criswell, L. A. (2004). Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis and Rheumatism, 50*(2), 400-412. doi:10.1002/art.20006

Gough, S. C., & Simmonds, M. J. (2007). The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Current genomics, 8*(7), 453–465. doi:10.2174/138920207783591690

Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., & Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics, 63*(6), 325-335. doi:10.1007/s00251-011-0513-0

Gregersen, P. K., Silver, J., & Winchester, R. J. (1987). The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis and Rheumatism, 30*(11), 1205-1213.

Han, B., Diogo, D., Eyre, S., Kallberg, H., Zhernakova, A., Bowes, J., Padyukov, L., Okada, Y., González-Gay, M.A., Rantapää-Dahlqvist, S. and Martin, J.,. (2014). Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *The American Journal of Human Genetics, 94*(4), 522-532.

Holers, V. M. (2013). Autoimmunity to citrullinated proteins and the initiation of rheumatoid arthritis. *Current Opinion in Immunology, 25*(6), 728-735. doi:10.1016/j.coi.2013.09.018

Holoshitz, J. (2010). The rheumatoid arthritis HLA-DRB1 shared epitope. *Current Opinion in Rheumatology, 22*(3), 293-298. doi:10.1097/BOR.0b013e328336ba63

Hosomichi, K., Shiina, T., Tajima, A., & Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *Journal of Human Genetics, 60*(11), 665-673. doi:10.1038/jhg.2015.102

Hoxha, M. (2018). A systematic review on the role of eicosanoid pathways in rheumatoid arthritis. *Advances in Medical Sciences, 63*(1), 22-29.

Huizinga, T. W., Amos, C. I., van der Helm-van Mil, A. H., Chen, W., van Gaalen, F. A., Jawaheer, D., Schreuder, G. M., Wener, M., Breedveld, F. C., Ahmad, N. and Lum, R. F. (2005). Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis and Rheumatism, 52*(11), 3433-3438. doi:10.1002/art.21385

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics, 14*(1), 33-38, 27-38.

Imboden, J. B. (2009). The immunopathogenesis of rheumatoid arthritis. *Annual Review of Pathology, 4*, 417-434. doi:10.1146/annurev.pathol.4.110807.092254

Janeway, C., Murphy, K. P., Travers, P., Walport, M., & Ehrenstein, M. (2008). Janeway's immunobiology.

Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. (1996). *Immunobiology: the immune system in health and disease* (Vol. 7): Current Biology London.

Jun, K. R., Choi, S. E., Cha, C. H., Oh, H. B., Heo, Y. S., Ahn, H. Y., & Lee, K. J. (2007). Meta-analysis of the association between HLA-DRB1 allele and rheumatoid arthritis susceptibility in Asian populations. *Journal of Korean Medical Science, 22*(6), 973-980. doi:10.3346/jkms.2007.22.6.973

Ka, S., Lee, S., Hong, J., Cho, Y., Sung, J., Kim, H. N., Kim, H. L. and Jung, J. (2017). HLAscan: genotyping of the HLA region using next-generation sequencing data. *BMC Bioinformatics, 18*(1), 258. doi:10.1186/s12859-017-1671-3

Kobayashi, K. S., & Van Den Elsen, P. J. (2012). NLRC5: a key regulator of MHC class I-dependent immune responses. *Nature Reviews Immunology, 12*(12), 813.

Kochi, Y., Suzuki, A., & Yamamoto, K. (2014). Genetic basis of rheumatoid arthritis: a current review. *Biochemical and Biophysical Research Communications, 452*(2), 254-262.

Koehn, C., Palmer, T., & Esdaile, J. (2002). *Rheumatoid arthritis: plan to win*: Oxford University Press.

Konda Mohan, V., Ganesan, N., Gopalakrishnan, R., & Venkatesan, V. (2017). HLA-DRB1 shared epitope alleles in patients with rheumatoid arthritis: relation to autoantibodies and disease severity in a south Indian population. *International Journal of Rheumatic Diseases, 20*(10), 1492-1498. doi:10.1111/1756-185x.12948

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution, 33*(7), 1870-1874. doi:10.1093/molbev/msw054

Kurko, J., Besenyei, T., Laki, J., Glant, T. T., Mikecz, K., & Szekanecz, Z. (2013). Genetics of rheumatoid arthritis - a comprehensive review. *Clinical Reviews in Allergy and Immunology, 45*(2), 170-179. doi:10.1007/s12016-012-8346-7

Lee, H. S., Lee, A. T., Criswell, L. A., Seldin, M. F., Amos, C. I., Carulli, J. P., Navarrete, C., Remmers, E. F., Kastner, D. L., Plenge, R. M. and Li, W. (2008). Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Molecular Medicine, 14*(5-6), 293-300. doi:10.2119/2007-00123.Lee

Lupus Foundation of America. (2015). *How lupus differs from arthritis*. [online] Available at: https://www.lupus.org/resources/how-lupus-differs-from-arthritis [Accessed 15 Apr. 2017].

MacGregor, A. J., Snieder, H., Rigby, A. S., Koskenvuo, M., Kaprio, J., Aho, K., & Silman, A. J. (2000). Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis and Rheumatism, 43*(1), 30-37. doi:10.1002/1529-0131(200001)43:1<30::aid-anr5>3.0.co;2-b

Marsh, S. (2011). Nomenclature for factors of the HLA system, update July 2010. *International journal of immunogenetics, 38*(1), 83-87.Matzaraki, V., Kumar, V., Wijmenga, C., & Zhernakova, A. (2017). The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology, 18*(1), 76.

Mayo Clinic. (2018). *Psoriatic arthritis - Symptoms and causes*. [online] Available at: https://www.mayoclinic.org/diseases-conditions/psoriatic-arthritis/symptoms-causes/syc-20354076 [Accessed 15 Feb. 2018].

Mizuki, N., & Kimura, M. (1996). [Gene structure of the human MHC region]. *Nihon Rinsho. Japanese Journal of Clinical Medicine, 54*(6), 1705-1717.

Morgan, A. W., Haroon-Rashid, L., Martin, S. G., Gooi, H. C., Worthington, J., Thomson, W., Barrett, J. H. and Emery, P. (2008). The shared epitope hypothesis in rheumatoid arthritis: evaluation of alternative classification criteria in a large UK Caucasian cohort. *Arthritis and Rheumatism, 58*(5), 1275-1283. doi:10.1002/art.23432

Morrison, A. C., Bare, L. A., Chambless, L. E., Ellis, S. G., Malloy, M., Kane, J. P., Pankow, J. S., Devlin, J. J., Willerson, J. T. and Boerwinkle, E. (2007). Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology, 166*(1), 28-35. doi:10.1093/aje/kwm060

Muazzam, A. G., Mansoor, A., Ali, L., Siddiqi, S., Hameed, A., Ajmal, M., & Mazhar, K. (2013). Association of HLA-DRB1 and -DQB1alleles and haplotypes with rheumatoid arthritis in a Pakistani population. *Arthritis Research & Therapy, 15*(4), R 95. doi:10.1186/ar4275

Mullen, M. P., Howard, D. J., Powell, R., & Hanrahan, J. P. (2009). A note on the use of FTA™ technology for storage of blood samples for DNA analysis and removal of PCR inhibitors. *Irish Journal of Agricultural and Food Research, 48*(1), 109-113.

Nakken, B., Papp, G., Bosnes, V., Zeher, M., Nagy, G., & Szodoray, P. (2017). Biomarkers for rheumatoid arthritis: From molecular processes to diagnostic applications-current concepts and future perspectives. *Immunology Letters, 189*, 13-18.

Ndunguru, J., Taylor, N. J., Yadav, J., Aly, H., Legg, J. P., Aveling, T., Thompson, G. and Fauquet, C. M. (2005). Application of FTA technology for sampling, recovery and molecular characterization of viral pathogens and virus-derived transgenes from plant tissues. *Virology Journal, 2*(1), 45.

O'Rielly, D. D., & Rahman, P. (2010). Pharmacogenetics of rheumatoid arthritis: Potential targets from susceptibility genes and present therapies. *Pharmacogenomics and Personalized Medicine, 3*, 15-31.

Oka, S., Furukawa, H., Kawasaki, A., Shimada, K., Sugii, S., Hashimoto, A., Komiya, A., Fukui, N., Ito, S., Nakamura, T. and Saisho, K. (2014). Protective effect of the HLA-DRB1* 13: 02 allele in Japanese rheumatoid arthritis patients. *PloS One, 9*(6), e99453.

Okada, Y., Kim, K., Han, B., Pillai, N. E., Ong, R.T.H., Saw, W. Y., Luo, M., Jiang, L., Yin, J., Bang, S. Y. and Lee, H. S. (2014). Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Human Molecular Genetics, 23*(25), 6916-6926. doi:10.1093/hmg/ddu387

Ollier, W., & Thomson, W. (1992). Population genetics of rheumatoid arthritis. *Rheumatic Diseases Clinics of North America, 18*(4), 741-759.

Parliament.vic.gov.au. (n.d.). [online] Available at: https://www.parliament.vic.gov.au/images/stories/committees/lawrefrom/forensics/submissions/Forensics-Whatman_2.pdf [Accessed 15 May 2017].

Raychaudhuri, S., Sandor, C., Stahl, E. A., Freudenberg, J., Lee, H. S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J. and Siminovitch, K. A. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genetics, 44*(3), 291.

Reynolds, R. J., Ahmed, A. F., Danila, M. I., Hughes, L. B., Consortium for the Longitudinal Evaluation of African Americans with Early Rheumatoid Arthritis Investigators, Gregersen, P. K., Raychaudhuri, S., Plenge, R. M. and Bridges Jr, S. L. (2014). HLA–DRB1–associated rheumatoid arthritis risk at multiple levels in African Americans: hierarchical classification systems, amino acid positions, and residues. *Arthritis & Rheumatology, 66*(12), 3274-3282.

Ruyssen-Witrand, A., Van Steenbergen, H. W., Van Heemst, J., Gourraud, P. A., Nigon, D., Lukas, C., Miceli-Richard, C., Jamard, B., Cambon-Thomsen, A., Cantagrel, A. and Dieudé, P. (2015). A new classification of HLA-DRB1 alleles based on acid-base properties of the amino acids located at positions 13, 70 and 71: impact on ACPA status or structural progression, and meta-analysis on 1235 patients with rheumatoid from two cohorts (ESPOIR and EAC cohort). *RMD Open, 1*(1), e000099. doi:10.1136/rmdopen-2015-000099

Sangha, O. (2000). Epidemiology of rheumatic diseases. *Rheumatology, 39*(suppl_2), 3-12. doi:10.1093/rheumatology/39.suppl_2.3

Santos, G. d. C. (2018). FTA Cards for Preservation of Nucleic Acids for Molecular Assays: A Review on the Use of Cytologic/Tissue Samples. *Archives of Pathology and Laboratory Medicine, 142*(3), 308-312. doi:10.5858/arpa.2017-0303-RA

Saxena, R., Plenge, R.M., Bjonnes, A.C., Dashti, H.S., Okada, Y., Gad El Haq, W., Hammoudeh, M., Al Emadi, S., Masri, B.K., Halabi, H. and Badsha, H. (2017). A Multinational Arab Genome-Wide Association Study Identifies New Genetic Associations for Rheumatoid Arthritis. *Arthritis & Rheumatology, 69*(5), 976-985. doi:10.1002/art.40051

Scally, S. W., Petersen, J., Law, S. C., Dudek, N. L., Nel, H. J., Loh, K. L., Wijeyewickrema, L. C., Eckle, S. B., van Heemst, J., Pike, R. N. and McCluskey, J. (2013). A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *Journal of Experimental Medicine, 210*(12), 2569-2582.

Small-Molecule Drug Discovery Suite 2019-2, Schrödinger, LLC, New York, NY, 2019.

Scrivo, R., Di Franco, M., Spadaro, A., & Valesini, G. (2007). The immunology of rheumatoid arthritis. *Annals of the New York Academy of Sciences, 1108*(1), 312-322.

Shimane, K., Kochi, Y., Suzuki, A., Okada, Y., Ishii, T., Horita, T., Saito, K., Okamoto, A., Nishimoto, N., Myouzen, K. and Kubo, M. (2013). An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. *Rheumatology (Oxford, England), 52*(7), 1172-1182. doi:10.1093/rheumatology/kes427

Shlotzhauer, T. L. M. D., McGuire, J. L. M. D., & Ziminski, C. M. M. D. (2014). *Living with Rheumatoid Arthritis* (3rd edition ed.). Baltimore, UNITED STATES: Johns Hopkins University Press.

Smolen, J. S., Aletaha, D., & McInnes, I. B. (2016). Rheumatoid arthritis. *The Lancet, 388*(10055), 2023-2038. doi:10.1016/S0140-6736(16)30173-8

Thorsby, E. (2009). A short history of HLA. *Tissue Antigens, 74*(2), 101-116. doi:10.1111/j.1399-0039.2009.01291.x

Trier, N., Izarzugaza, J., Chailyan, A., Marcatili, P., & Houen, G. (2018). Human MHC-II with Shared Epitope Motifs Are Optimal Epstein-Barr Virus Glycoprotein 42 Ligands-Relation to Rheumatoid Arthritis. *International Journal of Molecular Sciences, 19*(1). doi:10.3390/ijms19010317

Uçar, F., Karkucak, M., Alemdaroğlu, E., Çapkin, E., Yücel, B., Sönmez, M., Tosun, M. and Karaca, A. (2012). HLA-DRB1 allele distribution and its relation to rheumatoid arthritis in eastern Black Sea Turkish population. *Rheumatology International, 32*(4), 1003-1007.

Universite De Geneve; Faculte De Medecine (2015). *HLA genes and molecules*. [online] Medweb4.unige.ch. Available at: http://medweb4.unige.ch/immunologie/home/HSC/transplantation_barriers/structures_recognized/major_histocompatibility_antigens/genetic.php [Accessed 2 Jun. 2019].

van Boekel, M. A., Vossenaar, E. R., Van den Hoogen, F. H., & van Venrooij, W. J. (2001). Autoantibody systems in rheumatoid arthritis: specificity, sensitivity and diagnostic value. *Arthritis Research & Therapy, 4*(2), 87.

van der Woude, D., & Catrina, A. I. (2015). HLA and anti-citrullinated protein antibodies: Building blocks in RA. *Best practice & research Clinical rheumatology, 29*(6), 692-705.

van Deutekom, H. W., & Keşmir, C. (2015). Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics, 67*(8), 425-436. doi:10.1007/s00251-015-0849-y

van Drongelen, V., & Holoshitz, J. (2017). Human Leukocyte Antigen-Disease Associations in Rheumatoid Arthritis. *Rheumatic Diseases Clinics of North America, 43*(3), 363-376. doi:10.1016/j.rdc.2017.04.003

Verheul, M., Fearon, U., Trouw, L., & Veale, D. (2015). Biomarkers for rheumatoid and psoriatic arthritis. *Clinical Immunology, 161*(1), 2-10.

Viatte, S., Plant, D., Han, B., Fu, B., Yarwood, A., Thomson, W., Symmons, D.P., Worthington, J., Young, A., Hyrich, K.L. and Morgan, A.W. (2015). Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA, 313*(16), 1645-1656.

Warren, E. H., Zhang, X. C., Li, S., Fan, W., Storer, B. E., Chien, J. W., Boeckh, M. J., Zhao, L. P., Martin, P. J. and Hansen, J. A. (2012). Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. *Blood, 120*(14), 2796-2806. doi:10.1182/blood-2012-04-347286

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. doi:10.1093/bioinformatics/btp033

Weisman, M. H. (2011). *Rheumatoid Arthritis*. Cary, United States: Oxford University Press, Incorporated.

Wennerström, A., Vlachopoulou, E., Lahtela, L. E., Paakkanen, R., Eronen, K. T., Seppänen, M., & Lokki, M.-L. (2013). Diversity of extended HLA-DRB1 haplotypes in the Finnish population. *PloS One*, *8*(11), e79690.

West, S. (2014). *Rheumatology Secrets E-Book*: Elsevier Health Sciences.

Weyand, C. M., & Goronzy, J. J. (2000). Association of MHC and rheumatoid arthritis: HLA polymorphisms in phenotypic variants of rheumatoid arthritis. *Arthritis Research & Therapy*, *2*(3), 212.

Yamamoto, K., Okada, Y., Suzuki, A., & Kochi, Y. (2015). Genetic studies of rheumatoid arthritis. *Proceedings of the Japan Academy, Series B*, *91*(8), 410-422.