Association of Interleukin-23 Receptor (IL-23R) gene variants with Gout and Rheumatic Heart Fever (RHF)

Keresoma Leaupepe

Supervisor: Associate Professor Tony Merriman

A thesis submitted for the degree of Biomedical Science with Honours (BBiomedSc (Hons)) University of Otago, Dunedin, New Zealand

October 27
Abstract

Gout is an increasingly common form of inflammatory arthritis caused by the deposition of urate, leading to the formation of monosodium urate (MSU) crystals in joints and other body tissue. This results in subsequent recurrent acute inflammation attacks (1). Gout prevalence is increasing worldwide and has a particularly high prevalence in the Māori and Pacific populations of New Zealand (9.3 to 13.9% of Māori men and 14.9% of Pacific Island men) (2). Risk factors for gout development can be either genetic or environmental. The risk of gout is different between ancestral groups, suggesting that they have genetic differences (3).

Rheumatic heart fever (RHF) is a systemic auto-inflammatory disease that is caused by infection of the upper respiratory tract (mainly the throat) by group A β-haemolytic streptococci (GABHS). RHF happens via antigen molecular mimicry and cross reactivity mechanisms between the host and bacteria. Cross reactivity of antibodies and/or T cells stimulates recognition between the S.pyrogenes peptides and the host protein and leads to inflammation and autoimmunity (4). RHF incidence and prevalence has steadily declined in developed countries since the early 1900s. However, it remains a leading cause of morbidity and mortality among young individuals (6 – 15 years) in developing countries. The risk of RHF can be familial or environmental e.g. as poor housing conditions, crowding, and poor health knowledge (5, 6).

The IL23R gene codes for the interleukin 23 receptor. The receptor is located on the cell membrane of cells that are involved in the immune system, which provide defence mechanisms against infection and disease from foreign microbes. During the TH17 immune response, activation of IL23R from interaction with its subunit (IL23) initiates inflammation (7). Previous studies have shown that genetic variants derived from IL23R are associated with auto-inflammatory related diseases, such as rheumatoid arthritis, ankylosing spondylitis and inflammatory bowel disease. A study by Liu et al (2015) showed that the IL23R SNP rs7517847 minor allele G confers an association with gout in Chinese Han male population.
Our aim was to test IL23R gene variants (rs11209026, rs7517847 and rs11465804) for association with gout and RHF in European and Polynesian populations using case-control sample sets recruited within New Zealand and (for gout) Europe. To test this hypothesis, SNPs were genotyped using Taqman assay and statistical analysis was carried out using R studio logistic regression to test for association of SNPs with gout and RHF. Common confounders including ancestry, sex and age were adjusted for in the regression analysis.

Gout results revealed that rs11465804 and rs11209026 in both European and Polynesian were not significantly associated with gout. However, the rs7517847 minor allele (G) showed a significant association with gout in Polynesian (Polynesian OR = 0.85, P = 0.04) (European OR = 0.94, P = 0.53), which is consistent with the Lui et al (2015) findings. These data replicate the Liu et al (2015) findings and support the claim that IL23R has a causal role in gout in people with Polynesian ancestry. Hence, the IL23R pathway is a target for gout treatment in the Polynesian population.

RHF results revealed that only SNP rs11209026 shows evidence of association with a protective effect for the minor allele (A) (OR = 0.07, P-value = 0.002) after adjustment. Therefore the rs11209026 major allele (G) is in a susceptible direction. This provides evidence that IL23R has a casual role in RHF development risk in Polynesian people.
Acknowledgment

First and foremost, I would like to thank Almighty God for His immense blessings and for making my dreams come true.

Thank you to everyone who supported and helped me during the course of this study. I will always look back on this time in my life caringly, and feel grateful for the care and encouragement I received from so many people especially my family, relatives and friends. I would like to express my sincerest gratitude to my supervisor Associate Professor Tony Merriman for his enthusiasm, advice, the knowledge and patience throughout this project. Secondly, I would like to say thank you to the very knowledgeable and thoughtful lab staff, Ruth, Mandy, Murray and Marilyn, who helped me a great deal with my bench work and analyses. Thirdly, thanks to the students in the Merriman lab Hakimeh, Tanya, Amara, Humaira, Jarrod and Tom for being great company and fantastic to work alongside, and also for their help with my project. In particular, I would like to thank Mandy for her assistance throughout my project and for always being so generous with her time.

I would also like to thank to my wonderful family. Thank you to my parents (Keneti and Faaleleiga Leaupepe) and my siblings (Apela, Eleasaro, Hini and Sini) for their endless support during my mission to complete the write-up of this thesis. Also like to say thank you to NZAID for the scholarship and sponsoring my financial supports.

Special thanks to my tutor Anna Bundock for proof-reading my thesis chapters. Last but by no means least, I thank to my girlfriend Rosemarie for her non stop support during this academic years.
# Table of Contents

**Chapter 1: Introduction**

1.1 Complex Diseases ........................................................................................................16
1.1.1 Genetic Association Studies ..................................................................................16
1.1.2 Auto-inflammatory & Autoimmune Disease .........................................................17
1.2 Gout .............................................................................................................................18
1.2.1 Global Epidemiology of Gout ..............................................................................20
1.2.2 Gout Prevalence in New Zealand ..........................................................................21
1.2.3 Genetic differences between ethnicity sub-groups in NZ .......................................21
1.2.4 Synthesis of Urate ...............................................................................................21
1.2.5 Loss of Uricase Enzyme in Evolution ...................................................................22
1.2.6 Absence of Uricase Advantages .........................................................................23
1.2.7 Maintenance of Urate Levels ..............................................................................23
1.2.8 Urate Transporters ..............................................................................................24
1.2.9 Gout Stages ..........................................................................................................26
1.2.10 Gout Risk Factors ..............................................................................................28
1.2.11 MSU Crystal and Inflammation .........................................................................30
1.3 Gout Complications ....................................................................................................33
1.4 Gout Treatment ............................................................................................................33
1.4.1 Lifestyle Treatment .............................................................................................33
1.4.2 Medication Treatment ..........................................................................................34
1.5 Rheumatic Heart Fever ............................................................................................35
1.5.1 Global Epidemiology of RHF .............................................................................35
1.5.2 New Zealand Epidemiology of RHF ....................................................................38
1.5.3 Signs and Symptoms ............................................................................................39
1.5.4 Pathophysiology of RHF ......................................................................................40
Chapter 1: Introduction

1.5.5 Diagnosis ................................................................. 42
1.5.6 Prevention .................................................................. 43
1.5.7 Treatment ................................................................... 44

1.6 Interleukin-23 Receptor Complex ........................................ 45
   1.6.1 IL-23R genetic variants ............................................. 46

1.7 Aim of this Study ................................................................ 51

Chapter 2: Methods ................................................................ 52
1.8 Gout & RHF Case-Control Study ......................................... 53
1.9 Participant Recruitment & Ethical Approval ......................... 53
1.10 Study Population ................................................................ 53
1.11 Extraction of DNA from blood samples ............................. 54
   1.11.1 Lysis of the Red blood cell (RBC) ............................... 54
   1.11.2 Lysis of White blood cell (WBC) and DNA extraction ... 54
1.12 Quality control .................................................................. 55
1.13 SNP Selection ................................................................... 56
1.14 Constructing Linkage Disequilibrium (LD) using Haploview .......... 56
1.15 Haplotype Analysis........................................................... 57
1.16 TaqMan® genotyping Assay .............................................. 58
   1.16.1 TaqMan® Assay mechanism ...................................... 59
   1.16.2 Setting up the 384-well plate for TaqMan® genotyping using Light Cycler .............. 61
   1.16.3 Using the Light Cycler ............................................. 62
1.17 Entry of TaqMan® genotyping Data into Haploped File .......... 63
1.18 BC|SNPmax ..................................................................... 64
   1.18.1 Uploading SNPs Genotypes file on the BC|SNPmax .................. 64
   1.18.2 Affection data sets files ............................................ 65
1.19 Statistical analysis ............................................................ 65
   1.19.1 Hardy-Weinberg Equilibrium test using SHEsis software ...... 65
1.19.2 Logistic Regression using R studio software ................................................................. 67
1.19.3 R Studio Software ........................................................................................................ 68
1.19.4 Meta-Analysis ............................................................................................................ 70
1.19.5 Secondary Analysis ..................................................................................................... 72
1.19.6 Power Calculation ....................................................................................................... 74

Chapter 3: Results .................................................................................................................. 76

1.20 Association of IL23R SNPs of interest with Gout & Rheumatic Heart Fever ................. 77
1.20.1 Linkage Disequilibrium (LD) Plots of SNPs ($r^2$) ....................................................... 77

1.21 Gout Association Analysis .............................................................................................. 77
1.21.1 Meta-Analysis ............................................................................................................. 78
1.21.2 Secondary Analysis ..................................................................................................... 83
1.21.3 Power Calculation for Gout Dataset .......................................................................... 83

1.22 Rheumatic Heart Fever Association Analysis ............................................................... 84
1.22.1 Power Analysis for RHF Dataset ................................................................................ 85

1.23 Haplotype Analysis ........................................................................................................ 87

Chapter 4: Discussion .............................................................................................................. 88

1.24 Association Analysis of Gout Dataset ........................................................................... 89
1.24.1 RS11465804 & RS11209026 ....................................................................................... 90
1.24.2 RS7517847 ................................................................................................................ 91

1.25 Secondary Analysis ....................................................................................................... 93
1.26 Haplotype Analysis ........................................................................................................ 93

1.27 Association Analysis Rheumatic Heart Fever Dataset .................................................. 94
1.27.1 RS11465804 ............................................................................................................... 95
1.27.2 Rs7517847 ................................................................................................................ 95
1.27.3 RS11209026 ............................................................................................................... 96

1.28 Limitations in the Study ................................................................................................. 96

1.29 Conclusions .................................................................................................................. 97
Appendix ................................................................................................................................. 98

1.30 Appendix 1 ......................................................................................................................... 98
  1.30.1 R-studio command ......................................................................................................... 98

1.31 Appendix 2 ........................................................................................................................ 101
  1.31.1 Meta-Analysis Command: ......................................................................................... 101

1.32 Appendix 3 ........................................................................................................................ 103
  1.32.1 Secondary Analysis R Studio Command ...................................................................... 103

1.33 Appendix 4 ........................................................................................................................ 104
  1.33.1 Haplotype PLINK Command ..................................................................................... 104

References .................................................................................................................................. 106
List of Figures

Figure 1.2:1 James Gillray, 1799 painting of gout illustrating a foot being attacked. .................................................................19

Figure 1.2:2 Anatomy of the kidney, showing the structure of a nephron, made up of the glomerulus (the filter) and a series of tubules (tubule system). ..................................................................................24

Figure 1.2:3 Uric acid transport in the proximal renal tubule of the kidney by urate transporters. ...26

Figure 1.2:4 NALP3 inflammasome complex during gout inflammation. .................................................................32

Figure 1.5:1 Map presenting global prevalence of RHF from 1991 to present. (1) Fiji (2) Samoa (3) Tonga. ........................................................................................................................................37

Figure 1.5:2 Trends of RHF prevalence per 1000 persons from 1970 to 2008 of different region. ...38

Figure 1.5:3 NZ map illustrating geographic distribution of RHF prevalence per 100,000 in 2003. 39

Figure 1.6:1 IL23-IL23R Pathway .................................................................................................................................46

Figure 1.14:1 Linkage disequilibrium (LD) plot of 3 SNPs of interest in the IL23R gene in European population. ........................................................................................................................................57

Figure 1.16:1 Schmatic diagram illustrating an overview of TaqMan® genotyping assay. .............60

Figure 1.16:2 Example of cluster plot from the TaqMan® genotyping assay of the 3 SNP of interest. ..................................................................................................................................................62

Figure 1.17:1 Example of the master Haploped file. .................................................................................63

Figure 1.18:1 Uploading of genotypes file onto BC|SNPmax .................................................................65

Figure 1.19:1 Example calculation of HWE using SHEsis ........................................................................66

Figure 1.19:2 Example of SHEsis HWE calculation result of SNP rs11465804 .................................67

Figure 1.19:3 Example of input txt file for logistic regression analysis of SNP using R. .............68

Figure 1.19:4 Example of a Meta-analysis forest plot .........................................................................................71

Figure 1.19:5 An example of a meta-analysis output where the P-value for heterogeneity (red square). ......................................................................................................................................................72

Figure 1.19:6 Example of input txt file for secondary analysis of SNP using R .............................73

Figure 1.19:7 Power calculation figures generated on excel. .................................................................75
Figure 1.20: Linkage Disequilibrium (LD) plots that were generated from Haploview.

Figure 1.21: Meta-analysis plot for 3 SNPs of interest combining European and Polynesian data set in gout.
List of Tables

Table 1.6:1 A summary of Auto-inflammatory & Autoimmune disease that are associated with common IL23R SNPs. ........................................................................................................50
Table 1.15:1 Example of a input Map file. ........................................................................58
Table 1.15:2 Example of input Ped file. ..............................................................................58
Table 1.16:1 Volume of Master Mix per Reaction. ...............................................................61
Table 1.18:1 Example of SNP genotyping file for BC|SNPmax upload.................................64
Table 1.19:1 Example of an input file for meta-analysis of a SNP in European and Polynesian......70
Table 1.21:1 Minor allele frequency of 3 IL23R SNPs in Polynesian sub-groups (EP, WP and EPWP) for Gout and RHF sample set.........................................................................................79
Table:1.21:2 Meta-Analysis and Allele/Distribution data of IL23R SNP of interest for Gout Dataset. ........................................................................................................................................80
Table: 1.21:3 Allele/Genotype distribution and Odd Ratios/P-value of secondary analysis. ........83
Table 1.21:4 Power Output of Gout Dataset........................................................................84
Table 1.22:1 Power output for RHF in Polynesian Dataset. .......................................................85
Table 1.22:2 Allele and distribution data of IL23R SNPs for RHF dataset.................................86
Table 1.23:1 European Haplotype Result in Gout.................................................................87
Table 1.23:2 Polynesian Haplotype Result in Gout...............................................................87
Table 1.30:1 Example of an Genotype frequency table showing homozygous minor (G/G represent by 3), heterozygous (G/T represent by 2) and homozygous major (T/T represent by 1)...............98
Table 1.30:2 Example of a Allele frequency table show count of major and minor allele. .........99
Table 1.30:3 Result of unadjusted P-value (Pr (>l z l) of RS11465804_Genotype (Highlight yellow) ................................................................................................................................................99
Table 1.30:4 Result of unadjusted odds ratio (Highlight yellow) of RS11465804_Genotype .......99
Table:1.30:5 Result of unadjusted 95% (Highlight yellow) CI RS11465804_Genotype .............100
Table 1.30:6 Result of adjusted P-value (Highlight yellow) of RS11465804_Genotype ............100
Table 1.30:7 Result of adjusted Odd Ratio (Highlight yellow) of RS11465804_Genotype after adjustment

Table 1.31:1 Odd ratio, 95% CI interval, %W(fixed) and %W(random) value of NZ European and NZ Polynesian population individually.

Table 1.31:2 The combine result of 2 population. Showing Fixed and Random effects OR, 95% CI interval, z value and p-value.

Table 1.31:3 Heterogeneity test for 2 population.

Table:1.32:1 Result of P-value (Pr (> |z|) of RS11465804_Genotype (Highlight yellow).

Table:1.32:2 Result of unadjusted odds ratio (Highlight yellow) of RS11465804_Genotype.

Table:1.32:3 Result of unadjusted 95% (Highlight yellow) of CI Urate_Cau_rs112.
Chapter 1: Introduction
1.1.1 Complex Diseases

Complex diseases or polygenetic disease are diseases that are not only influenced by inherited genetic variants but also environmental factors (8). Common known complex diseases includes diabetes, obesity, gout and cardiovascular disease e.g. rheumatic heart disease. Complex diseases have multiple genetic variants associated with risk of the disease. Understanding the biological basis as well as the mechanism of action of these genetic variants is a crucial step for developing proper and effective treatments and interventions for these diseases. Unfortunately, finding these genetic variants is not easy. Known common variants typically only account for ~10% of heritability variance with considerable heritability variance as not yet identified (9).

1.1.2 Genetic Association Studies

A genetic association study can help explain the biological basis of a complex disease (10). These studies are conducted to investigate if a certain genetic variant has a significant association with a disease (11). There are two types of genetic association studies: (1) family-based studies and (2) population-based studies, which are commonly known as case-control studies (12).

A family-based type study determines a inheritable genetic variants from parents to their offspring (13). An allele that increases the risk of developing disease, is likely to be inherited by the offspring in a population with the disease. A population based study measures the frequencies and distribution of single-nucleotide polymorphism (SNP) alleles between cases and controls. Cases refer to those who have the disease and controls refer to unaffected individuals randomly recruited from the selected population. If a SNP allele (especially a minor allele) frequency is increased in cases, the risk of disease development is also increased. A gene variant that has a functional effect on a certain trait or has a linkage disequilibrium (LD) with another functional variants will likely have an association with a certain phenotype.
1.1.3 Auto-inflammatory & Autoimmune Disease

Auto-inflammatory disease is often thought to be autoimmune disease. Recent studies that were previously known as autoimmune disease are now known to be into auto-inflammatory disease. Both, auto-inflammatory and autoimmune disease have an impaired immune system response. Auto-inflammatory diseases have a malfunctioning innate immune system while autoimmune diseases patients have an impaired adaptive immune system (14, 15). Therefore, their causal mechanism is different, as well as their treatment options, long-term health risks, and potential complications. However, sometimes both diseases can share the same symptoms (14).

1.1.3.1 Auto-inflammatory disease

Auto-inflammatory disease is due to innate immune dysfunction. The immune system becomes activate without cause (meaning the immune system responds without antibodies to activate it). As a result, the innate immune reaction becomes over reactive. The over reactive response induces varying degrees of inflammation that can happen anywhere in the body. Fever is a common symptom of this effect. Over reactive inflammation can also initiate disease symptoms in muscles, joints, skin and internal organs at the same time. This includes disease clinical gout, inflammatory bowel disease, Crohn’s disease, ankylosin spondylitis, ulcerative colitis, osteoarthritis and psoriatic arthritis. Auto-inflammatory disease can be caused by genetic mutation which induces regulation impairment of pro-inflammatory cytokines, including interleukin-12 and TNF-alpha (14-17).

1.1.3.2 Auto-immune disease

Auto-immune disease occurs when the adaptive or acquired part of the immune system has mistakenly attacked something that is identified as a threat to the body. The autoimmune mechanism allows the immune response to release antibodies and T cells that attack its own tissues and induce inflammation. Inflammation often occurs at an initial inflammation site, but it can progress systemically later on to other body sites. Examples of autoimmune disease are rheumatic heart fever (RHF), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), sjogren syndrome, hashimoto thyroiditis, juvenile (type 1) diabetes, polymyositis, multiple sclerosis and scleroderma (18-20).
1.2 Gout

First documented by the Egyptians in 2640 BC, gout was one of the earliest diseases to be considered as a clinical entity. Gout is an increasingly common form of inflammatory arthritis caused by the deposition of urate, leading to the formation of monosodium urate (MSU) crystals in both joints and other body tissue. This results in subsequent recurrent acute inflammation attacks (21). Characteristics of gout include red, tender, hot and painful swollen joints. Its occurrence is mostly restricted to the big toe or the first metatarsophalangeal joints; but other joints such as fingers, knees, heels and wrists may also be affected. The slow-forming urate deposits in and around the joints (tophi) result in a deformed appearance. If left untreated, recurrent gout attacks may become severe and can lead to the development of chronic gout, characterized by permanent deformation of joints (22-25).

Clinically, gout can be categorised as either primary or secondary, depending on its aetiology. Formation of gout from high levels of serum urate in the blood (hyperuricemia) is classified as primary gout, and is due to overproduction of serum urate or/and under excretion of uric acid. In addition, abnormal urate metabolism, via consumption of purine-rich foods such as alcohol and seafood, is also a cause of primary gout, due to the conversion of purines to uric acid (26, 27). Secondary gout is associated with the effects of other medical conditions and chronic comorbidities, including diabetes, kidney disease, obesity, cardiovascular disease and metabolic syndrome. Moreover, side effects from medications such as diuretics, aspirin and cyclosporine are also categorized as a cause of secondary gout (28-30).
Figure 1.2: James Gillray, 1799 painting of gout illustrating a foot being attacked.

The image delivers a feeling of extreme pain caused by gout (Taken from The Gout by James Gillray. Published May 14th 1799., public domain)
1.2.1 Global Epidemiology of Gout

The prevalence of gout has increased dramatically worldwide over recent times, and varies greatly depending on a number of genetic, environment and physiological factors (31). Studies have shown that the Polynesian (Māori /Pacific) population has the highest prevalence worldwide, and that the prevalence in the European population is increasing (32). Gout is an age-related disease, commonly affecting older people (> 45), and is found to be more prevalent in men and women that have undergone menopause. The prevalence peak is at an age range of 75 – 84 with a male: female prevalence ratio of approximately 3 – 4:1 (33).

Previous studies have shown that approximately 3-4% of European men suffer from gout in westernized countries and that in 2007 and 2008 8.3 million adults in the U.S had gout (34-36). Buckley (2011), reported that the prevalence of gout in Asian ethnicities in China was 3.6/1,000 in 2002 and 5.3/1,000 in 2004 (32, 37, 38). It has been documented that the prevalence in modern Thai and Chinese was historically low prior to 1958, but has recently increased from 0.46% to 0.83% in men and 0 to 0.23% in women (39). Moreover, other Asian populations from Vietnam, Jammu, Thailand and India have a prevalence ranging from 0.14 – 0.19% with a significantly higher prevalence among the Aboriginal than non-aboriginal population from Taiwan (9.1% and 0.3% respectively) (32, 40-42).

The increasing prevalence has been largely attributed to changes in lifestyle, diet and, comorbidities, with the differential prevalence thought to be a result of different environmental and genetic predisposition factors that different ethnic groups are exposed to. For example, the different gout prevalence in Asian groups’ compared to European populations is thought to be due to changes of lifestyle, increased longevity and different diet composition between the two groups. Genetic differences are also to be expected between Chinese and European. As for Taiwanese Aboriginal and Polynesian populations, increasing prevalence is attributed to a genetic predisposition for high serum urate concentrations.
1.2.2 Gout Prevalence in New Zealand

In NZ Māori populations, the prevalence of gout is thought to have increased since the 1840’s, when Europeans arrived in NZ. The prevalence increasing is thought to be due to lifestyle components brought to NZ by the Western people. Following their arrival, the percentage of NZ Māori (6.06%) and Pacific people (7.63%) over 20 years of age that have gout is now high (43). People of Polynesian decent also have the highest incidence of chronic gout, tophaceous disease and related joint inflammation disorders (44).

1.2.3 Genetic differences between ethnicity sub-groups in NZ

Information about Pacific migratory patterns is crucial for genetic association studies of clinical gout as multiple genetic variants from various genes involve in serum urate regulation and more importantly, genetic variant frequency and distribution differ between different populations. This means they have differences in biological basis as well a molecular processes underlying gout. The colonization of South Pacific islands during migratory events by Western countries not only caused a bottleneck effect on Pacific population but also established genetic differences between Western Polynesian (people from Tonga, Samoa, Niue, Tokelau) and Eastern Polynesian (people from Māori, Cook Island, Hawaii, New Zealand) (45).

1.2.4 Synthesis of Urate

Urate is synthesised from the catabolism of ribose-5-phosphate, fructose and nucleic acid.

1.2.4.1 Ribose-5-phosphate

Ribose-5-phosphate transformed into phospho-ribosyl pyrophosphate (PRPP), will then convert later on into Inosine monophosphate (IMP) (46). IMP commonly acts as an intermediate molecule in purine biosynthesis. It has the ability to transform into guanosine monophosphate (GMP) or adenosine monophosphate (AMP) when necessary (47). AMP and GMP nucleotides are used for deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) synthesis.
1.2.4.2 Nucleic Acids

The GMP nucleotide is transformed into guanosine. The guanosine is then degraded via a 2 step process. Firstly, the guanosine phosphorylase enzyme initiates production of free guanine by phosphorylating the nucleoside. Then the deamination process where guanine deaminase enzyme converts the free guanine to xanthine. Conversely, the AMP nucleotide converts to adenosine by nucleotidase enzyme. The adenosine then degraded via a 2 step reaction. Firstly, adenosine deaminase enzyme binds on to adenosine to produce IMP. IMP then converts to inosine. Inosine is degrade by phosphorylation of nucleoside which leads to the production of hypoxanthine (48). The hypoxanthine is then transformed to xanthine and then finally to urate which is excreted in the urine (49).

1.2.4.3 Fructose

Fructose is converted to fructose-1-phosphate by fructokinase in the liver through a phosphorylation reaction (50). Phosphorylation of fructose causes a decrease of intracellular phosphate and adenosine triphosphate (ATP) levels and initiates inhibition of protein synthesis. AMP is generated and transformed to IMP. IMP then converts to urate (50). This means increasing AMP and IMP levels stimulate catabolic pathways which cause over-production of excess urate/uric acid (51).

1.2.5 Loss of Uricase Enzyme in Evolution

Uricase is an enzyme that initiates oxidative degradation of urate to allantoin. Allantoin is an extremely soluble chemical compound which is freely excreted in the urine (52). Humans and other higher primates do not have uricase enzyme, (53) as it was lost due to mutation during hominoid evolution. Wu and co-workers (54) showed that there are three mutations in the human, chimpanzees and gorilla’s uricase gene. These mutations include: two nonsense mutations, one in codon 33 and the other in codon 187 and another mutation was located in the splice acceptor signal exon 3 (54, 55). The absence of uricase enzyme causes increased urate level in the human body (55).
1.2.6 Absence of Uricase Advantages

Absence of uricase does however, have an advantageous impact on humans compared to other primates that have uricase. Positive effects derived from increased urate levels includes antioxidants effects, helping to maintain blood pressure and a neuroprotection function (56). The antioxidant property of urate provides protection against free radical activity and also helps decrease age-specific cancer rate (57, 58). A role for urate in neuroprotection is through protection from different disease such as multiple sclerosis and dementia including Alzheimer's disease and Parkinson’s disease (56).

1.2.7 Maintenance of Urate Levels

Most of uric acid excretion occurs at the kidney. Hence, the kidneys play a key role in the regulation of serum urate levels in the body (59). Renal under-excretion of uric acid will lead to hyperuricaemia (60). Approximately 90% of hyperuricaemia is accounted for by renal under excretion and ~10% is accounted for over-production of urate (61). The kidney filters uric acid through the glomerulus into a tubule system which includes the proximal tubule, loop of Henle and distal tubule. This tubule system will reabsorb the uric acid or pass it to the collecting ducts for excretion in the urine (Figure 1:2.2). Typically, in human ~90% of uric acid that pass to the kidney is reabsorbed while 5% to 10% is excreted (62).
1.2.8 Urate Transporters

The kidney apical membrane contains a numbers of transporters (Figure 1:2.3) including solute carrier family 5 member 12 (SLC5A12), sodium-anion transporters, solute carrier family 5 member 8 (SLC5A8) and solute carrier family 22 member 12 (SLC22A12), which is also called the urate transporter 1 (URAT1). Their fundamental role is to provide anions for URAT1 to transport uric acid.
Transporters involved in uric acid reabsorption include the organic anion transporter 4 (OAT4), which is an organic anion-dicarboxylate exchanger, and the short isoform of solute carrier family 2 member 9 (SLC2A9v2). Transporters responsible for uric acid secretion into the lumen are the ATP-binding cassette sub-family G member 2 (ABCG2), sodium phosphate transporter 1 (NPT1) and multi drug resistance protein 1 (MRP4). Ichida and colleagues (63) showed that ABCG2 is involved in uric acid excretion in the kidney. PDZ domain-containing 1 (PDZK1) is a protein that controls and regulates most uric acid transporters found in the kidney apical membrane. Other uric acid transporters also located at the basolateral side are organic anion transporter 1 (OAT1) and organic anion transporter 3 (OAT3). The long isoform of solute carrier family 2 member 9 (SLC2A9v1) which is also called the glucose transporter 9 (GLUT 9) has the ability to remove uric acid from the cell (64).
Figure 1.2:3 Uric acid transport in the proximal renal tubule of the kidney by urate transporters.

Taken from Merriman (2015) (65).

1.2.9 Gout Stages

The stage of gout are divided into four categories which are: asymptomatic hyperuricaemia, acute gout, intercritical gout and chronic tophaceous gout (66).

1.2.9.1 Asymptomatic Hyperuricaemia

This stage of gout is clinically categorised by a high serum urate levels (hyperuricaemia effect), however the patient will not report any symptoms. Hyperuricaemia effect is the major risk factor for gout and is defined as serum urate >0.42 mmol/L (67). However, not all individuals with
hyperuricaemia have gout (68). Hyperuricaemia a result from renal under-excretion of uric acid, over-production of excess urate or even both (69). Hyperuricaemia will lead to the deposition of needle-shaped monosodium urate crystals (MSU) around the affected tissues.

Overproduction of urate can be either exogenous (high intake of purine, seafood, fructose, alcohol containing food) or endogenous (increase in purine nucleotide breakdown as in a patient diagnosed with leukaemia and patient having cytotoxic therapy) (70-72). Under-excretion of renal acid is the most common cause of gout which is predominately caused by the impairment in kidney transporters. Such defects includes decreased level of glomerular filtration and increased tubular reabsorption.

There are currently drugs that are able to increase the level of uric acid in the body such as cyclosporine and diuretics (73). There are also drugs that decrease uric acid in the body such as salicylates and uricosuric (74).

1.2.9.1.1 Complication of Hyperuricaemia

Previous studies have shown hyperuricaemia to be a cause of gout. Hyperuricaemia is associated with obesity, renal dysfunction, hypertension, hyperlipidaemia and endothelial dysfunction (75-77). A study by Fang et al (2000) (78) shows that hyperuricaemia has evidence of independent association with cardiovascular mortality. Elevated levels of serum urate are often common in individuals with metabolic syndromes (MS) (79).

1.2.9.2 Acute Gout

Acute gout is characterised by MSU crystals forming in the joints, which lead to the activation of inflammation causing symptoms such as severe pain, immobility, redness, tenderness and warmth at affected area (66). The attacks mostly occur at the first metatarsophalangeal joint and often occur during night time. However, attacks can occur at other sites of the body including knees, tarsal and sub tarsal joints, ankles, wrists and small joints of the hands (80). Attacks of acute gout are often initiated by stress, trauma, excess alcohol and purine contain food consumption, infection, surgery, dehydration, diet and use of diuretics (27, 81).
1.2.9.3 **Intercritical Gout**

This stage is characterised by the interval between attacks (66). In this stage the level of urate can be very low, however, it can still cause re-attacks of gout. The affected joints appeared to be normal but over time the attacks and symptoms repeatedly occur and can affect other joints (82).

1.2.10 **Gout Risk Factors**

Gout pathogenesis has a number of risk factors. Common risk factors include sex, age, ethnicity, medications, diet and alcohol consumption.

1.2.10.1 **Sex**

Serum urate levels in males rise at puberty (73). Occurrence of gout is very rare in young women. This is thought to be because of the presence of oestrogen which has a uricosuric effect. Oestrogen also has the ability to decrease absorption of renal uric acid by reducing the expression of urate transporters that facilitated uric acid reabsorption (83-85). However, the prevalence of gout increases in women after menopause due to decreased oestrogen levels and increased level of uric acid (86). Common primary symptoms of gout can also develop in postmenopausal women that take diuretic drugs (87). Mikuls *et al* (2005) showed that the distributions of gout between elderly patients for both sexes is equivalent (88).

1.2.10.2 **Age**

Ageing is a crucial development risk factor for gout in both sexes. This risk maybe due to increasing serum urate level due to renal function impairment, the increased use of diuretics and occurrence of various co-morbidities associated with clinical gout (86). The prevalence of gout cases of all-ages in Aotearoa New Zealand Health Tracker (ANZHT) population was 2.69%. Prevalence of gout is higher in Pacific and Māori older men and women. About one-quarter of Pacific men and one-third of Māori men aged 65 and over are affected with gout. For Māori and Pacific women, about 12% and 25% of those aged 65 and over have gout (43).
1.2.10.3 Ethnicity

Māori and Pacific people of New Zealand, Pacific people from Western, Eastern Polynesian and Taiwanese aborigines are the most common ethnicities diagnosed with gout and hyperuricaemia (43, 89). Garcia et al (1997) showed that the Taiwanese aborigines genetic basis for familial gout was associated with elevated serum urate level (hyperuricaemia) and gout (90). The variation of renal uric acid transporters genes in Māori and Pacific people in New Zealand has been shown to cause under-excretion of renal acid which leads to gout development (91).

1.2.10.4 Purines Consumption

Foods that are rich in purines can increase serum urate level which is why they are associated with gout risk (70). Another study by Choi and colleagues showed that purines contained in red meat and seafood increase the risk of developing gout, while purines found in vegetables do not have evidence of association with gout risk (27).

1.2.10.5 Fructose Consumption

Many studies have illustrated that high fructose intake is associated with both gout and hyperuricaemia, as well as other co-morbidities such as insulin resistance, obesity, type 2 diabetes and metabolic syndrome (71, 92-96). A study by Choi et al (2008) (71) showed that risk of gout has a significant association with increased consumption of food such as fructose, sugar sweetened drinks, fructose containing in fruits (orange, apple) and fruit juices. Another study by Choi (2010) concludes that risk of gout is raised up to 74% in women that have one serving of sugar-sweetened soda per day, compared to women who have less than one serving per month (97). They also showed that diet soft drinks do not have any evidence of association with elevated level of gout risk.

1.2.10.6 Alcohol Consumption

Choi and colleagues (2004) showed evidence of an association between beer/liquor, especially purine-rich beverages, with increased development risk of gout and hyperuricaemia (81). However, no study has shown any association of wine with increased risk of gout and hyperuricaemia (81, 98). The biology underlying the gout and alcohol association is the fact that beer is rich in guanosine (81).
Alcohol intake is also associated with increased lactate generation and triggering of ketoacidosis (99). The keto acid will not only compete with uric acid for secretion but also trigger proximal tubular uric acid absorption by activating transporter URAT1 which has an organic anion exchange function (70).

1.2.10.7 Obesity
Elevated serum urate levels are associated with obesity and is thought to be due to the obesity association with insulin-resistance and therefore (70) high levels of insulin in the blood (hyperinsulinemia effect). This causes increased renal tubule reabsorption of uric acid (100). In addition, weight gain over time shows evidence of association with hyperuricaemia (101) and losing weight decreased development risk for gout and hyperuricaemia (102).

1.2.10.8 Medication
A numbers of drugs have been shown to have an association with gout such as diuretics, low-dose aspirin and medication used for organ transplantation (103). Low doses (1-2g/day) of aspirin have shown to cause a reduction of uric acid excretion while at high dose (>3g/day) it has a uricosuric effect. This indicates that aspirin has a bimodal effect on serum urate levels (104). Moreover, use of diuretics and low dose aspirin simultaneously worsens its effect on serum urate levels (105)

1.2.10.9 Kidney and other Organ Transplants
A study shows that 13% of transplant recipients develop gout and around 50% of recipients have hyperuricaemia (106). Kidney transplant receipt initiates reduction of uric acid secretion and decreases glomerular filtration rate. Moreover, drugs often prescribed for transplant recipients such as diuretics, some anti antimicrobials and immunosuppressant drugs (i.e. cyclosporine) exacerbate the effect on elevated serum urate level and consequently have a high association with gout and hyperuricaemia (107, 108).

1.2.11 MSU Crystal and Inflammation
The MSU crystal was first identified in the 18th century as a major cause of gout, however the molecular mechanism of MSU initiated gout was not known until recently. The deposition of MSU
crystals occurs in joints and soft tissues such as cartilage and fibrous tissues where it can avoid contact with inflammatory and immune related mediators (109).

According to Fabio et al (2006) (110) a MS- related crystal called calcium pyrophosphate dehydrate (CPPD) crystals or pseudogout also has the ability to initiate gout inflammation. When MSU crystals shed from their restricted sites into joint bursa they become a highly phlogistic compound, and may become a danger signal to the body. This signal leads to macrophage and monocytes being recruited to the site, and activating the NLRP3 inflammasome (Figure1:2.4) in the innate immune response (111). This results in caspase-1 activation, which is important in processing and secretion of interleukin (IL)-1B. IL-1B not only initiates the production of inflammatory mediators, including tumor necrosis factor (TNF)-α and IL-8, but also activates pro-inflammation cytokines. These cytokines and mediators stimulate infiltration of neutrophils into the inflamed joints and discharge inflammatory initiator compounds such as proteolytic enzymes, oxygen radicals, arachidonic metabolites and cytokines (112, 113). Previous studies have shown that cytokines such as TNF-α, IL-1B, IL6 and IL18, are significantly associated with joint disease aetiology. These cytokines, IL18 in particular, can attract neutrophils and macrophages to affected joints, and induce various pro-inflammatory effects (114).

IL-1B is a key regulator of the pro-inflammatory response in gout. Kingsbury et al (2011) (111), used a murine gout model with deficient IL-1 receptors and a wildtype mouse treated with IL-1 inhibitors to show that the injection of MSU crystals into joints significantly induced inflammation. This demonstrated that the inflammasome (Figure1:2.4) has a crucial role as a connection between well known gout stimulus (MSU crystal) and pathological hallmarks of acute gout attack (neutrophil influx).
Figure 1.2:4 NALP3 inflammasome complex during gout inflammation.

MSU/CPPD crystals are recognizable by the pattern recognition receptors of the innate immune response, which involve Toll Like Receptors (TLRs), and are phagocytosed by monocyte and macrophages. MSU crystals intracellularly are recognized by the NLRP3 inflammasome, which controls self-oligomerization and results in oligomerization of NLRP3 and cleavage/activation of procaspa-1 to caspase-1. Cathepsin B, ROS, and K+ efflux initiated by assembly of ATP is also implied to be involved in the activation and oligomerization of NLRP3 after MSU internalization. Caspase-1 inactive proIL-1β, transcribed in a NF-kB-dependent behavior subsequent TLR stimulation to made activated IL-1β, which is secreted into the extracellular joint fluid. Then the IL-1β activates IL-1 receptors on endothelial cells and local macrophages within the joint, resulting in signal transduction and gene activation result in to the secretion of proinflammatory cytokines and chemokines. These in turn stimulate leukocytes into the joint, strengthening the inflammatory cascade. Taken from: Kingsbury et al. (2011) (111).
1.3 Gout Complications

Many studies demonstrate significant association of gout with other illness including cardiovascular disease, kidney disease, hypertension, Type 2 Diabetes and metabolic syndrome (115-119). According to a study conducted in gout patients in a UK cohort, gout patients also have significant association with comorbidities such as coronary artery disease, hypertension and diabetes (120, 121). These additional complications makes gout difficult to manage and treat effectively.

1.4 Gout Treatment

Clinical gout symptoms have significant effect on the quality of life of the patient. Therefore, effective treatment, including changes in lifestyle and medication, must be employed to reduce symptoms of acute gout and avoid risk development of chronic gout and bone deformation.

1.4.1 Lifestyle Treatment

Dietary changes recommended for gout management include a low calorie diet with less red meat, less carbohydrates and avoid foods that contain large amount of fructose. Sugar-sweetened drinks and alcohol beverage are also to be avoided. These changes have had a significant effect on gout and other diseases that are associated with it such as diabetes and hypertension (120). Eating low fat dairy foods is shown to significantly decrease serum urate concentration and lower risk of gout (27, 98). Moreover, previous studies showed that consumption of some fruits, vegetables and coffee also have significant association with lowering serum urate concentration and a decreased risk of gout (122). Consumption of cherries as well as cherry products can reduce gout symptoms by preventing activation of the inflammatory pathway. Cherries inhibit inflammation by decreasing C-reactive protein and nitric oxide (NO) contained in the blood (123). Choi and colleagues (27) revealed that consumption of food rich in proteins, minerals and vitamins such as nuts and legumes have a huge benefit of lowering risk of gout.
1.4.2 Medication Treatment

First line of medication often used to treat clinical gout symptoms are non-steroidal anti-inflammatory drugs (NSAIDS) (2). NSAIDS reduce gout attacks and symptoms such as swelling, redness, pain and fever by preventing prostaglandin synthesis that initiates inflammation (124, 125). On the other hand, corticosteroids are used for acute gout treatment in patients that cannot take oral drugs and cases that have renal impairment or gastrointestinal bleeding (126). Corticoid steroids reduce gout symptoms by disrupting regulation of cytokines action in immune response which therefore inhibit inflammation (127, 128). Colchicine drugs have been used for more than 2000 years for gout treatment as they prevent cell growth through binding to tubulin dimers (129). Interference of colchicine with tubulin dimers inhibit microtubules polymerization and down-regulated membrane dependent mechanism including chemotaxis and phagocytosis processes (130). Evidence suggests that colchicine at high concentrations suppresses urate crystal stimulation of the NLRP3 inammasome (131). Chronic gout treatment often aims to reduce serum urate concentration to < 0.36 mmol/L, prevent frequency and reoccurrence of gout attacks by eliminating existing MSU crystals and preventing production of new crystals (132). Urate-lowering treatment can be used to treat both renal acid under-excretion as well as its over-production. Commonly allopurinol medication is used to lower urate production which prevents gout by blocking the conversion of hypoxanthine compound to urate. As a result, serum urate level decreases in the blood and urine, while the excretion of uric acid precursors, hypoxanthine and xanthine, is increased (133). Another common urate-lowering drug used for gout prevention is called Febuxostat. It prevents gout by inhibiting the oxidized and the reduced form of the xanthine oxidase enzyme (134). Other medication including probenecid, sulfinpyrazone and benzbromarone are used to prevent gout by increasing uric acid excretion. Probenecid inhibits the reabsorption function of proximal renal tubular while benzbromarone inhibits the reabsorption action of SLC2A9 and disrupt URAT1 regulation (134, 135).
1.5 Rheumatic Heart Fever

Rheumatic heart fever (RHF) is a significant burden of disease both in New Zealand and around the world. Its specific mechanism of pathogenesis is not quite discovered. RHF is a systemic auto-inflammatory disease that develops as a result of an unusual immune response to group A β-haemolytic streptococci (GABHS) infection. Major manifestations often develop 3 weeks post infection. The infection initially occurs in the upper respiratory tract where the streptococcus infects the throat, and result in a sore throat (4). The bacteria then enter the bloodstream and relocate to the heart, causing key manifestations such as cardiac valve stenosis (136-138). The bacteria can also promote arthritis in joints and inflammation of other organs (139). RHF inflammation can affect the heart, joints, skin and brain. Symptoms include fever, multiple painful joints, involuntary muscle movements, and a rare non itchy pink rash commonly known as erythema marginatum. Permanent valve damages will show as rheumatic heart disease. The damaged valve will manifest as heart failure and increase the risk of atrial fibrillation and valve infections.

RHF incidence and prevalence has steadily declined in developed countries since the early 1900s. However, it remains a leading cause of morbidity and mortality among young individuals (6 – 15 years) in developing countries (5). Studies have shown that the disease still has high incidence rates in some developed countries, such as New Zealand and Australia, due to large indigenous populations (Māori, Pacific Island, Aborigines) (138). This suggests that genetic and familial components of RHF are not the only factors underlying susceptibility, but that environmental factors such as poor housing conditions, crowding, and poor health knowledge are also involved. These environmental factors explain why low socioeconomic communities and developing nations such as Polynesian/Pacific Island and Māori have the highest incidence and prevalence on RHF than developed nations (140).

1.5.1 Global Epidemiology of RHF

The frequency of RHF is estimated by annual incidence and prevalence of cases. However, some countries find it difficult to obtain this data due to limited financial resources available to conduct an
epidemiology study. Most studies in RHF frequency are conducted using school children, hospital admission and discharge diagnoses.

There have been a number of epidemiological studies in RHF conducted over the past decades on developing and developed countries. An overview of these studies illustrates that the prevalence in developing countries is increasing dramatically, while the prevalence in developed countries is decreasing. It was estimated in the early 1900s in developed countries RHF frequency was 0.26% while the present proportion is 0.03% (141). However, its incidence is still estimated to be over 15 million cases worldwide with ~282,000 new cases and about 233,000 deaths each year (5). RHF frequency worldwide is common among 5 – 30 years old. Moreover, ~25 – 40% of cardiac cases admitted to the hospital are because of RHF (142-144).

According to the WHO (World Health Organisation) report (2004) (145), RHF mortality rate is greatly varying from 0.5 per 100,000 in Denmark, to 8.2 per 100,000 in China (146). The disability-adjusted life years (DALY\(^1\)) from RHF is also varied from 27.4 per 100,000 population of Region of Americas to 173.4 per 100,000 population in South East Asia (147).

Studies in school children showed that the Western Pacific has the highest prevalence of RHF especially in Samoa (77.8 per 1000 population). The Cook Islands (18.6 per 1000 population), and NZ Māori population (6.5 per 1000 population) also show an increase prevalence compared to NZ non- Māori (0.9 per 1000 population). The Region of Americas had the lowest prevalence, especially Cuba (Havana) with 0.2-2.9 per 1000 population. This shows that the incidence, prevalence and mortality rates of RHF varies widely between ethnicity, countries and even between sub-groups in the same countries (63-65).

A study by Seckeler et al (2011) (5) reported the result of 164 reviewed articles, 79 of which were reviews from hospital admissions or discharges and 8 were from population-based screening of national health registries prospective disease surveillance of Acute Rheumatic fever (ARF) and RHF.

---

\(^1\) Is the measurement of overall disease burden and it is expressed as the number of years lost due to poor health, disability or premature death. It was developed as a way of linking the overall health and life expectancy of different countries.
They demonstrated that RHF frequency (per 1000 people) varied between regions and between different times periods (Figure 1:5.1 and Figure 1:5.2).

![Map of global prevalence of RHF from 1991 to present. (1) Fiji (2) Samoa (3) Tonga.](image)

**Figure 1.5:1 Map presenting global prevalence of RHF from 1991 to present.** (1) Fiji (2) Samoa (3) Tonga.

*It appears that most of the Western Pacific countries has at least 4/1000 persons with RHF. Taken from Seckeler et al (2005).*
Figure 1.5:2 Trends of RHF prevalence per 1000 persons from 1970 to 2008 in different regions.

(A) The Americas (B) Europe (C) Africa (D) Eastern Mediterranean (E) Western Pacific, and (F) Southeast Asia. Square dots show prevalence frequency highlight from the literature review. It appears that RHF trend of all region is increasing except for Europe, where it appears to be declining. Taken from Seckeler et al (2005)

1.5.2 New Zealand Epidemiology of RHF

According to the New Zealand (NZ) Guidelines for Rheumatic fever (2009) (148), the rate of RHF in NZ in 2003 was 3.8/100,000; which was much larger than the rate of other Western countries e.g. Australia, Canada, Japan, Norway, Switzerland, and the United States. In 2003, approximately 141 new cases were reported. Most of these cases were in the age group 10 to 14 (n=82) and age group 5 to 9 (n=27). These 141 new cases included 70 cases from Māori and 58 cases among Pacific Island individuals, and were geographically located in lower socioeconomic areas in the North Island (Figure 1.5:3) including parts of Auckland, Waikato, Northland, Bay of Plenty, Rotorua, Gisborne, Hawke’s
Bay and Porirua. This areas are reported to have a high proportion of people of Māori /Pacific/Polynesian descent (149).

Figure 1.5.3 NZ map illustrating the geographic distribution of RHF prevalence per 100,000 in 2003. Taken from New Zealand Guidelines for Rheumatic Fever (2009).

1.5.3 Signs and Symptoms

Common symptoms of RHF often appear two-four weeks after throat infection and include fever, multiple joints pain, and involuntary muscle movements. Rarely symptoms include erythema marginatum also arise (137). Severe RHF cases often have permanent damage of the heart valves,
especially the mitral valves. This valve damages mostly occurs after multiple attacks of RHF. Heart valve damage increases the risk of developing atrial fibrillation and valves infection (4).

1.5.4 Pathophysiology of RHF

Group A beta-hemolytic streptococcal (GABHS) is a species classified by the Lancefield method of grouping haemolytic bacteria, which is based on the carbohydrate arrangement of antigens on their cell wall. Hence, Streptococcus species that are Group A are called Streptococcus pyogenes (150). *S.pyogenes* infect the upper respiratory tract and are known to be involved in the immune response that leads to RHF. This is achieved via an antigen molecular mimicry mechanism and cross reactivity mechanism (type II hypersensitivity) between the host and bacteria. This occurs when the host and the *S.pyogenes* share the same peptide epitope, which initiates an autoimmune response of the body’s local immune response against its own tissue. Cross reactivity of antibodies and/or T cells stimulate recognition between the *S.pyrogenes* peptides and the host protein which leads to inflammation and autoimmunity (151-153). Patients at this stage may develop symptoms such as carditis, fever, polyarthritis and other rare symptoms including subcutaneous nodules, chorea and erythema marginatum (154).

During RHF, *S.pyrogenes* relocate from the upper respiratory tract and develop colonies on the heart valve tissue and activate the autoimmune mechanism. Valvular endothelium is a well-known site of lymphocyte-induced damage. CD4+ T cells are the key effectors of heart tissue autoimmune reactions in RHF (155, 156). During RHF, the expression of vascular cell adhesion protein 1 (VCAM-1), a protein that is upregulated by self-antigen-specific antibodies, helps lymphocytes increase their adhesion. This suggests that VCAM-1 also contributes to the inflammation and valvular scarring associated with RHF (157). Normally, the self-reactive B cells remain silent in the periphery that means that T cell do not stimulate it. However, in RHF cases, antigen-presenting cells such as B-cells present the *S.pyrogenes* antigen to CD4-T cells that eventually differentiate to T2 helper cells (i.e. Th17). T2 helper cells stimulate B-cell activation which will go on and activated the production and secretion of antibodies to attack *S.pyrogenes* cell walls. These antibodies stimulate complement
mechanism and activate effector cells such as macrophages, natural killer cells, neutrophils that attack *S.pyrogenes* cell walls. This effector cells and complement may have also attack myocardium and joints, resulting in the RHF symptoms (156).

*S.pyrogenes* has a cell wall that contains polymers and M protein that are highly antigenic. The M protein of *S.pyrogenes* cross reacts with the cardiac myofiber protein myosin, heart muscle glycogen and smooth muscle cells of arteries. This cross reaction induces cytokine release and tissue destruction (158). The inflammation mechanism occurs through direct attachment of complement and Fc receptor mediators recruitment of neutrophils and macrophages. Neutrophils and macrophage are effector cells that attack body tissues and *S.pyrogenes* cell wall that caused initiation of inflammation. The inflammation forms Aschoff bodies that contain swollen eosinophilic collagen, lymphocytes and macrophages which can be seen under light microscopy (159).

Guilherme *et al* (2012) (160) suggested that the human body has a natural protection against the *S.pyrogenes* via genetically-controlled adaptive and innate immune system mechanisms. The innate immune system develops the first line of defence against *S.pyrogenes* while initiating the adaptive response. This is done by processing and presenting the antigen to the T cell, and through secretion of several cytokines, initiating complement activation (159). They suggest that genetic variants within genes that encode for effector molecules of the innate and adaptive immune systems may contribute to the pathogenesis of RHF. Such genes include toll like receptor 2, tumor growth factor-beta and IL23R. Most genetic studies of RHF illustrate that the major genetic predisposition factor associated with RHF susceptibility is a component of MHC class II molecules that are found on lymphocytes and on antigen-presenting cells. In particular, the DR and DQ alleles on chromosome 6 have been associated with RHF susceptibility (161). A number of studies have shown that the human leukocyte antigen (HLA) locus is the major locus that causes susceptibility to RHF (162). RHF susceptibility is dependent on certain and varied combinations of DR alleles which are associated with an increased susceptibility to valvular lesions. The susceptibility of RHF is due to the mechanism of HLA
molecules in presenting of antigen to T cell receptors which leads to activation of the immune response (161).

Besides DQ, some other loci have been associated with RHF. Stanevicha et al (2003) (163) found the cytokine TNF-α locus on human chromosome 6 to be associated with RHF, with a high expression level of TNF-α enhancing valvular tissue inflammation. Schafranski et al (2008) (164) illustrate that mannose-binding lectin (MBL) gene variants, a inflammatory protein involved in pathogen detection, is also associated with RHF, often leading to mitral valve stenosis.

1.5.5 Diagnosis

Diagnosis of RHF is determined using the modified Jones criteria documented in 1944 by T. Duckett Jones, MD and revised by the American Heart Association (154, 165). The Jones criteria diagnoses RHF when two major criteria or one major criteria and two minor criteria present together with streptococcal infection. Streptococcal infection is detected by increasing antistreptolysin O titre level \(^2\) (166). Kumar and Tandon (2013) suggest that the Jones’ criteria revised by American Heart Association in 1992 is inadequate to diagnosis RHF due to echocardiographic and Doppler (E&D) studies which detect subclinical carditis in cases with acute rheumatic fever (167).

1.5.5.1 Major Criteria

There are 5 major criteria which are polyarthritis, subcutaneous nodules, carditis, erythema marginatum and Sydenham’s chorea.

- **Polyarthritis** – Is the inflammation of large joints, it often begins in the legs then spread upwards and to other body parts.

- **Carditis** – Inflammation of the myocardium which is capable of developing congestive heart failure. This causes lack of breath, heart valves murmur and a pericarditis rub.

\(^2\) Antibody used to target against streptolysin O, an immunogenic, oxygen-labile hemolytic toxin made by strains of group A and most strains from groups C and G streptococci. Used for RHF diagnosis.
• **Subcutaneous nodules** – Collagen fibres accumulate firmly over and around bones or tendons. These nodules often occur outside the elbow, on the back of the wrist and the front of the knee.

• **Erythema marginatum** - A permanent pink rash that starts on the chest or arms as a macula, which then spreads outward. The rash form a clear appearance in the middle that looks like a ring. It will continue to spread outward and merge with other rings, which ultimately form a snake-like structure. This rash typically does not occur on the face and its effect is exacerbated in hot conditions.

• **Sydenham's chorea** - is a condition categorised by rapid, clumsy jerking movements that mainly affect the face, arms, hands and feet. This symptom is normally latent, it occurring very late in the disease and at least three months after infection.

### 1.5.5.2 Minor Criteria

There are 6 minor criteria for RHF diagnosis.

- **Fever** – RHF cases have a 38.2–38.9 °C (100.8–102.0 °F) temperature
- **Arthralgia** – Characterise by a pain at joints without swelling
- **Elevated level of erythrocyte sedimentation rate or C reactive protein**
- **Leucocytosis** – white blood count exceeding its normal range
- **ECG (electrocardiography)** - RHF cases will show features of heart blocks and prolonged PR interval.

- **Have previous series of RHF or any inactive heart disease.**

### 1.5.6 Prevention

Prevention of reappearance of RHF is accomplished by eliminating the acute infection and prophylaxis using antibiotics. The American Heart Association recommends that dental health has to be maintained, and that people with medical history of bacterial endocarditis, a heart transplant, artificial heart valves, or any kind of congenital heart defects must require to take long-term antibiotic prophylaxis (154, 168).
1.5.7 Treatment

Treatment of RHF focuses on the management and reduction of inflammation using anti-inflammatory medications, including aspirin and corticosteroids. Antibiotics are also used to treat those who have positive cultures for throat streptococcus. Salicylate medication are also often used to manage inflammation. The common used salicylate medication is aspirin which is given at high dose (100 mg/kg/day). However, aspirin has a number of side effects (mainly in children and teenagers) such as gastritis, salicylate poisoning and Reye's syndrome. Therefore, alternative medication must be considered when using aspirin or any aspirin containing product. Such alternative treatments include ibuprofen to suppress the pain and discomfort, and corticosteroids used to stop any moderate or severe inflammatory reaction caused by RHF (169-171). Steroid therapy is also used to minimise tissue scarring and avoid development of other RHF symptoms such as mitral valve stenosis. For cases having one attack of RHF, monthly injection (sometime 28 days interval) of penicillin for 5 years is prescribed for prevention of further RHF symptoms. However, if the patient has signs of carditis, penicillin injection treatment is performed for over 40 years. Recurrence of RHF is treated using low-dose of antibiotics such as penicillin, sulfadiazine, or erythromycin (171, 172).

1.5.7.1 Vaccine

Vaccine treatment is not available currently for prevention against S.pyogenes infection. Development of a vaccine is difficult due to the number of S.pyogenes strains abundant in the environment. Clinical trials and financial limitations also limit vaccine development (173, 174).

1.5.7.2 Infection

Patients that are positive for cultures of S.pyogenes are treated with penicillin except for those who allergic to penicillin. The most commonly used penicillin that was suggested by the Oxford Handbook of Clinical Medicine for RHF treatment is benzathine benzylpenicillin (175).

1.5.7.3 Inflammation

Corticosteroids are often used for inflammation treatment while salicylates are used to stop the pain (4).
1.5.7.4 Heart Failure

Heart failure resulting from RHF (such as carditis and congestive heart failure) are treated with ACE inhibitors, diuretics, beta blockers, and digoxin (176, 177).

1.6 Interleukin-23 Receptor Complex

The IL23R gene is located at chromosome 1p31, and codes for the interleukin 23 receptor. This receptor is located on the cell membrane of cells that are involved in the immune system, such as natural killer (NK) cells, monocytes, dendritic and T cells. These cells provide defence mechanisms against infection and disease from foreign microbes (178). The interleukin 23 receptor has a subunit interleukin 23 (IL23), which forms a protein-protein interaction together with the receptor, acting as a lock and key combination when the immune system is activated (Figure 1:6.1) (179). Previous studies have shown that IL23 is a member of the IL12 family as it is formed by the combination of the IL-12 p40/p19 subunits and is mostly synthesised by activated macrophage and dendritic cells. IL23 is a pro-inflammatory cytokine that triggers and regulates chemical signals, promoting inflammation by inducing secretion of IL-1 and TNF from stimulated dendritic cells and macrophages (112, 180, 181).

IL-1 and TNF activate the innate immune system and initiate dendritic cells to release IL-12 (112, 182). Additionally, IL23 also performs a role in the secretion of IL-17 from Th17 cells, which leads to the secretion of various other enhancer inflammation factors, such as IL-6 and TNF-α (183). The protein interaction of IL23R-IL23 initiates and guides the inflammation by inducing a local response of the immune system; activating the natural killer (NK) cell, initiating the T-cell mediated inflammatory response, stimulating the production of antibodies and promoting the assembly of inflammatory mediators, such as IL1, IL6, and TNF-α (184, 185). Hence any genetic variants present in IL23R can be predicted to contribute to immune/autoimmune related diseases. The most common autoimmune related diseases shown to have the IL23R pathway genetically contributing to risk are Crohns disease, rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis (186-190).
Figure 1.6: IL23-IL23R Pathway.

IL-23 (p40 and p19 subunits) binds with the IL-23R complex (IL-12Rβ1 and IL-23R subunits), which is associated with Jak family members including Tyk2 and Jak2. IL-23 binding to IL23R activates Jak2-mediated phosphorylation of tyrosine residues situated at the intracellular domain of the IL-23R subunit. Phosphorylated tyrosine residues act as a landing site for STAT3 molecules, which in turn are phosphorylated. Phospho-STAT3 proteins homodimerize and move into the nucleus which stimulates transcription of cytokines, such as IL-17A, IL-17F, IL-22 and IFN-γ, TNF-α. Adapted from Cesare et al (2009) (191).

1.6.1 IL-23R genetic variants

Few studies have documented the association of particular IL-23R variants and gout (112), and there are currently no studies that have shown evidence of IL-23R variants association with RHF. However, many studies have been published documenting association of IL23R variants with various auto-inflammatory/auto-immune related disorders such as Crohn’s disease, rheumatoid arthritis,
ankylosing spondylitis, psoriasis arthritis and inflammatory bowel disease (Refer to Table 1.6:1 for summary of common IL23R variants association with auto-inflammatory/autoimmune disease).

1.6.1.1 SNPs of Interest

1.6.1.1.1 RS11465804

Burton and co-workers (192) showed that rs11465804 (a synonymous SNP) is one of several IL23R gene variants that are significantly associated with ankylosing spondylitis (AS) in a European sample set of over 1,000 patients (Odd Ratios = 0.68, P-value = 0.0002). The major allele (T) appeared to be the susceptible allele for AS. Duerr et al (2006) (193) revealed that rs11465804 is one of many IL23R SNPs that showed minor allele G having a protective association with Crohn’s disease in non-Jewish and Jewish samples sets (OR < 1, P-value < 0.05).

The Duerr study also showed that rs11465804 linked with rs11209026 confer an association on Crohn’s disease. A study by Einarsdottir and colleagues (194) showed rs11465804 to have a susceptible (minor allele G) association with IBD (including Ulcerative Colitis (UC) and Crohn’s Disease (CD)) (P-value = 0.009, OR = 1.38) in the Swedish population. The study also revealed that SNP rs11465804 haplotypes with other IL23R SNPs (i.e. rs11209026, rs10489629 and rs201841) which showed significant association with celiac disease (P-value = 0.00006) and psoriasis (P-value = 0.03) in the Finnish population.

1.6.1.1.2 RS11209026

RS11209026 is a coding non-synonymous SNP (c.1142G>A, p.Arg381Gln) with a minor allele (A) showed by many studies to have protective association with most auto-inflammatory/auto-immune related disease. Duerr et al (2006) (178) reveal that the rs11209026 major allele (G), appeared to have a strong susceptible association with IBD including CD and UC development in Jewish and non-Jewish patients (P-value < 0.05) except for UC in Jewish population (P-value > 0.05). This indicates that the minor allele (A) has a protective effect against IBD, CD in Jewish/non-Jewish and UC in the
non-Jewish population. The association of rs11209026 with UC in Jewish population was not significant (P-value = 4.91 × 10⁻¹). A study by Weersma and colleagues (195) also showed a protective effect of the rs11209026 minor allele (A) with IBD, CD and UC in 518 Dutch case-control patients (IBD OR = 0.19, CI = 0.10-0.37, P-value = 6.6 E-09; CD OR = 0.14, CI = 0.06-0.37, P-value = 3.9E-07 and UC OR = 0.33, CI = 0.15-0.73, P-value = 1.4E-03). They also showed that rs11209026 has no significant association evidence with celiac disease. Brown (2008) (196) shows that the rs11209026 minor allele A also has a protective effect against ankylosing spondylitis in over 1,000 European patients (Carriers OR = 0.63, P-value = 2.8x10e⁻⁵). A haplotype of this SNP with SNP rs7530511 (197) showed evidence of association with psoriasis. Moreover, a study also showed a protective association of rs11209026 minor allele (A) with psoriasis (198, 199). Hollis-Moffatt and colleagues (199) conducted a study to test six IL23R variants which included rs11209026 for significant association with rheumatoid arthritis in a New Zealand cohort. Their data revealed that rs11209026 does not have evidence of association with RA (OR = 1.01, P-value = 0.86) however one SNP (rs13431510) emphasised a significant association with RA (OR = 1.14, P-value = <0.001). Chen-Xu and colleagues (186) replicated this data in European sample set. There results revealed that only a Spanish sample set formed a significant association of rs13431510 with RA (OR = 1.14, P-value = 0.03) while UK samples had an OR trending in the susceptible direction (OR = 1.15, P-value = 0.06). Other population set i.e. Australasian and Norwegian had odd ratios of 1.00; P-value = 0.98, and OR 1.00; P-value = 0.97 respectively.

1.6.1.1.3 RS7517847

Lui et al (2015) (112) is currently the only study that has documented an association of a IL23R gene variant with gout. This variant is a synonymous SNP rs7517847. This study showed the rs7517847 minor allele (G) to be a susceptible risk factor for gout development in Chinese Han male population (OR = 1.22, P-value = 0.047). A study by Tung and colleagues (200) suggest that the rs7517847 minor allele G is specifically associated with Crohn’s disease but not associated with UC or others type of IBD.
Another study (201) showed that rs7517847 had the most significant association (protective effect of minor allele G for rs7517847 and A for rs11209026) with CD (OR = 0.65, P-value = 4.9x10(-9)) compare to SNP rs11209026 (OR = 0.43, P –value = 6.65x10(-6)) and it is statistically independent from SNP rs11209026. They also mentioned that rs7517847 has no significant association with UC (OR = 0.63, P-value = 0.008). Li and co-workers (202) meta-analysis supports the result that rs7517847 and rs11209026 minor alleles have a protective effect against CD development. This means that major allele G (rs11209026) and T (rs7517847) import susceptibility to CD development. A meta-analysis study conducted by Lee and colleagues (203) revealed a significant association of rs7517847 with ankylosing spondilitis in Europeans and Asians population.
Table 1.6.1 A summary of Auto-inflammatory & Autoimmune disease that are associated with common IL23R SNPs.

IL23R SNP of interest highlight yellow.

<table>
<thead>
<tr>
<th>IL23R SNP</th>
<th>Auto-inflammatory &amp; Autoimmune disease</th>
<th>Risk Allele</th>
<th>Chromosome position</th>
<th>Source/Previous Studies</th>
</tr>
</thead>
</table>
1.7 **Aim of this Study**

This literature review highlighted the current gap in the knowledge of gout and RHF in relation to IL-23R genetic variants of interest (rs11465804, rs7517847, rs11465804). While the association of IL23R with both gout and RHF has not been confirmed by any previous studies, it has been shown with related auto-inflammatory disease. Hence, the aim of this project is to test IL23R gene variants for association with gout and rheumatic heart fever in the European and Polynesian (Pacific/Māori) population. This will be done using case-control sample sets recruited from within New Zealand and Europe (gout only) and rheumatic heart fever in the Polynesian (Pacific/ Māori) population using case-control sample sets recruited from within New Zealand.
Chapter 2: Methods
1.8 Gout & RHF Case-Control Study

1.9 Participant Recruitment & Ethical Approval

Participants were educated about this case-control study and how their information would be used. Participation was voluntary and full consent for use of DNA was obtained. The NZ Multi-Region Ethics committee granted the ethical approval for this study. Gout cases were all confirmed by a rheumatologist or a General Practitioner (GP); satisfying the criteria for clinical gout from the American College of Rheumatology (ACR) for preliminary diagnostic conditions of gout (204). Cases were recruited from 2006 to 2013 from the Auckland, Waikato and Canterbury regions of New Zealand. A control sample set was recruited from Auckland, Canterbury and Otago regions and had no self-reported history of acute gout. An additional ~1000 gout cases of European ancestry were recruited throughout Europe. The total number of cases was 2003 and the total number of controls was 2069. In addition, questionnaires were provided to all participants regarding their information including ethnicity (Polynesian ancestry), age, sex, medical history and family history. Some of this information was used for confounders’ adjustment in this study.

For RHF, all cases (total number = 207) were also confirmed by a rheumatologist or a GP; using the criteria in the American College of Rheumatology (ACR) for preliminary diagnostic conditions of RHF (204). All cases were recruited in 2013 by Pedro Ming and Dinny Lennon from Auckland Penicillin Clinics and most cases were adults of Polynesian descent. Controls (total number = 164) used for RHF were also used for gout controls. Questionnaires were also provided for all RHF participants regarding their ethnicity (Polynesian ancestry).

1.10 Study Population

The study cohort was divided into two major groups: (1) European (European Ancestry) and (2) Polynesian (Māori and Pacific Island Ancestry) ethnicities. The Polynesian population was further divided into sub-groups: (a) Western Polynesian (WP) include people from Tonga, Samoa, Niue and
Tokelau, (b) Eastern Polynesian (EP) involved people from Māori and Cook Islands and (c) a mixture of the two sub-groups (EPWP). This was done as a previous study (205) showed allele and genotyping distribution of various genetic variants often vary between ancestral ethnicities. Therefore, analyses of this study could result in finding false positive association of an allele/locus with gout/RHF (205). To overcome a potential false result, the numbers of Polynesian grandparent ancestry was accounted for as a confounder. Polynesian genetic status are different due to travelling events that occurred during colonisation of their lands (45, 206, 207).

1.11 Extraction of DNA from blood samples

DNA was extracted from blood samples by technical staff working in the Merriman laboratory in the Biochemistry department, Otago University.

1.11.1 Lysis of the Red blood cell (RBC)

Blood (~ 10 ml) was collected from participants (case and controls) using tubes with EDTA to help avoid coagulation. Blood was then transferred into falcon tubes (50 ml) and washed using lysis buffer till reached a total volume of 40 ml. Tubes were mixed thoroughly and centrifuged for about 15 min at 2500 rpm. In this process, RBCs were lysed allowing the haem to dissolve in the supernatant. The supernatant was removed and the pellet was resuspended in the lysis buffer and mixed by centrifuged for another 15 minutes at 2500 rpm. The supernatant was discarded again. This step was repeated until the haem was completely discarded and the pellet change colour to white. Conducting this step is important because the haem is recognised as a PCR inhibitor. White blood cells (WBCs) pellet of DNA was then extracted instantly or frozen at 25°C.

1.11.2 Lysis of White blood cell (WBC) and DNA extraction

1.11.2.1 Digestion of protein and WBC lysis

A total of 3.5 ml of 6 mol/l guanadinium hydrochloride (GuHCl) was added to the white blood cells pellet and the mixture was mixed well. A total of 250 ml of 7.5 mol/l ammonium acetate (NH4Ac),
50 ml of 10 mg/ml proteinase K and 250 ml of 20% Na sarcoyl was each added immediately and mixed thoroughly. The solution was then incubated in the 37°C room and left overnight.

1.11.2.2 Removal of protein from cell

The solution was reassigned in a 15 ml tube containing 2 ml of prechilled chloroform (CHCl₃). Then the tube was mixed by vortexing and left for a minute, then centrifuged it for 3 minutes at 2500 rpm. After centrifugation the solution formed two layers. The bottom layer (organic layer) contained proteins and the top layer (aqueous layer) contained DNA. The top layer was collected for DNA and 10 ml of cold absolute ethanol was added immediately to the DNA in a 15 ml tube. The solution was mixed by inversion of the tube. This allowed the DNA to precipitate. The sample was then centrifuged at 3000 rpm for 15 minutes. The supernatant was removed and the pellet was rinsed twice using 4 ml of 70% ethanol and was centrifuged again for 5 minutes at 3000 rpm. The supernatant was discarded and the pellet was left to dry overnight in room temperature.

1.11.2.3 DNA quantification and dilution

The pellet was resuspended in 200 ml TE buffer of pH 7.5 (10 mmol/l, 0.1 mmol/l) and incubated at 55°C for 1 hour. Optical densitometry was used to determine the concentration of DNA using 260nm on the Nanodrop Spectrophotometer (ND- 1000). DNA was diluted to 200ng/μL concentration using 10mM Tris/1mM EDTA and samples were then stored in the freezer. For genotyping purposes, the sample was diluted to 8ng/μL using TE buffer.

1.12 Quality control

The role of quality control is to repeat ~10% of samples to ensure no genotyping error occurred, correct genotyping calling and to check for misaligned 96-well plates. Quality controls were made by aliquoting 8 samples from all case-control 96-well box containing samples that were recruited for the study into a separate 96-well plate. Quality controls were conducted after every TaqMan® genotyping assay.
1.13 SNP Selection

Three genetic variants derived from IL23R were chosen for this study. SNP selection was based on the literature review regarding auto-inflammatory and auto-autoimmune related diseases such as Crohn’s disease, rheumatoid arthritis, ankylosing spondylitis, psoriasis arthritis and inflammatory bowel disease that appeared to be associated with IL23R gene variants (193, 208-210). An Ensemble search ([http://asia.ensembl.org/Homo_sapiens/Info/Index](http://asia.ensembl.org/Homo_sapiens/Info/Index)) looking for functional IL23R variants was performed. SNP information such as population allele frequency, association with inflammatory disease was obtained from Ensemble data. The Polynesian sample sets SNP and genotype data was estimated using data of Han Chinese population, as studies have showed that Oceanic population originated from continental East Asia (45). Therefore, SNP and genotype information of the Polynesian population was assumed to be similar to the Chinese HapMap data. The 3 IL23R SNPs chosen were RS11465804 (G/T), RS7517847 (G/T), and a non-synonymous SNP RS11209026 (A/G) (Arg381Gln).

1.14 Constructing Linkage Disequilibrium (LD) using Haploview

The Haplotype Map (HapMap) ([http://www.hapmap.org/index.html.en](http://www.hapmap.org/index.html.en)) is a catalogue of common DNA variations from four populations: Han Chinese from Beijing, China (CHB), Yoruba from Ibadan, Nigeria (YRI), European from Utah, USA residents with Northern and Western European ancestry (CEU), and Japanese from Tokyo, Japan (JPT). Build 35 and 36 platforms of the HapMap were used through the Haploview programme version 4.2 software to make linkage disequilibrium (LD) plots (Figure 1.14:1) for the 3 IL23R SNPs. The LD plot was generated only for the gout sample set. The pairwise LD for the IL23R SNPs was measured using $r^2$ values. An $r^2$ value of 0 means that there was no LD between markers, $r^2$ of 100 shows that there was complete LD between markers, $r^2$ > 70 shows a high LD, 40-70 was considered as a moderate LD, $r^2$ < 40 means LD is low. Linkage disequilibrium was used to investigate if the 3 SNPs of interest had any correlation with each other in order to evaluate if haplotype analysis should be done.
Each box shows the extent of LD between a single pair of markers. The black boxes indicate strong LD ($r^2 = 79$), the grey boxes represent low LD ($r^2 = 3$). Numbers inside these boxes indicate the $r^2$ value between marker pairs.

### 1.15 Haplotype Analysis

A haplotype is a set of SNP alleles inherited together and found on the same chromosome. Haplotype analysis was conducted on SNPs rs11209026 and rs11465804 as they have high LD with each other in European and Polynesian sample sets. The result from this analysis can be used to increase power to discover any association of this SNPs with gout.

Haplotype analysis was performed using PLINK software. In this analyses, two input files (.ped & .map file) were created on excel (Table 1.15:1, Table 1.15:2). These files were import into PLINK software and specific commands (Appendix 4) were used to calculate P-value, OR, case and control minor allele frequency of SNPs (rs11465804 & rs11209026) alleles combination.
Table 1.15: Example of input Map file.

First column are the chromosome no, second column are the SNPs, third column showed genetic distance (morgans) – but can use 0 for all and last column are the SNP position.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RS11465804</td>
<td>0</td>
<td>67702526</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>RS11209026</td>
<td>0</td>
<td>67705958</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.15:2 Example of input Ped file.

Column 1 = Family ID, Column 2 = Individual ID, Column 3 = Paternal ID (can used 0), Column 4 = Maternal ID (can used 0), Column 5 = Sex (1 = male, female = 2, other = unknown), Column 6 = Phenotype (1 = controls, cases = 2, unknown = 0), Column 7 & 8 = rs11465804 Genotype, Column 9 & 10 = rs11209026 Genotype.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1006</td>
<td>A1006</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A1016</td>
<td>A1016</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A289</td>
<td>A289</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A314</td>
<td>A314</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A323</td>
<td>A323</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A328</td>
<td>A328</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A336</td>
<td>A336</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A351</td>
<td>A351</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A375</td>
<td>A375</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A377</td>
<td>A377</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

1.16 TaqMan® genotyping Assay

TaqMan® assay is an extremely sensitive technique used for genotyping genetic variants correctly. A light cycler 480 version 15 machine (Roche Applied Science, Indianapolis, USA) was used for TaqMan® genotyping assay. All TaqMan® genotyping assay are designed to work under universal
reaction conditions. TaqMan® genotyping assays were ordered from Applied Biosystems (http://www.appliedbiosystems.com).

1.16.1 TaqMan® Assay mechanism

The 6FAM dye-MGB and VIC dye-MGB labelled probes are the two fluorogenic probes used by the TaqMan® genotyping assay process. The VIC dye is connected to the allele 1 probe at the 5’ end and the 6FAM dye is connected to the allele 2 probe at 5’ end (211). Each SNP had probes specifically designed for them situated between the two PCR primers. The Taqman probe consisted of quencher (3’ end) and flurophore reporter (5’ end) which is placed in close proximity to each other to avoid emission of flurophore signals while the probe is inactivated/intact (212) (Figure 1.16:1). The probe will be activated when its allele-specific sequence is correctly complementary to the variants allele sequence during PCR amplification step. The probe will then bind to the target DNA strand and then be degraded by the 5’-nuclease action of the Taq polymerase molecule during extension of the DNA strand from the PCR primers (213, 214). The reporter flurophore will then separate from the probe and initiate signal (Figure 1.16:1). The formation of PCR products can be measured by the density/amount of fluorescence emitted by probe reporter (FAM & VIC). Identifying genotype can be determined by measuring each signal released by reporters, detecting allele(s) of the target region that exist in the samples (214).
Figure 1.16: Schmatic diagram illustrating an overview of TaqMan® genotyping assay.

The probe binds to the allele specific sequence in the target DNA during amplification step. This binding cause the reporter fluorophore dye to separate from the probe and the quencher by the action of the 5′-nuclease Taq polymerase. Releasing of reporter which eventually caused emission of fluorescent signals. Taken from Bass et al (2010) (215).
1.16.2 Setting up the 384-well plate for TaqMan® genotyping using Light Cycler

A 2µL aliquot of 6-10ng/µL DNA was loaded into a 384-well plate, a shield placed on top of plate and the plate centrifuged at 1000rpm, 1min. The plate was covered in the 37º room or drawer for minimum of 2 days to ensure DNA is dried down. Master Mix (Kapa mix + Water + Probe for 3 SNP of interest) (Table 1.16:1) was prepared and vortexed immediately. A 5µL aliquot of Master Mix was placed into each 384-well using a dispenser and the plate by using a swiper to seal it properly. The plate was spin to ensure the Master Mix was on the plate well bottom. Then the plate was wrapped in foil to protect from sunlight and store in the freezer till it was ready to run.

*Preparation of Master Mix reaction

Per Sample → 2.5µL Kapa mix + 2.43µL Water + 0.07µL Probe

Defrost probe from freezer first and aliquot Kapa Mix and water into a label tube. Then vortex the solution instantly and spin down probe before adding in the tube. Vortex the solution again afterwards.

Table 1.16:1 Volume of Master Mix per Reaction.

<table>
<thead>
<tr>
<th></th>
<th>¼ plate</th>
<th>½ plate</th>
<th>1 plate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard Reaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5µL + Dried DNA)</td>
<td>1 Reaction</td>
<td>108 Reaction (96 samples)</td>
<td>208 Reaction (192 samples)</td>
</tr>
<tr>
<td>2 x Kapa Mix</td>
<td>2.5 µL</td>
<td>270 µL</td>
<td>520 µL</td>
</tr>
<tr>
<td>Water</td>
<td>2.43 µL</td>
<td>262.44 µL</td>
<td>505.44 µL</td>
</tr>
<tr>
<td>40 x SNP assay</td>
<td>0.07 µL</td>
<td>7.56 µL</td>
<td>14.56 µL</td>
</tr>
<tr>
<td>Total</td>
<td>5 µL</td>
<td>540 µL</td>
<td>1040 µL</td>
</tr>
</tbody>
</table>
1.16.3 Using the Light Cycler

Turn the light cycler on and open the Light Cycler 480 software, enter password/username and check correct database is selected. Choose ‘New Experiment from template’ then choose ‘Kapa mix protocol SNP genotyping’. Place the prepared plate in Light Cycler. To start run, press ‘start run’. Automatically you will be notified to name your plate and where you want to place your running file. Label subsets by selecting ‘Subset editor’ and apply pre-existing template. Give the subset a name and then select relevant wells by holding down ‘control’ and highlighting with the mouse. Then select ‘Apply’. Repeat this for each of the subsets. Running of the sample will take ~45 minutes. When TaqMan® genotyping is complete the result will display on a table and on a cluster plot (Figure 1.16:2). Result from TaqMan® assay were recorded in a haploped file (Figure 1.17:1).

![Cluster Plot Example](image.png)

**Figure 1.16:2 Example of cluster plot from the TaqMan® genotyping assay of the 3 SNP of interest.**

*The blue triangle clusters represent homozygous allele X (major allele in this case), the red triangles represent heterozygotes allele which are X and Y (both allele), the green triangles represents homozygous allele Y (minor allele in this case), pink triangle represent unknown samples and the grey...*
triangles are fail/negative samples. The pink triangles occur when a sample fails to cluster together with any genotype group, therefore no genotype can be assigned for these samples or DNA failures. The grey triangles occur when there is no fluorescence detected by the assay, these are most likely to be blanks in DNA plates. The X-axis represent Fluorescence with wavelength of 483-533 and Y-axis represent wavelength of 523-568. NB: Sometimes, Taqman assay assigned alleles with wrong colours, in this case manual editing must performed to provide allele with correct colours.

1.17 Entry of TaqMan® genotyping Data into Haploped File

The Merriman Lab staff created a master haploped file (Figure 1.17:1) that consisted of samples details such as sample ID, sample number/box, box position, gout affection, age, sex and ancestry details. This file was used to record SNPs genotyping data which helped to ensure accurate and consistent calling of all samples genotyping result.

![Table of TaqMan® genotyping Data](image)

**Figure 1.17:1 Example of the master Haploped file.**

Showing samples information such as Patient/Sample ID, Box number, Box position, Sex, Ethnicity (ancestry information), gout affection (affected or not affected)) and genotyping information
represented by number 1 and 2. Homozygous X allele → label 1,1; Heterozygous allele → X & Y label 1,2; and Homozygous allele Y → label 2,2. Blanks samples highlighted in red were used to check there was no misplacement of samples genotyping and was also used to check for correct orientation of the plate during genotyping.

1.18 BC|SNPmax

BC|SNPmax is a database provided by Biocomputing Platforms Ltd at http://www.bcplatforms.com/. It helps to analysis and manage large numbers of SNP genotyping data as well as SNP clinical phenotype data. SNP information such as marker maps, pedigrees, SNP quality information and analysis result are also kept in this database.

1.18.1 Uploading SNPs Genotypes file on the BC|SNPmax

A complete SNP genotyping file (Table 1.18:1) was completed by adding allele genotypes to the genotype template file prepared by Mandy Phipps Green (Merriman lab staff). The SNP genotyping was uploaded file onto BC|SNPmax.

**Table 1.18:1 Exampe of SNP genotyping file for BC|SNPmax upload**

<table>
<thead>
<tr>
<th>PATIENT</th>
<th>Rs11465804</th>
<th>Rs11465804</th>
<th>Rs11209026</th>
<th>Rs11209026</th>
<th>Rs7517847</th>
<th>Rs7517847</th>
</tr>
</thead>
<tbody>
<tr>
<td>A4001</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>A4002</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A4003</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A4004</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>A4005</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A4007</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

The two genotyping files Gout (European & Polynesian) sample set and RHF (Polynesian) sample set file were uploaded to BC|SNPmax (Figure 1.18:1).
1.18.2 Affection data sets files

Affection data sets files were created for both gout and RHF genotyping sample set by Mandy Phipps Green (Merriman lab staff). These files contained the gout/RHF affection status which were labelled as 0 if the individual does not have gout/RHF and 1 if the individual has gout/RHF.

1.19 Statistical analysis

1.19.1 Hardy-Weinberg Equilibrium test using SHEsis software

The Hardy-Weinberg equilibrium (HWE) test of all genotype samples was calculated using a web-based software called SHEsis (http://analysis.bio-x.cn/myAnalysis.php) (216) (Figure 1.19:1). HWE was conducted on all 3 SNPs DNA in the European, Polynesian and Polynesian sub-groups. The HWE theorem states that all allele and genotype distributions in any population should remain the same from generation to generation in a large, random-mating population not subjected to influences such as mate choice, mutation, selection or migration (217). Therefore, calculation of HWE can be used to
inform any genotype errors or problems with calling of genotypes in any DNA box. The ideal HWE had a p-value >0.05 which means that a SNP is consider to be in HWE. If the HWE p-value is <0.05 this means the SNP is not in HWE (Figure 1.19:2). Samples found not to be in HWE were repeated for confirmation. Previous studies explain some possible reasons that cause SNPs to deviate from HWE, including mutation selection, small population size, disease association, non-random mating, genetic drift and meiotic drive (217, 218).

Figure 1.19:1 Example calculation of HWE using SHEsis.

Tick site analysis option, enter number of sites (no. of SNP), the marker names (name of SNP with population) and the input data of case and control SNP genotype with a reference column.
Figure 1.19: Example of SHEsis HWE calculation result of SNP rs11465804.

Minor (G) & major (T) allele frequency (green box), genotype frequency (blue box), unadjusted Odds ratio (95% CI), and HWE test p-value for both case and controls (orange box). In this case HWE p-value for case:control (0.14:0.15) is not significant, meaning that this SNP genotyping is within the HWE.

1.19.2 Logistic Regression using R studio software

Logistic regression is a model that analyses a dataset in which independent variables can be used to determine an outcome. In logistic regression analysis, the dependent variable can have only two possible outcomes while the independent variables may be either continuous or discrete. Logistic regression is commonly used in genetic research, where the outcome variable indicates occurrence or absence of a disease (219).

Logistic regression model was used to investigate an association of SNPs of interest with gout and RHF. These analyses were carried out using R studio software. The input text files were created as shown in Figure 1.19:3 below using Excel.
Figure 1.19.3 Example of input txt file for logistic regression analysis of SNP using R.

Affection status (label 0 for controls and 1 for cases), SNP genotype (label 1,1 for homozygous major; 2,2 for heterozygous and 3,2 for homozygous minor), ETHCLASS (ethnicity class → label 1 for EP, 2 for WP, 9 for EPWP and 3 for European), CAUvsPOLY (Subject either European or Polynesian → label 1 for European and 2 for Polynesian) and the 3 confounder: age, sex, GPancestry (samples with >0.75 GPANCESTRY are highly Polynesian ancestry) are shown.

1.19.3 R Studio Software

Firstly, convert the input file format from excel to txt, open R and import input txt file (Figure 1.19:3).

Name the input txt file and create subset for European, Polynesian, EP (Eastern Polynesian includes people from Hawaii, Cook Island and New Zealand), WP (Western Polynesian include people from Samoa, Niue and Tuvalu) and EPWP (combination of Eastern and Western Polynesian). Due to the fact that allele and genotype frequencies can differ between Polynesian populations, it is important to assess their allele and genotype frequency in sub-groups → EP, WP and EPWP ancestry. The GPANCESTRY column in the input txt file represent Polynesian ancestry. Any sample with a GPANCESTRY value >0.75 has higher Polynesian ancestry. If the allele frequency of each subgroup is similar or varies by a small amount therefore, we can analyse them all as Polynesian but if not they
must analyse individually (EP, WP and EPWP sub-groups). Repeat the same command (Appendix 1) for WP and EPWP sub-groups. Used command (Appendix 1) to calculate unadjusted and adjusted P-value, odd ratio, 95% confident interval of European, Polynesian and Polynesian subgroups.

1.19.3.1 Interpretation of P-value

Fisher’s p-value was used to measure the statistical significance of SNPs association with gout/RHF. A P-value of <0.05 was consider to be statistically significant but a P-value >0.05 was consider as non-significant (220).

1.19.3.2 Interpretation of Odd Ratio (OR)

An OR of 1 means that case group allele frequency is similar to control group. This means no association occurred between the SNP and the disease An OR > 1 indicates that the minor allele is over-represented in cases that in controls, therefore cases minor allele may be associated with susceptibility to develop the disease. An OR < 1 indicates that the minor allele is under-represented in cases compared with controls. This suggest that minor allele is in a protective direction against a disease. An OR of 2 means that the risk of a disease development is increased by 2 times when exposed to a risk allele in a SNP (221).

1.19.3.3 95% Confidence Interval (95% CI)

A confidence interval has a upper and a lower limit that estimate array of values for an unknown parameter such as an OR. The confidence limit determines how big or small the true effect of the unknown parameter might be. A confidence intervals at the 95% level, means that 95% of the time, the confidence interval will contain the true value of the unknown parameter (222).

1.19.3.4 Adjusting for Confounding factors

If a covariate is associated with a SNP in a case-control association studies, there is a possibility that the covariates may have an influence on the association status between genetic marker and the disease. Therefore, adjustment of covariates in any genetic study is important in order to avoid false association. The three most common confounders in the association studies for gout are sex, age, Polynesian ancestry. This is because older people and Polynesian people are more likely to develop
gout and it is more frequent in men than in women. Hence these factors were adjusted in gout sample sets. For the RHF sample set, only Polynesian ancestry confounder was used for adjustment because some patient do not have age and sex recorded. Furthermore, there is no sex bias in RHF and all controls are RHF-free adults so age will not be a confounder.

1.19.4 Meta-Analysis

A meta-analysis was used to combine the data sets of European and Polynesian population in order to increase the overall power of the study. Meta-analyses of all genetic variants were conducted using R studio software. P-values and Odds ratios of genetic variants in European and Polynesian including subgroups WP, EP and EPWP were combined. A heterogeneity test was also conducted to estimate the differences between the data set. The meta-analysis used 2 statistical models: (1) a fixed effects model, (2) a random effects model. In the fixed effects model it is believed that all studies have a similar effect size and come from a common population. In the random effects model, it is believed that the true effect can differ from study to study (223, 224)

1.19.4.1 Meta-analysis R-studio

Firstly, the input data set was prepared using excel as shown in Table 1.19:1. The excel file was converted into a txt file and imported into the R software and the analysis run for European and Polynesian population by inserting command. The overall effect P-value, the heterogeneity P-value and a forest plot (Figure 1.19:4 & Figure 1.19:5) was calculated from this command (Appendix 2).

Table 1.19:1 Example of an input file for meta-analysis of a SNP in European and Polynesian.

Beta value is a log of the odd ratio and SE would be its standard error. NB: Meta-analysis was perform only on Gout sample set only because RHF sample set only had Polynesian population.

<table>
<thead>
<tr>
<th>STUDY</th>
<th>BETA</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European</td>
<td>-0.040363</td>
<td>0.162483</td>
</tr>
<tr>
<td>NZ Polynesian</td>
<td>-0.1717378</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Output Result:

<table>
<thead>
<tr>
<th>Study</th>
<th>TE</th>
<th>seTE</th>
<th>OR</th>
<th>95%-CI</th>
<th>W(fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ Caucasian</td>
<td>-0.040</td>
<td>0.1625</td>
<td>0.960</td>
<td>[0.698; 1.321]</td>
<td>64.7%</td>
</tr>
<tr>
<td>NZ Polynesian</td>
<td>-0.172</td>
<td>0.2200</td>
<td>0.842</td>
<td>[0.547; 1.296]</td>
<td>35.3%</td>
</tr>
</tbody>
</table>

Overall (95% CI)

*Heterogeneity: Q=0.2, df=1, p=0.631*

0.917 [0.710; 1.185] 100%

---

**Figure 1.19: Example of a Meta-analysis forest plot.**

The grey boxes represent individual data sets which vary in size depending on the % weight (label W(fixed)) and they contribute to the overall analysis. The higher the percentage weight the bigger the box will be and the larger the influence of that data set on the overall OR. The solid line in the graph centre is known as the line of no effect with the value of 1. The lines passing through boxes horizontally represent the 95% confidence interval. Overall OR (95% CI) is highlighted in green. Overall meta-analysis is represented by the diamond on the plot. The middle of the diamond represent the OR and it width represent 95% CI. NB: If the diamond pass through the line of no effect, therefore the different between data set case-controls is considered to be not significant.
Figure 1.19:5 An example of a meta-analysis output where the P-value for heterogeneity (red square).

Therefore, the overall effect P-value is from the Fixed effect model (Blue square). If the heterogeneity test is significant then Random effects model P-value would be the ideal overall effect P-value to use (yellow square).

1.19.5 Secondary Analysis

A secondary analysis was conduct on the 3 SNPs dataset to examine any association between the controls with higher serum urate level (> 0.41 mmol/L) and gout in European and Polynesian population. The main purpose of this analysis was to test if controls with hyperuricaemia are associated with gout in IL23R SNP of interest data set. If the association is statistically significant, therefore hyperuricaemia controls have IL23R locus with variants that have susceptible effects. If no evidence of association found this indicates that hyperuricaemia controls have IL23R protective variants against gout.
1.19.5.1 R-studio of Secondary Analysis

The secondary analysis was conducted using R-studio software. An input file was prepared on excel (Figure 1.19.6). This file was converted into txt format before importing into R. Command shown in Appendix 3 were used to calculate P-values, odds ratio and 95% CI interval of 3 SNP data set.

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>SEX</th>
<th>AGE</th>
<th>AFFECTION</th>
<th>URATE1MONTH</th>
<th>HUvsGOUTAFFECTION</th>
<th>ETHNICITY</th>
<th>GPANCESTRY</th>
<th>RS13465804_1</th>
<th>RS13465804_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT0038</td>
<td>1</td>
<td>57</td>
<td>1</td>
<td>0.57</td>
<td>1.10 samoan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0056</td>
<td>1</td>
<td>55</td>
<td>1</td>
<td>0.54</td>
<td>1.05 NZ Maori 0.5 Caucasian</td>
<td>0.5 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0071</td>
<td>1</td>
<td>53</td>
<td>1</td>
<td>0.47</td>
<td>1.0875 NZ Maori 0.125 Caucasian</td>
<td>0.875 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0081</td>
<td>1</td>
<td>59</td>
<td>1</td>
<td>0.48</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0094</td>
<td>1</td>
<td>61</td>
<td>1</td>
<td>0.55</td>
<td>1.075 CI Maori 0.125 Tahitian 0.125 Caucasian</td>
<td>0.875 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0095</td>
<td>2</td>
<td>76</td>
<td>1</td>
<td>0.5</td>
<td>1.10 CI Maori</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0099</td>
<td>2</td>
<td>62</td>
<td>1</td>
<td>0.46</td>
<td>1.10 CI Maori</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0099</td>
<td>1</td>
<td>70</td>
<td>1</td>
<td>0.45</td>
<td>1.075 CI Maori 0.25 Caucasian</td>
<td>0.75 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0101</td>
<td>1</td>
<td>43</td>
<td>1</td>
<td>0.42</td>
<td>1.05 Puleapukan 0.5 CI Maori</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0105</td>
<td>1</td>
<td>48</td>
<td>1</td>
<td>0.65</td>
<td>1 CI Maori (% unknown)</td>
<td>T T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0109</td>
<td>2</td>
<td>59</td>
<td>1</td>
<td>0.47</td>
<td>1.0125 NZ Maori 0.875 Caucasian</td>
<td>0.125 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0112</td>
<td>1</td>
<td>40</td>
<td>1</td>
<td>0.53</td>
<td>1.025 Puleapukan 0.75 CI Maori</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT0138</td>
<td>1</td>
<td>75</td>
<td>1</td>
<td>0.64</td>
<td>1.10 NZ Maori</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS007</td>
<td>1</td>
<td>53</td>
<td>1</td>
<td>0.43</td>
<td>1.10 tongan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS007</td>
<td>1</td>
<td>62</td>
<td>1</td>
<td>0.48</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS009</td>
<td>1</td>
<td>71</td>
<td>1</td>
<td>0.53</td>
<td>1.10 Caucasian</td>
<td>0.5 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS010</td>
<td>1</td>
<td>58</td>
<td>1</td>
<td>0.64</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS013</td>
<td>1</td>
<td>71</td>
<td>1</td>
<td>0.42</td>
<td>1.0875 NZ Maori 0.125 Caucasian</td>
<td>0.875 G</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS014</td>
<td>1</td>
<td>59</td>
<td>1</td>
<td>0.57</td>
<td>1.10 samoan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS049</td>
<td>1</td>
<td>43</td>
<td>1</td>
<td>0.69</td>
<td>1.10 tongan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS052</td>
<td>1</td>
<td>66</td>
<td>1</td>
<td>0.47</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS053</td>
<td>1</td>
<td>85</td>
<td>1</td>
<td>0.45</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS058</td>
<td>1</td>
<td>70</td>
<td>1</td>
<td>0.64</td>
<td>1.10 NZ Maori</td>
<td>1 G</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS059</td>
<td>1</td>
<td>63</td>
<td>1</td>
<td>0.61</td>
<td>1.075 samoan 0.25 Chinese</td>
<td>0.75 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS060</td>
<td>1</td>
<td>39</td>
<td>1</td>
<td>0.64</td>
<td>1.075 samoan 0.25 Chinese</td>
<td>0.75 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS062</td>
<td>1</td>
<td>80</td>
<td>1</td>
<td>0.5</td>
<td>1.05 NZ Maori 0.25 Lebonese 0.25 Caucasian</td>
<td>0.5 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS063</td>
<td>1</td>
<td>51</td>
<td>1</td>
<td>0.55</td>
<td>1.05 samoan 0.5 Tokelauan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS065</td>
<td>1</td>
<td>55</td>
<td>1</td>
<td>0.5</td>
<td>1.10 tokelauan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS072</td>
<td>1</td>
<td>49</td>
<td>1</td>
<td>0.45</td>
<td>1.075 CI Maori 0.125 Chinese 0.125 Caucasian</td>
<td>0.75 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS074</td>
<td>1</td>
<td>68</td>
<td>1</td>
<td>0.51</td>
<td>1.10 Caucasian</td>
<td>1 G</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS075</td>
<td>1</td>
<td>47</td>
<td>1</td>
<td>0.47</td>
<td>1.10 Caucasian</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS076</td>
<td>1</td>
<td>38</td>
<td>1</td>
<td>0.57</td>
<td>1.10 tongan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS094</td>
<td>1</td>
<td>43</td>
<td>1</td>
<td>0.53</td>
<td>1.10 samoan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS100</td>
<td>1</td>
<td>46</td>
<td>1</td>
<td>0.98</td>
<td>1.10 samoan</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS302</td>
<td>1</td>
<td>52</td>
<td>1</td>
<td>0.43</td>
<td>1.10 Niuean</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS312</td>
<td>1</td>
<td>46</td>
<td>1</td>
<td>0.42</td>
<td>1.10 Niuean</td>
<td>1 T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure: 1.19.6 Example of input txt file for secondary analysis of SNP using R.

Affection status (label 0 for controls and 1 for cases), URATE1MONTH (mention amount of serum urate in each sample), HUvsGOUTAFFECTION (label patient with >0. mmol/L serum urate level with 1 and sample < 0.41 with 0), SNP genotype (label T:T for homozygous major; T:G for heterozygous and G:G for homozygous minor), GPancestry (samples with >0.75 GPANCESTRY are highly Polynesian ancestry) are shown and Ethnicity column shown ethnicity of patients. NB: 365 European participants were found to be hyperuricaemia and 248 Polynesian participants.
1.19.6 Power Calculation

A power calculation was performed in order to test whether the sample size of the study has enough power to detect statistically significant associations. A P-value < 0.05 was used in the power calculation to indicate that sample sets had a significant levels, meaning that it have enough power to detect an association with gout (OR ≥ 1). Nevertheless, the study power was decreased by SNPs with a low minor allele frequency and small sample size data set. In other cases, some variants have lower minor allele frequency but still show high power. The level of effect was estimate from odd ratios 1.2 (low effect from Liu et al (2015) shows SNP rs7517847 minor allele (G) has an OR of 1.2 for gout), 1.5 (moderate effect) and 2.0 (strong effect). Power calculation figure shown in Figure 1.19:7.
Figure 1.19:7 Power calculation figures generated on excel.

Yellow box represent control MAF, Red box represent odd rations that used to estimate the level of effects (e.g. 1.2, 1.5 and 2), Green box is the number is a factor by which the number of cases has to be multiplied to get total number of controls, Purple box represent total numbers of cases and orange box represent the output power.
Chapter 3: Results
1.20 Association of IL23R SNPs of interest with Gout & Rheumatic Heart Fever

1.20.1 Linkage Disequilibrium (LD) Plots of SNPs ($r^2$)

Linkage disequilibrium plots of 3 SNPs revealed that SNP rs11465804 and SNP rs11209026 had strong LD ($r^2 = 79$) in the European dataset. The same two SNPs also show moderate LD values in the Polynesian sample set ($r^2 = 43$) (Figure 1.20:1). However, the LD value between SNP rs7517847 and SNP rs11465804 were very low in European and have nothing in Polynesian ($r^2 = 3$, $r^2 = 0$ respectively). This suggests that rs7517847 and rs11465805 are not correlated in either population. On the other hand, SNPs rs11209026 and rs11465804 in both population are correlated and associated with each other, hence they are in linkage disequilibrium.

![Linkage Disequilibrium (LD) plots](image)

*Figure:1.20:1 Linkage Disequilibrium (LD) plots that were generated from Haploview. Showing the LD structure (displayed as $r^2$) between SNPs rs11465804 and rs11209026 in (A) European population and (B) Polynesian population.*

1.21 Gout Association Analysis

Genotype and allele distributions of IL23R SNPs of interest (rs11465804, rs11209026 and rs7517847) are presented in Table 1.21:2. This shows that two SNPs (rs11465804 and rs11209026) showed no
significant association with gout in both European and Polynesian population after adjustment for confounders (sex, age and ancestry grand-parents) (rs11465804 European OR = 1.04, P-value = 0.82; Polynesian OR = 0.84, P-value = 0.59 rs11209026 European OR = 1.09, P-value = 0.83, Polynesian OR = 0.85, P-value = 0.67 respectively). For the SNP rs7517847 there is evidence of significant association with gout in the Polynesian population. The SNP rs7517847 of Polynesian data set shows a protective effect of the major allele (T), meaning that the minor allele (G) susceptibility to gout (European OR = 0.95, P-value = 0.53; Polynesian OR = 0.85, P-value = 0.04). The Polynesian data set for rs7517847 was divided into 3 sub-groups (WP, EP and EPWP) to account for the possible presence of stratification admixture within Polynesian due to the heterogeneity of allele frequency distribution in Polynesian ancestries (Table 1.21:1, Table 1.21:2). Statistical analysis revealed that only the WP population show evidence of association with a susceptible effect of the minor allele (G) (WP OR = 0.76, P-value = 0.04).

The HWE test calculated from SHEsis for all SNPs in European and Polynesian population appeared to be within Hardy-Weinberg equilibrium (P-value > 0.05) except for SNP rs11465804 in the Polynesian case data set (P-value = 0.01) (Table 1.21:2). This means a genotyping error might have occurred or because of the occurrences of mutation selection, small population size, non-random mating, genetic drift and meiotic drive.

Minor allele distribution of rs11465804 and rs11209026 in European and Polynesian population is relatively low compared to rs7517847 (Table 1:21.1). Moreover, the homozygous risk allele frequency of rs7517847 is relatively higher compare to rs11465804 and rs11209026 except rs11465804 Polynesian controls and rs11209026 Polynesian population which are monomorphic. As a result rs7517847 is highly possible to associate with gout.

1.21.1 Meta-Analysis

Meta-analysis was conducted on the 3 SNPs by combining the European and Polynesian sample sets to increase their data overall power. The 3 SNPs were not significantly heterogeneous between sample sets, meaning the datasets are relatively similar (rs11465804 \( P_{\text{hetero}} = 0.568 \), rs11209026 \( P_{\text{hetero}} = \))
0.550, rs7517847 P_{hetero} = 0.399) (Table 1.21:2). However, meta-analysis showed no evidence of significant association of data set (rs11465804 P_{overall effect} = 0.958, rs11209026 P_{overall effect} = 0.752, rs7517847 P_{overall effect} = 0.196) (Table 1.21:2 & Figure 1.21:1).

**Table 1.21:1 Minor allele frequency of 3 IL23R SNPs in Polynesian sub-groups (EP, WP and EPWP) for Gout and RHF sample set.**

<table>
<thead>
<tr>
<th>SNP &amp; Population</th>
<th>GOUT</th>
<th>RHF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNP</strong> &amp; <strong>Population</strong></td>
<td><strong>MAF (%)</strong></td>
<td><strong>MAF (%)</strong></td>
</tr>
<tr>
<td><strong>RS11465804</strong></td>
<td><strong>G</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>EP</td>
<td>1.55</td>
<td>1.70</td>
</tr>
<tr>
<td>WP</td>
<td>0.57</td>
<td>0.00</td>
</tr>
<tr>
<td>EPWP</td>
<td>0.34</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>RS11209026</strong></td>
<td><strong>A</strong></td>
<td><strong>A</strong></td>
</tr>
<tr>
<td>EP</td>
<td>0.91</td>
<td>0.80</td>
</tr>
<tr>
<td>WP</td>
<td>0.22</td>
<td>1.51</td>
</tr>
<tr>
<td>EPWP</td>
<td>0.23</td>
<td>1.32</td>
</tr>
<tr>
<td><strong>RS7517847</strong></td>
<td><strong>G</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>EP</td>
<td>65.12</td>
<td>62.52</td>
</tr>
<tr>
<td>WP</td>
<td>57.22</td>
<td>55.00</td>
</tr>
<tr>
<td>EPWP</td>
<td>59.01</td>
<td>51.50</td>
</tr>
</tbody>
</table>
### Table 1.21:2 Meta-Analysis and Allele/Distribution data of IL23R SNP of interest for Gout Dataset.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Cases</th>
<th>Meta-analysis</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>HWE</th>
<th>HWE controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS11465804</td>
<td></td>
<td>European</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SNP</td>
<td>MAF</td>
<td>Controls</td>
<td>MAF</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>839</td>
<td>92 (0.099)</td>
<td>98 (0.052)</td>
<td>844 (0.882)</td>
<td>110 (0.115)</td>
<td>116 (0.061)</td>
<td>0.85 (0.64 to 1.11)</td>
<td>1.04 (0.74 to 1.46)</td>
</tr>
<tr>
<td>Polynesian</td>
<td>914</td>
<td>20 (0.021)</td>
<td>22 (0.012)</td>
<td>1047 (0.967)</td>
<td>36 (0.033)</td>
<td>36 (0.017)</td>
<td>0.72 (0.41 to 1.22)</td>
<td>0.84 (0.44 to 1.16)</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td></td>
<td>Z-value = -0.053, Overall effect P-value = 0.958, Heterogeneity P-value = 0.568, OR (95% CI) = 0.99 (0.74 to 1.34), Genotyping (%) success rate = 99.21%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS11209926</td>
<td></td>
<td>European</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SNP</td>
<td>MAF</td>
<td>Controls</td>
<td>MAF</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>826</td>
<td>104 (0.111)</td>
<td>112 (0.060)</td>
<td>847 (0.884)</td>
<td>107 (0.112)</td>
<td>115 (0.060)</td>
<td>0.96 (0.74 to 1.34)</td>
<td>1.09 (0.79 to 1.50)</td>
</tr>
<tr>
<td>Polynesian</td>
<td>920</td>
<td>16 (0.017)</td>
<td>16 (0.009)</td>
<td>1061 (0.970)</td>
<td>33 (0.030)</td>
<td>33 (0.015)</td>
<td>0.56 (0.29 to 1.03)</td>
<td>0.85 (0.41 to 1.79)</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td></td>
<td>Z-value = 0.316, Overall effect P-value = 0.752, Heterogeneity P-value = 0.550, OR (95% CI) = 1.05 (0.78 to 1.41), Genotyping (%) success rate = 99.63%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS7517847</td>
<td></td>
<td>European</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SNP</td>
<td>MAF</td>
<td>Controls</td>
<td>MAF</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>282</td>
<td>439 (0.477)</td>
<td>199 (0.216)</td>
<td>837 (0.455)</td>
<td>260 (0.273)</td>
<td>491 (0.516)</td>
<td>0.94 (0.82 to 1.07)</td>
<td>0.95 (0.82 to 1.12)</td>
</tr>
<tr>
<td>Polynesian</td>
<td>168</td>
<td>410 (0.454)</td>
<td>326 (0.361)</td>
<td>1062 (0.587)</td>
<td>162 (0.157)</td>
<td>484 (0.468)</td>
<td>0.92 (0.81 to 1.02)</td>
<td>0.85 (0.73 to 0.99)</td>
</tr>
<tr>
<td>Region</td>
<td>EP</td>
<td>WP</td>
<td>EPWP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>79</td>
<td>191</td>
<td>1 (0.030)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.202</td>
<td>0.488</td>
<td>0.515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>121</td>
<td>15 (0.455)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.309</td>
<td>0.309</td>
<td>0.555</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>433</td>
<td>433</td>
<td>47 (0.712)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.554</td>
<td>0.554</td>
<td>0.712</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>53</td>
<td>7 (0.101)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.148</td>
<td>0.191</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>324</td>
<td>122</td>
<td>38 (0.551)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.471</td>
<td>0.439</td>
<td>0.551</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>262</td>
<td>103</td>
<td>24 (0.348)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.381</td>
<td>0.371</td>
<td>0.348</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>848</td>
<td>328</td>
<td>86 (0.623)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.616</td>
<td>0.590</td>
<td>0.623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.94</td>
<td>0.92</td>
<td>1.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79 to 1.11</td>
<td>0.74 to 1.14</td>
<td>0.79 to 3.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.76</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79 to 1.21</td>
<td>0.59 to 0.98</td>
<td>0.51 to 2.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.47</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>0.47</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.04</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>0.82</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meta-analysis  

Z-value = - 1.294, Overall effect P-value = 0.196, Heterogeneity P-value = 0.399, OR (95% CI) = 0.93 (0.83 to 1.04), Genotyping (%) success rate = 96.64%
Figure 1.21: Meta-analysis plot for 3 SNPs of interest combining European and Polynesian data set in gout.

(A) RS11465804 plot, (B) RS7517847 plot, (C) RS11209026 plot. NB: Meta-analysis for Polynesian sub-group only conducted in RS7517847 because is the only significant SNP.
1.21.2 Secondary Analysis

Secondary analysis was performed to test whether controls with hyperuricaemia (have hyperuricaemia ≥ 0.42 mmol/L but not affected with gout) have significant association with gout in IL23R SNP of interest dataset. This was done to determine if controls with hyperuricaemia in IL23R SNP of interest dataset have a locus that contains a variant with protective effects against gout. Table 1.21:3 shows that no SNPs showed any sign of significant association with gout in both population. This indicated that controls with hyperuricaemia may have IL23R locus that produced genetic variant with protective effects against gout in Europeans and Polynesian population.

Table: 1.21:3 Allele/Genotype distribution and Odd Ratios/P-value of secondary analysis.

Calculated from control sample sets in European and Polynesian population for IL23R SNPs.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Controls</th>
<th>MAF</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS11465804</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>TT</td>
<td>844</td>
<td>(0.882)</td>
<td></td>
</tr>
<tr>
<td>Polynesian</td>
<td>GT</td>
<td>110</td>
<td>(0.115)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>3</td>
<td>(0.003)</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>116</td>
<td>(0.061)</td>
<td>(0.54 to 1.95)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS11209026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>GG</td>
<td>847</td>
<td>(0.884)</td>
<td></td>
</tr>
<tr>
<td>Polynesian</td>
<td>GA</td>
<td>107</td>
<td>(0.112)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>4</td>
<td>(0.004)</td>
<td>0.867</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>115</td>
<td>(0.060)</td>
<td>(0.50 to 1.61)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS7517847</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>TT</td>
<td>260</td>
<td>(0.273)</td>
<td></td>
</tr>
<tr>
<td>Polynesian</td>
<td>GT</td>
<td>491</td>
<td>(0.516)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>200</td>
<td>(0.210)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>891</td>
<td>(0.468)</td>
<td>1.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.79 to 1.41)</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.21.3 Power Calculation for Gout Dataset

Power calculations (Table 1.21:4) shows that all populations of SNP rs7517847 estimate from all three odd ratios (European OR 1.2 = 0.79, OR 1.5 = 0.99, OR 2 = 1.0; Polynesian OR 1.2 = 0.79, OR 1.5 = 0.99, OR 2 = 1.0 respectively). This means that the population size and minor allele in rs7517847 dataset is enough to find an association with gout in both population. Power calculation of other two
SNPs (rs11465804 and rs11209026) is relatively low compared to SNP rs7517847 power because of a low minor allele frequency (MAF). The European sample set for both of the SNPs has higher power than the Polynesian sample set (rs11565804 European OR 1.2 = 0.29, OR 1.5 = 0.90, OR 2 = 0.99; Polynesian OR 1.2 = 0.12, OR 1.5 = 0.44, OR 2 = 0.91). This indicates that both of this SNPs have lower minor allele frequency therefore a chance of finding any association of this SNPs with gout is quite low.

Table 1.21:4 Power Output of Gout Dataset.

The power output was estimated from the range of Odd Ratios (low effect, medium effect and high effect) and minor allele frequency (MAF) of controls.

<table>
<thead>
<tr>
<th>SNP &amp; Population</th>
<th>Controls MAF</th>
<th>Power Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odd Ratios</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>RS11465804</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>0.061</td>
<td>0.29</td>
</tr>
<tr>
<td>Polynesian</td>
<td>0.017</td>
<td>0.12</td>
</tr>
<tr>
<td>RS11209026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>0.060</td>
<td>0.28</td>
</tr>
<tr>
<td>Polynesian</td>
<td>0.015</td>
<td>0.12</td>
</tr>
<tr>
<td>RS7517847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>0.468</td>
<td>0.79</td>
</tr>
<tr>
<td>Polynesian</td>
<td>0.610</td>
<td>0.79</td>
</tr>
</tbody>
</table>

1.22 Rheumatic Heart Fever Association Analysis

The allele and genotyping distribution present in Table 1.22:2 shows SNPs rs11465804 and rs7517847 do not associate with RHF in the Polynesian population after adjustment for grandparent ancestries confounding. Interestingly, rs11465804 shows evidence of association with RHF before confounder adjustment (OR = 0.11, P-value = 0.004). This means case and control sample sets are poorly matched in the numbers of Polynesian ancestry grandparent they have. On the other hand, SNP
rs11209026 shows evidence of association with a protective effect for the minor allele (A) (OR = 0.07, P-value = 0.002) after adjustment. Therefore the rs11209026 major allele (G) is in a susceptible direction. The HWE of all 3 SNP data set were within Hardy-Weinberg equilibrium and rs7517847 has the most minor and homozygous minor allele distribution while rs11465804 and rs11209026 are monomorphic.

### 1.22.1 Power Analysis for RHF Dataset

The power calculation of the RHF dataset in Polynesian population (Table 1.22:1) shows power outputs of Odd ratio 1.2, 1.5 and 2 are relatively low. This is because the RHF dataset has a low size. However, SNP rs7517847 seems to have high power outputs compare to the other two SNPs. This may be due to rs7517847 having a greater allele frequency than rs11465804 and rs11209026.

<table>
<thead>
<tr>
<th>Table 1.22:1 Power output for RHF in Polynesian Dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The power output was estimate from range of Odd Ratios (low effect, medium effect and high effect) and minor allele frequency (MAF) of controls.</td>
</tr>
<tr>
<td>SNPs</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>RS11465805</td>
</tr>
<tr>
<td>RS11209026</td>
</tr>
<tr>
<td>RS7517847</td>
</tr>
</tbody>
</table>
Table 1.22: Allele and distribution data of IL23R SNPs for RHF dataset.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Cases</th>
<th>MAF</th>
<th>Controls</th>
<th>MAF</th>
<th>Unadjusted OR (95% CI)</th>
<th>Adjusted OR (95% CI)</th>
<th>P value</th>
<th>Unadjusted P value</th>
<th>Adjusted P value</th>
<th>HWE case</th>
<th>HWE controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS11465804</td>
<td>TT</td>
<td>GT</td>
<td>GG</td>
<td>G</td>
<td>2 (0.990) 9 (0.903)</td>
<td>10 (0.097) 10 (0.049)</td>
<td>0.11</td>
<td>0.004</td>
<td>0.922</td>
<td>0.94</td>
<td>0.6</td>
</tr>
<tr>
<td>Polynesian</td>
<td>202</td>
<td>2 (0.010)</td>
<td>2 (0.005)</td>
<td>93 (0.903)</td>
<td>10 (0.097)</td>
<td>10 (0.049)</td>
<td>0.11 (0.02 to 0.42)</td>
<td>0.91 (0.09 to 5.91)</td>
<td>0.004</td>
<td>0.922</td>
<td>0.94</td>
</tr>
<tr>
<td>RS11209026</td>
<td>GG</td>
<td>GA</td>
<td>AA</td>
<td>A</td>
<td>3 (0.015) 1 (0.927)</td>
<td>8 (0.073) 8 (0.037)</td>
<td>0.13</td>
<td>0.07 (0.008 to 0.35)</td>
<td>0.009</td>
<td>0.002</td>
<td>0.92</td>
</tr>
<tr>
<td>Polynesian</td>
<td>198</td>
<td>3 (0.985)</td>
<td>3 (0.007)</td>
<td>101 (0.927)</td>
<td>8 (0.073)</td>
<td>8 (0.037)</td>
<td>0.13 (0.02 to 0.51)</td>
<td>0.07 (0.008 to 0.35)</td>
<td>0.009</td>
<td>0.002</td>
<td>0.92</td>
</tr>
<tr>
<td>RS7517847</td>
<td>TT</td>
<td>GT</td>
<td>GG</td>
<td>G</td>
<td>70 (0.354) 236 (0.596)</td>
<td>18 (0.173) 54 (0.519)</td>
<td>1.09</td>
<td>1.02 (0.62 to 1.66)</td>
<td>0.621</td>
<td>0.947</td>
<td>0.92</td>
</tr>
<tr>
<td>Polynesian</td>
<td>32</td>
<td>96 (0.485)</td>
<td>70 (0.354)</td>
<td>236 (0.596)</td>
<td>18 (0.173)</td>
<td>54 (0.519)</td>
<td>1.09 (0.77 to 1.55)</td>
<td>1.02 (0.62 to 1.66)</td>
<td>0.621</td>
<td>0.947</td>
<td>0.92</td>
</tr>
</tbody>
</table>
1.23 **Haplotype Analysis**

Haplotype analysis was conducted on rs11465804 interactions with rs11209026 due to the strong and moderate LD ($r^2$) in European ($r^2 = 79$) and Polynesian ($r^2 = 43$) population. Result shown in Table 1.23.1 and Table 1.23.2 is that haplotype of rs11465804 and rs11209026 in European and Polynesian showed no evidence of association with gout (P-value > 0.05).

**Table 1.23:1 European Haplotype Result in Gout**

<table>
<thead>
<tr>
<th>Allele Combination (rs11465804/rs11209026)</th>
<th>Case MAF</th>
<th>Control MAF</th>
<th>Odd Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A</td>
<td>0.049</td>
<td>0.056</td>
<td>0.884</td>
<td>0.404</td>
</tr>
<tr>
<td>T/A</td>
<td>0.009</td>
<td>0.006</td>
<td>1.441</td>
<td>0.318</td>
</tr>
<tr>
<td>G/G</td>
<td>0.002</td>
<td>0.006</td>
<td>0.429</td>
<td>0.138</td>
</tr>
<tr>
<td>T/G</td>
<td>0.938</td>
<td>0.931</td>
<td>1.111</td>
<td>0.431</td>
</tr>
</tbody>
</table>

**Table 1.23:2 Polynesian Haplotype Result in Gout**

<table>
<thead>
<tr>
<th>Allele Combination (rs11465804/rs11209026)</th>
<th>Case MAF</th>
<th>Control MAF</th>
<th>Odd Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A</td>
<td>0.006</td>
<td>0.011</td>
<td>0.526</td>
<td>0.082</td>
</tr>
<tr>
<td>T/A</td>
<td>0.003</td>
<td>0.003</td>
<td>0.793</td>
<td>0.693</td>
</tr>
<tr>
<td>G/G</td>
<td>0.006</td>
<td>0.006</td>
<td>1.021</td>
<td>0.965</td>
</tr>
<tr>
<td>T/G</td>
<td>0.985</td>
<td>0.979</td>
<td>1.411</td>
<td>0.172</td>
</tr>
</tbody>
</table>
Chapter 4: Discussion
1.24 Association Analysis of Gout Dataset

Single nucleotide polymorphisms (SNPs) and gene studies provided knowledge and understanding into genetic predisposition of certain diseases, information about genetic regions involved in certain disease and insight to how target genes will respond to drugs. The clinical and biological interactions between genes and risk factors for complex disease such as Polynesian ancestry, age and sex are examined using gene studies. SNP studies help improve our knowledge about the genetic basis of complex disease such as gout and RHF. Previous studies show that non-synonymous SNPs (have amino acid changes) and SNPs within regulatory regions have the most effect on certain diseases traits (225, 226).

Gout is a type of auto-inflammatory arthritis that is typified by the deposit of MSU crystals either around or inside an affected joint. The MSU crystal is phagocytized by an active monocytes after the NLRP3 inflammasome is triggered. This caused activation of the caspase-1 subunit of the inflammasome, leads to the secretion of active IL-1β cytokine. This cytokine then stimulates the production and secretion of inflammatory mediators and pro-inflammatory cytokines that initiate the inflow of neutrophils into the affected joints. These neutrophils secret toxin substances that eventually caused gout symptoms (113).

IL23R has a subunit called IL23 subunit, it is a pro-inflammatory cytokine that is mostly made by active macrophages and dendritic cells. IL23R has a heterodimer structure made up of IL-12p40 and IL-23p19 molecule, and the levels of mRNA expression of p40 and p19 subunit have been proven to have elevated level in inflammation locations. IL-23R is mostly expressed in active memory T cells, NK cells, monocytes, macrophages and in dendritic cells compared to the other IL23 receptor (i.e. IL-12RBl). The IL-23 subunit has a crucial role in regulation of gout inflammatory mechanism by binding to IL-23R, and forming a regulatory complex, it promotes the T-cell-mediated inflammatory response. It also works alongside other related inflammatory cytokines to stimulate the innate immune response that leads to gout stimulation, and is involved in IL-17 activation (IL-17 is secreted from Th17 cells), and activates the release of other inflammatory mediators and cytokines, which includes:
IL-6, IL-8 and TNF-α). This eventually worsens the gout inflammatory reaction and caused chronic inflammation, chronic gout at the same time causes tissue damage (113, 227, 228).

Because the IL-23R rs7517847 (G/T), rs11465804 (G/T) and rs11209026 (A/G) polymorphism is associated with RA, ankylosing spondylitis, inflammatory bowel disease (i.e. Crohn’s disease, Ulcerative colitis and Celiac disease) and other systemic autoimmune diseases, we examined if it was also associated with gout in European and Polynesian sample sets.

**1.24.1 RS11465804 & RS11209026**

Analysis of SNP rs11465804 and rs11209026 in the gout data set shows no evidence of association with gout in European and Polynesian population (RS11465804 European OR = 0.85, P-value = 0.24; Polynesian OR = 0.72, P-value = 0.19; RS11209026 European OR = 0.96, P-value = 0.78; Polynesian OR = 0.56, P-value = 0.05 respectively). Adjustment for confounders (i.e. Polynesian ancestry, sex and age) using a logistic regression mode was conducted and revealed that both SNPs still shows no evidence of association in both sample sets (RS11465804 European OR = 0.96, P-value = 0.82; Polynesian OR = 0.84, P-value = 0.59; RS11209026 European OR = 1.03, P-value = 0.83; Polynesian OR = 0.85, P-value = 0.67 respectively). Meta-analysis of European and Polynesian population data sets support the SNPs non-significant association with gout (RS11465804 OR = 0.99, P\text{overall} = 0.958; RS11209026 OR = 1.05, P\text{overall} = 0.752). According to the power calculation result, both SNPs power output was relatively low compared to SNP rs7517847 (RS114658043 European OR 1.2, Polynesian OR 1.2 = 0.12 = 0.29, RS11209026 European OR 1.2 = 0.28; Polynesian OR 1.2 = 0.12 respectively). This indicates that rs11465804 and rs11209026 have either low minor allele frequency or sample sets. This may have been why we were not able to detect any significant association with gout in both sample sets. However, European power output is slightly greater than Polynesian this suggests that European have more MAF as well as numbers of their sample sets than Polynesian.

Moreover, SNP rs11465804 and rs11209026 has not been associated with hyperuricaemia or gout in any previous studies that been documented to date. However, many previous studies demonstrate the association of either major or minor allele of this two SNPs with other related auto-inflammatory
disease such as inflammation bowel disease, psoriasis, Crohn's disease, ankylosing spondylitis, ulcerative Colitis, psoriasis and celiac disease. Refer to Section 1.6.1.1.1 (rs11465804) and 1.6.1.1.2 (rs11209026) and Table 1.6:1 for previous study demonstrating association of this related auto-inflammatory disease with IL23R SNP rs11465804 and rs11209026.

1.24.2 RS7517847

Analysis of SNP rs7517847 in the gout data set showed no evidence of association with gout in Polynesian and European populations (European OR = 0.94, P-value = 0.35; Polynesian OR = 0.92, P-value = 0.18 respectively). However, after adjustment for confounders the SNP showed a protective association of the major allele T with gout in Polynesian population but still no sign of association with European population was seen(European OR = 0.95, P-value = 0.53; Polynesian OR = 0.85, P-value = 0.04 respectively). Therefore, rs7517847 minor allele G has a susceptible association with gout. As the Polynesian population showed evidence of association with gout in rs7517847 we divided the, Polynesian population into 3 sub-groups (i.e. WP, EP, and EPWP) and performed logistic regression analysis (with adjustment for confounders). This was done to test whether the Polynesian sub-group was more susceptible to develop clinical gout. This analysis was conducted due to the fact that allele and genotype distribution in Polynesian people are different. The Polynesian sub-groups analysis showed that EP (including people from Hawaii, Cook Island and New Zealand) and EPWP population have no evidence of association with gout while the WP population (include people from Samoa, Tonga, Niue and Tokelau) showed evidence of association (OR = 0.76, P –value = 0.04). RS7517847 meta-analysis did not show significant association of Polynesian and WP population with gout (OR = 0.93, Poverall = 0.196).

A power calculation showed rs7517847 have a very high power output for European, Polynesian, EP, WP except for EPWP population (European OR 1.2 = 0.79, Polynesian OR 1.2 = .79, EP OR 1.2 = 0.55, WP OR 1.2 = 0.49, EPWP OR 1.2 = 0.09 respectively). This may mean that European, Polynesian, EP, and WP dataset had a high amount of minor allele frequency and also have a large population size except for EPWP population which have a small population size. This explains why
WP analysis detected an association with gout while EPWP have such a low power output and show no sign of association with gout. A numbers of studies show evidence of association of either minor or major allele of rs7517847 with other related auto-inflammatory disease especially Crohn’s disease and ankylosing spondylitis (Refer to Section 1.6.1.1.3 and Table 1.6:1).

The susceptible association of rs7517847 minor allele G with gout in Polynesian population replicate the data published by Liu et al (2015). This study was conducted on 400 patients with gout and 582 controls recruited from Affiliated Hospital of Qingdao University Medical College (Qingdao, Shandong, China). The aim of the study was to illustrate if IL23R SNP rs7517847 (G/T) have an association with gout in Chinese Han male. The SNPs were genotyped using TaqMan probes assay. They conduct an association analysis using the x² test. They also performed genotype-phenotype analysis. Logistic regression mode was used to adjust for age in the study sample set. The association analysis result revealed that SNP rs7517847 minor allele G and genotype G/G had a susceptible association with gout (minor genotype frequency G/G OR = 1.2, P-value = 0.034; minor allele frequency OR = 1.2, P-value = 0.040) in Chinese Han population. The genotype-phenotype analysis showed that cases with G/G genotype have shorter disease time than cases than cases with G/T, and GT + TT genotype. This analysis also demonstrated that waist to hip ratio of cases with G/G genotype are lower than cases with G/T and T/T genotype. Therefore, they concludes that IL23R might have a significant association with gout in the Chinese Han male population, however their findings should be confirmed using larger sample sizes and other independent populations with variety of ethnicities.

A study by Kageyama and colleagues (2007) (229) showed that the soluble TNF-α receptor called etanercept, has a promising effect on serum levels of IL-16, IL-17, IL-23, and macrophage inflammatory protein-3alpha (MIP-3alpha) in rheumatoid arthritis (RA) cases. Their finding showed that IL-23, had elevated levels of urate in RA synovial fluid cases and reduced urate level in SF of osteoarthritis, pseudogouty arthritis and gouty arthritis cases. This suggests that IL-23 have a function in the RA pathogenesis.
1.25 Secondary Analysis

This analysis was conducted to examine the association between controls participants that had high serum urate level ≥ 0.42 mmol/L with gout in IL23R SNP of interest dataset. Result reveals that all three SNPs dataset in European and Polynesian population showed no significant association with gout. This finding suggests controls with hyperuricaemia have IL23R locus with variants that have protective effects on gout development.

1.26 Haplotype Analysis

Haplotype analysis was conducted between SNP rs11465804 and SNP rs11209026 because of their strong LD ($r^2$) in European ($r^2 = 79$) and moderate $r^2$ in Polynesian ($r^2 = 43$) population. Strong LD means two or more SNPs are associated and inherited from the same chromosomes and their association could either increased risk for a certain disease or reduced risk. SNPs with weak LD ($r^2 < 40$) (i.e. SNP rs7517847 with $r^2 = 3$ in European and 0 in Polynesian) indicated that a SNP acts independently for either increasing or decreasing risk of a disease. Haplotype result showed all allele combination of rs11465804 and rs11209026 do not statistically significantly confers an association with gout in both European and Polynesian population. However, European allele combination of T/A, T/G and Polynesian allele combination of G/G and T/G have odd ratios trending in the susceptible direction (OR > 1).
1.27 Association Analysis Rheumatic Heart Fever Dataset

RHF is a systemic auto-inflammatory disease that develops as a result of an unusual immune response due to group A-haemolytic streptococci (GABHS) infection. The infection initially occurs in the upper respiratory tract where the streptococcus infects the throat, resulting in a sore throat. *S.pyogenes* infect the upper respiratory tract and are known to be involved in the immune response that leads to RHF. This is achieved via an antigen molecular mimicry mechanism and cross reactivity mechanism between the host and bacteria. Cross reactivity of antibodies and/or T cell stimulate recognition between the *S.pyrogenes* peptides and the host protein leads to inflammation and autoimmunity. CD4+ T cells are the key effectors of heart tissue autoimmune reactions in RHF. During RHF, the expression of vascular cell adhesion protein 1 (VCAM-1), a protein that is upregulated by self-antigen-specific antibodies, helps lymphocytes increase their adhesion. This suggests that VCAM-1 also contributes to inflammation and valvular scarring associated with RHF (145). Normally, the self-reactive B cells remain silent in the periphery that means that T cell do not stimulate it. However, in RHF cases antigen presenting cell such as B cell present *S.pyrogenes* antigen on to CD4-T cells that eventually differentiate to T2 helper cell (i.e. Th17 cell that secrete IL-17 and IL-22 cytokines). Subsequently, T2 helper cell stimulate B cell activation which will go on and activate the production and secretion of antibodies to stimulate cells (e.g. natural killer cell, macrophage and neutrophil) that attack *S.pyrogenes* cell walls. This cells may also attack myocardium and joints which produced RHF symptoms.

Because the IL-23R rs11465804 (G/T), rs11209026 (A/G) and rs7517847 (G/T) polymorphism is associated with RA, ankylosing spondylitis, inflammatory bowel disease (i.e. Crohn’s disease, Ulcerative colitis and Celiac disease) and other systemic autoimmune diseases we examined if it was also associated with RHF in Polynesian sample sets. In addition many studies illustrate the possible role of Th17 cell in the IL23R genetic variants that leads to the development of related auto-inflammatory disease such as Crohn’s disease, psoriasis, ulcerative colitis and inflammatory bowel disease (230-233). This association is caused because IL23 subunit of IL23R has a role in the secretion
of IL-17 from Th17 cells which leads to the secretion other enhancer inflammation factors which eventually elevate inflammation reaction and initiated auto-inflammatory disease.

1.27.1 RS11465804

RS11465804 appeared to confer a significant association with RHF in the Polynesian population (OR = 0.11, P-value = 0.0047). However, after adjustment for confounders (i.e. Polynesian ancestry) using logistic regression model the association lost its significant with RHF (OR = 0.90, P-value = 0.92). This indicates that participants within case and controls are poorly matched. Poorly matched case and controls means that they have a different number of people with Polynesian ancestry background. 

Note that the RHF dataset was only adjusted for Polynesian ancestry only. Power calculation reveals that rs11465804 has quite low power. This is because of a small population size and low minor allele frequency. This may be the reasons as to why significant association was not found. No previous study illustrate that rs11465804 is associated with RHF only studies that showed association of rs11465804 variants with other auto-inflammatory and autoimmune disease such as Rheumatoid arthritis, Crohn’s disease and ankylosing spondylitis Refer to Section 1.6.1.1.1 and Table 1.6:1 for previous studies demonstrating association of rs11465804 and related auto-inflammatory and autoimmune disease. Moreover, a study by Schmechel and colleagues (2008) (230) showed that IL23R variants (i.e. rs11465804) genotypes have a significant effect on IL—223 serum expression level in Crohn’s disease cases. Therefore, they concludes, that Th17 cell function have a susceptible risk for Crohn’s disease.

1.27.2 Rs7517847

No evidence of association was shown before or after adjustment. Power output for rs7517847 was also quite low but it higher than the other two SNPs. This suggest that rs7517847 have more minor alleles but population size is small. To confirm that this SNP do not associated with RHF, we need a large population size. No previous study illustrate that rs7517847 is associated with RHF only studies

3 Cytokines released by Th17 cell
that showed association of rs7517847 variants with other auto-inflammatory and autoimmune disease. Refer to Section 1.6.1.1.3 and Table 1.6:1 for previous studies demonstrating association of rs7517847 and related auto-inflammatory and autoimmune disease. Furthermore, SNP rs7517847 was one of the 10 IL23R SNPs used in Schmechel and colleagues study that concluded: IL23R variants (i.e. rs7517847) have a confer association with IL-22 serum expression level in Crohn’s patients therefore, Th17 cell function have a susceptible risk for Crohn’s disease (230).

1.27.3 RS11209026

Evidence of a protective association with RHF was shown before and after adjustment (unadjusted OR = 0.13, P-value = 0.009; adjusted OR = 0.07, P-value = 0.002). Hence, the rs11209026 major allele (G) has a susceptible effect for RHF in Polynesian population. No previous study showed this data before however, few previous studies showed that rs11209026 major allele (G) have a susceptible effect on other auto-inflammatory and auto-immune related disease such as rheumatoid arthritis and Crohn’s disease on selected population while minor allele (A) appeared to be in protective direction (193, 195, 196, 209) (Refer to Section 1.6.1.1.2 and Table 1.6:1). However, the power output of rs11209026 is really low because of small population size and low minor allele frequency. Therefore, a large sample set is needed to validate this significant association. In addition, rs11209026 gene variants was one of the 10 IL23R SNPs used in Schmechel and colleagues study that concluded that: IL23R variants (i.e. rs11209026) confer an association with IL-22 serum expression level in Crohn’s patients therefore, Th17 cell function have a susceptible risk for Crohn’s disease (230).

1.28 Limitations in the Study

The gout and RHF datasets used in this study have have small population sizes. The main issue with small population studies is the interpretation of results, mainly the analysis of P-values and confidence intervals. For rare variants, the small population sizes can fail to identify any true association (234). Also, interpretation of results from small studies can yield false positive results or can overestimate the strength of the true association (235). Any non-significant association that was observed could be
due to the low sample size in each population subset. The recruitment of samples is still on-going, which will increase the power of this study.

1.29 Conclusion

The analyses conducted in this study showed that SNP rs7517847 minor allele (G) has a significant association (after adjustment with confounders) with gout in the Polynesian people. Therefore, this result supports a role of IL23R in the development of gout in Polynesian population and in Western Polynesian peoples. This replicate Liu et al (2015) findings that SNP rs7517847 minor allele (G) showed evidence of association with gout in Chinese Han male population however, need a large sample size and variety of ethnicities to confirm the significant of the findings (112). Haplotype analysis (rs11465804 and rs11209026) of the gout data set did not demonstrate significant association in both population suggesting that this haplotype analysis did not improve the analysis.

The RHF dataset analyses showed that SNP rs11209026 major allele (G) has a significant association (after adjustment with confounders) with RHF in the Polynesian population. This indicates that IL23R may have a causal role for RHF development in Polynesian people. On the other hand, SNP rs11465804 lost its association with RHF after adjustment with ancestry confounders. This showed that case and control individuals are sub-optimally matched in terms of Polynesian ancestry distributions.

These findings may help improve the knowledge and understanding on the genetic and biological basis underlying gout and RHF in the Polynesian population and have implications for management and treatment for these debilitating diseases. Hence, IL23R pathway is a target for gout and RHF treatment in people of Polynesian ancestry. There is perhaps value in individualising gout and RHF management/treatment because of genetic differences.
Appendix

1.30 Appendix 1

1.30.1 R-studio command

Naming input file

> View (IL23R.GENOTYPES. AND.PHENOTYPES.SEPT2015)
  > View (IL23R)

Create Population subset

> European <- subset (IL23R, CAUvsPOLY==1)
  > Polynesian <- subset (IL23R, CAUvsPOLY==2)

> EP <- subset (IL23R, ETHCLASS==1)
  > WP <- subset (IL23R, ETHCLASS==2)
> EPWP <- subset (IL23R, ETHCLASS==9)

Calculation Allele and Genotyping Distribution of Polynesian sub-groups

> EP_highGP <- subset (EP, GPANCESTRY>=0.75)

> table (EP_highGP$RS11465804_Genotype)

Output Result:

Table 1.30:1 Example of an Genotype frequency table showing homozygous minor (G/G represent by 3), heterozygous (G/T represent by 2) and homozygous major (T/T represent by 1).

<table>
<thead>
<tr>
<th></th>
<th>TT</th>
<th>GT</th>
<th>GG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>721</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.30:2 Example of a allele frequency table show count of major and minor allele.

Frequency is calculated by dividing allele count by the total number of alleles.

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>23</td>
<td>0.015478</td>
</tr>
<tr>
<td>T</td>
<td>1463</td>
<td>0.984522</td>
</tr>
<tr>
<td>Total</td>
<td>1486</td>
<td></td>
</tr>
</tbody>
</table>

Calculating unadjusted P-value

```r
> Cau_rs114 = glm (AFFECTION~RS11465804_Genotype, data= European, family=binomial)
> Summary (Cau_rs114)
```

**Output Result:**

Call: glm (formula = AFFECTION ~ RS11465804_Genotype, family = binomial, data = European)

**Table 1.30:3 Result of unadjusted P-value (Pr (>l z l) of RS11465804_Genotype (Highlight yellow)**

| Coefficients:     | Estimate | Std. Error | z value | Pr (>|z|) |
|-------------------|----------|------------|---------|----------|
| (Intercept)       | 0.2099   | 0.1654     | 1.269   | 0.204    |
| RS11465804_Genotype | -0.1670  | 0.1425     | -1.172  | 0.241    |

Calculating unadjusted Odds Ratio.

```r
> exp (coefficients (Cau_rs114))
```

**Output Result:**

**Table 1.30:4 Result of unadjusted odds ratio (Highlight yellow) of RS11465804_Genotype**

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>RS11465804_Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2335139</td>
<td>0.8462311</td>
</tr>
</tbody>
</table>
Calculating 95% Confident Interval

> exp (confint (Cau_rs114))

Output Result:

Table: 1.30: 5 Result of unadjusted 95% (Highlight yellow) CI RS11465804_Genotype

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.8924103</td>
<td>1.707586</td>
</tr>
<tr>
<td>RS11465804_Genotype</td>
<td>0.6391413</td>
<td>1.118317</td>
</tr>
</tbody>
</table>

Calculating adjusted P-value

Confounder variables which are AGE, SEX, and GPANCESTRY (ancestry) are adjusted for in Gout samples set but for RHF samples set, only GPANCESTRY will be adjusted for.

> Cau_rs114adj = glm (AFFECTION~RS11465804_Genotype+AGE+SEX, data=European, family=binomial)

> summary (Cau_rs114adj)

Output Result:

Call: glm (formula = AFFECTION ~ RS11465804_Genotype + AGE + SEX, family = binomial, data = European)

Table 1.30: 6 Result of adjusted P-value (Highlight yellow) of RS11465804_Genotype.

It has been adjusted with AGE + SEX +GPANCESTRY

|                | Estimate | Std Error | z value | Pr(>|z|) |
|----------------|----------|-----------|---------|----------|
| (Intercept)    | -1.274796 | 0.312722  | -4.076  | 4.57e-05 *** |
| RS11465804_Genotype | -0.040363 | 0.162483  | -0.248  | 0.804    |
| AGE            | 0.053452  | 0.003561  | 15.011  | <2e-16 ** |
Calculating adjusted Odd Ratio [command]

> exp (coefficients (Cau_rs114adj))

Output Result:

Table 1.30:7 Result of adjusted Odd Ratio (Highlight yellow) of RS11465804_Genotype after adjustment

<table>
<thead>
<tr>
<th>Intercept</th>
<th>RS11465804_Genotype</th>
<th>AGE</th>
<th>SEX</th>
<th>GPANCESTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2794880</td>
<td>0.9604409</td>
<td>1.0549066</td>
<td>0.2629620</td>
<td>0.143256</td>
</tr>
</tbody>
</table>

Calculating adjusted 95% Confident Interval [Command]

1.31 Appendix 2

1.31.1 Meta-Analysis Command:

> rs114meta<read.delim("~/Desktop/rs114meta.txt", stringsAsFactors=FALSE
> Library (meta)
> rs114metareresults=metagen
  (TE=BETA,seTE=SE,studlab=STUDY,sm="OR",data=rs114meta)
> rs114metareresults

This command will calculate Fixed effect model and Random effects model OR, Z-value, P-value and 95% CI. Also calculated the Heterogeneity test.
Output Result:

Table 1.3.1 Odd ratio, 95% CI interval, %W(fixed) and %W(random) value of NZ European and NZ Polynesian population individually.

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95%-CI</th>
<th>%W(fixed)</th>
<th>%W(random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European</td>
<td>0.9604</td>
<td>[0.6985; 1.3206]</td>
<td>64.71</td>
<td>64.71</td>
</tr>
<tr>
<td>NZ Polynesian</td>
<td>0.8422</td>
<td>[0.5472; 1.2962]</td>
<td>35.29</td>
<td>35.29</td>
</tr>
</tbody>
</table>

Number of studies combined: k=2

Table 1.3.2 The combine result of 2 population. Showing Fixed and Random effects OR, 95% CI interval, z value and p-value.

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95%-CI</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect model</td>
<td>0.9169</td>
<td>[0.7097; 1.1846]</td>
<td>-0.6636</td>
<td>0.507</td>
</tr>
<tr>
<td>Random effects model</td>
<td>0.9169</td>
<td>[0.7097; 1.1846]</td>
<td>-0.6636</td>
<td>0.507</td>
</tr>
</tbody>
</table>

Quantifying heterogeneity:

\[
tau^2 = 0; \quad H = 1; \quad I^2 = 0\%
\]

Table 1.3.3 Heterogeneity test for 2 population

<table>
<thead>
<tr>
<th>Test of heterogeneity:</th>
<th>Q d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.23</td>
<td>0.631</td>
</tr>
</tbody>
</table>

Note: If the heterogeneity result (Table 1.3.3) is significant (P-value < 0.05) random effect model Odd ratio and P-value will use. If it not significant (P-value > 0.05) fixed effects model OD and P.value will use (Table 1.3.2).

For this example, heterogeneity result is no significant meaning that differences between data set is not huge. Therefore, fixed effects OR (95% CI) and P-value are ideal data.

Prepare a forest plot (Figure 1.21:1) of the meta-analysis result using this command:

```
ml = metagen (TE=BETA, seTE=SE, studlab=STUDY, sm="OR", data=sexagebmi_me)
```
ml 

forest(ml, comb.random=FALSE, print.Q=TRUE, print.pval.Q=TRUE, print.ta u2=FALSE, print.I2=FALSE, hetlab="Heterogeneity:", xlab="OR", digits=3, text.fixed="Overall (95% CI)", smlab=" ")

1.32 Appendix 3

1.32.1 Secondary Analysis R Studio Command

P-value command:

Call:

glm(formula = HUVsGOUTAFFECTION ~ RS11465804_Genotype, family = binomial, data = European)

Output:

Call: glm (formula = AFFECTION ~ RS11465804_Genotype, family = binomial, data = European)

Table:1.32:1 Result of P-value (Pr (>l z l) of RS11465804_Genotype (Highlight yellow)

| Coefficients: | Estimate | Std. Error | z value | Pr (>|z|) |
|---------------|----------|------------|---------|---------|
| (Intercept)   | 0.2099   | 0.1654     | 1.269   | 0.204   |
| RS11465804_Genotype | -0.1670 | 0.1425     | -1.172  | 0.935   |

Odd Ratio command:

> exp(coefficients(Urate_Cau_rs114))

Output:

Table:1.32:2 Result of unadjusted odds ratio (Highlight yellow) of RS11465804_Genotype

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>RS11465804_Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2335139</td>
<td>0.9732311</td>
</tr>
</tbody>
</table>
95% Confidence Interval Command:

> exp(confint(Urate_Cau_rs112))

Output:

**Table:** Result of unadjusted 95% (Highlight yellow) of CI Urate_Cau_rs112.

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.8924103</td>
<td>1.707586</td>
</tr>
<tr>
<td>RS11465804_Genotype</td>
<td>0.6391413</td>
<td>1.118317</td>
</tr>
</tbody>
</table>

### 1.33 Appendix 4

#### 1.33.1 Haplotype PLINK Command

Locate were your Ped and Map at command

```bash
cd /Users/tanyaflynn/Executables/plink
```

N.B: you can drag and drop the folder onto terminal to make it automatically write the file path for you.

To get case/control frequencies:

```
plink --ped CAUdata.ped --map CAUdata.map --hap-assoc --hap-window 2 --noweb --missing-phenotype -9 --allow-no-sex --out haplotypesKeresoma
```

To get the odds ratio and P-value:

```
plink --ped CAUdata.ped --map CAUdata.map --hap-logistic --hap-window 2 --noweb --missing-phenotype -9 --allow-no-sex --out haplotypesKeresoma2
```

N.B: in the file called haplotypesKeresoma2.assoc.hap.logistic you want to look at the OR and P columns
N.B: If you want to look at all possible haplotypes, not just the most frequently seen ones add the command --mhf 0 to both those commands.
References


149. Institute of Environmental Science and Research Limited NaodiN, Zealand. Annual Report WMoH.


189. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.


