GENE EXPRESSION AND METHYLATION OF MATRIX METALLOPROTEINASE-13 AND OSTEOARTHRITIS
&
BURDEN OF RISK ALLELES FOR COMPLEX TRAITS IN THE NEWFOUNDLAND AND LABRADOR POPULATION

by

© Patricia Emily Rose Harper

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

Graduate Program in Human Genetics, Faculty of Medicine

Memorial University of Newfoundland

October 2015

St John's Newfoundland and Labrador

Abstract

PART I

To evaluate the role of *MMP-13*, a catabolic enzyme, in osteoarthritis (OA) pathogenesis, we utilized a case-control study design to analyze cartilage samples obtained from 48 OA patients and 30 healthy controls. We found that *MMP-13* expression was significantly increased in OA-affected cartilage compared to healthy cartilage. A CpG site in the promoter of *MMP-13* was also significantly demethylated in osteoarthritic cartilage but was not correlated with the gene expression. The results suggest a potential role for *MMP-13* in OA pathogenesis.

PART II

Founder populations, such as Newfoundland & Labrador (NL), often have higher prevalences of genetic disease, making them ideal for disease gene mapping. Mean frequencies of common disease-associated alleles from over 140 diseases were evaluated in the NL population. The mean risk allele frequencies of coronary heart disease, celiac disease and bladder cancer were elevated in NL when compared to an out-bred population. The elevated disease allele frequencies were in concert with the high prevalence of the diseases in NL. This novel approach to studying disease risk in a population may provide new insights into genetic risk of complex conditions and could be used to inform health policy.

Acknowledgements

I would like to thank my supervisor Dr. Guangju Zhai for accepting me into the program. Your support and guidance throughout my time as a graduate student helped me to succeed and develop my understanding and love of genetics. To my supervisory committee members, Dr. Roger Green and Dr. Guang Sun, thank you for your constructive advice and encouragement with my research.

To my lab mates, thank you for the valuable advice and resources in support of our team. Especially Maggie Liu whose organization and energy helped keep things in the lab running (relatively) smoothly and provided important feedback in the lab and on my work.

To the study participants, hospital staff, surgeons, nurses and everyone who contributed their resources to helping the project, thank you. This work was supported by the Canadian Institutes of Health Research, Research & Development Corporation of NL and Memorial University of Newfoundland.

Lastly, thank you to my family, Anne, Mark, Sarah, Philip, and my friends for their love and encouragement through this long process. Your unwavering confidence in me is greatly appreciated.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

ACAAV	Antineutrophil cytoplasmic antibody-associated vasculitis
ACR	American College of Rheumatology
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
AF	Atrial fibrillation
AGE	Advanced glycation end products
ALL	Acute Lymphoblastic Leukemia
ALS	Amyotrophic Lateral Sclerosis
AS	Ankylosing Spondilitis
ASTN2	Astrotactin 2
BCC	Basal Cell Carcinoma
BMI	Body mass index
BMP	Bone Morphogenic Protein
bp	base pair
BRCA	Breast cancer gene
Ca	Cancer
cAMP	Cyclic adenosine monophosphate
CEPH	Centre d'Etude Polytechnique
CEU	Northern Europeans from Utah
CHD	Coronary Heart Disease
CH_3	Methyl group
CHST11	Carbohydrate Sulfotransferase 11
CLL	Chronic Lymphocytic Leukemia
COL12A1	Collagen, type XII, alpha 1
CpG	Cystosine-phosphate-Guanidine
CRC	Colorectal Cancer
CREB	cAMP response element-binding protein

CRP	c-reactive protein
DMM	Destabilization of the medial meniscus
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide
DOTL1	Disruptor of telomeric silencing- like protein
DVWA	A protein with double von Willebrand factor A domain
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPMR	Progressive epilepsy with mental retardation
ERK	Extracellular signal-regulated kinases
ESR	Erythrocyte sedimentation rate
FADS	Fetal akinesia deformation sequence
FDA	Food and Drug Administration
FILIP1	Filamin A interacting protein 1
FTO	Fat mass and obesity-associated protein
g	Gravitational force
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GDF5	Growth/Differentiation Factor 5
GWAS	Genome-Wide Association Study
HHGS	Histological Histochemical Grading System
HLA	Human Leucocyte Antigen
HREA	Health Research Ethics Authority
IBD	Inflammatory Bowel Disease
IBD (E-O)	Inflammatory Bowel Disease (Early-onset)
IL-1	Interleukin-1
JSN	Joint space narrowing

kb	kilobases
KLHDC5	Kelch Domain Containing 5
K-L	Kellgren-Lawrence
LD	Linkage Disequilibrium
LOAD	Late-onset Alzheimer's Disease
LN_2	Liquid nitrogen
LoF	Loss-of-function
MAP	Mitogen-activated protein kinase
Mb	Megabase
MCF2L	MCF2 cell line derived transforming sequence-like, an oncogene
MI	Myocardial Infarction
miRNA	micro-RNA
mM	millimolar
MM	Multiple Myeloma
MMP	Matrix Metalloproteinase
mL	millilitres
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
mtDNA	mitochondrial DNA
NCOA3	Nuclear Receptor coactivator 3
NFOAS	Newfoundland and Labrador Osteoarthritis Study
NL	Newfoundland and Labrador
NSAID	Non-steroidal anti-inflammatory drugs
OA	Osteoarthritis
OACH	Osteoarthritis Cartilage Histopathology
OARSI	Osteoarthritis Research Society International
OI	Osteogenesis Imperfecta

OMIM	Online Mendelian Inheritance in Man database
PCH2	Pontocerebellar hypoplasia type 2
PCR	Polymerase Chain Reaction
pH	power of Hydrogen
PTHrP	parathyroid-hormone-related-protein
PTHLH	parathyroid hormone-like hormone
QC	Quality Control
RA	Rheumatoid Arthritis
RCDP1	rhizomelic chondrodysplasia punctata type 1
ROS	Reactive Oxygen Species
RQ	Relative quantity
rt-PCR	real-time polymerase chain reaction
RUNX2	Runt-related transcription factor 2
RNA	Ribonucleic acid
SAM	S-adenosyl methionine
SAP	Shrimp alkaline phosphatase
SENP6	sentrin specific peptidase 6
SLE	Systemic Lupus Erythematosus
SMAD3	Mothers against decapentaplegic 3
SNP	Single Nucleotide Polymorphism
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TET	Ten eleven translocation protein
TGF-β	Transforming growth factor-beta
TP63	Tumour protein p63
TNF	Tumour Necrosis Factor
TSFM	Translation Elongation Factor, Mitochondrial gene

TSS	Transcription Start Site
T1D	Type I Diabetes
T2D	Type II Diabetes
UC	Ulcerative colitis
UK	United Kingdom
US	United States
vLINCL	Variant form of late infantile neuronal ceroid lipofuscinosis
VT	Venous Thrombosis
WHO	World Health Organization
WOMAC	Western Ontario and McMaster university osteoarthritis index
5mC	5-methyl-cytosine
5hmC	5-hydroxymethyl-cytosine
µL	microliter

List of Appendices

1 PART I: GENE EXPRESSION AND METHYLATION IN THE PROMOTER OF MATRIX METALLOPROTEINASE-13 AND OSTEOARTHRITIS

1.1 CHAPTER 1: Introduction

1.1.1 Osteoarthritis

1.1.1.1 Definition

Osteoarthritis (OA) is described as a chronic disorder of the muscoskeletal system. It most commonly affects joints of the knee, hip, hand, foot and spine. Characteristics of OA include the degradation of cartilage, synovial inflammation, subchondral bone remodelling, osteophyte (bone spur) growth, degeneration of ligaments and hypertrophy of the affected joint (Poole, 2012). These processes result in inflammation, significant pain, deformation and a progressive loss of joint function (Arthritis Alliance of Canada, 2011).

1.1.1.2 Diagnosis

OA is diagnosed by physical examination in conjunction with radiographic evidence by X-Ray or magnetic resonance imaging (MRI). Several scoring systems exist for assessment of radiographic OA; one of the most commonly used is the Kellgren-Lawrence scale (K-L grade) developed in 1957. It uses five grades (0 – 4) to assess OA severity based primarily on joint space narrowing (JSN) and osteophyte growth. A grade of 2 or more is classified as radiographic OA (Table 1) (Kellgren & Lawrence, 1957).

Table 1.1 The Kellgren-Lawrence System for Classification of Osteoarthritis

Grade 0	No radiographic features of osteoarthritis
Grade 1	Doubtful narrowing of joint space, possible osteophytic growth
Grade 2	Possible joint space narrowing and definite osteophyte
Grade 3	Definite narrowing of joint space, multiple moderate osteophytes and minor sclerosis and possible deformity of bone contour
Grade 4	Marked joint space narrowing, large osteophytes, severe sclerosis and definite deformity of bone contour

Another commonly used scoring system is the Osteoarthritis Research Society International (OARSI) atlas, which uses semi-quantitative scoring to establish a grade from 0 to 3. Osteophyte growth and JSN are evaluated separately and then combined for a

final grade (R. D. Altman, Hochberg, Murphy, Wolfe, & Lequesne, 1995; R. D. Altman & Gold, 2007; Culvenor, Engen, Oiestad, Engebretsen, & Risberg, 2014). To reach the cut-off, one of three separate criteria must be met, either a JSN grade of 2 or higher, sum osteophyte growth of grade 2 or higher, or JSN grade 1 combined with osteophyte grade 1, and includes lateral and medial tibiofemoral compartments separately (R. D. Altman et al., 1995; R. D. Altman & Gold, 2007; Culvenor et al., 2014). The OARSI atlas method of evaluating JSN and osteophyte growth separately provides more flexibility than the K-L grade in the diagnosis of radiographic OA.

Laboratory tests are often ordered to support a diagnosis of OA or eliminate other alternatives. Tests for C–reactive protein (CRP) or erythrocyte sedimentation rate (ESR) are among those often used, although they are only effective as indicators of inflammation and are not site or disease specific (S. Singh, Kumar, & Sharma, 2014). The American College of Rheumatology (ACR) developed criteria for diagnosing OA that includes clinical and laboratory features to identify symptomatic OA patients (R. Altman et al., 1991). An example of the criteria used to diagnose hip OA is as follows. Classification of a patient with hip OA includes pain with 1) internal hip rotation of ≥ 15°, hip morning stiffness for ≤ 60 minutes, and greater than 50 years of age, or 2) internal hip rotation of < 15° and ESR of ≤ 45 mm/hour or hip flexion of ≤ 115 °. When radiographic evidence is included with clinical evidence, pain with at least 2 of the 3 following criteria: osteophytes, JSN, and ESR < 20 mm/hour, a doctor can establish a diagnosis.

The Western Ontario and McMaster university osteoarthritis index (WOMAC scores) can also be used to assess pain, disability and joint stiffness for knee OA (see appendix). Histological grading systems have also been developed to assess the pathology of the articular surface. The OA Histological Histochemical Grading System (HHGS) introduced by Mankin *et al*. in 1971, and the Osteoarthritis Cartilage Histopathology (OACH) scoring system established in 1998, are commonly used. The Mankin score uses a sum of four categories; surface structure, cells, safranin-O staining and tidemark

integrity to give a scale range between 0 and 14 (Pearson, Kurien, Shu, & Scammell, 2011), while OACH is a product of grade and stage, giving a scale of 0 to 24 (Pritzker et al., 2006).

1.1.1.3 Epidemiology of Osteoarthritis

OA is the most common form of arthritis. According to the World Health Organization (WHO), OA affects approximately 9.6% of men and 18% of women globally (World Health Organization, 2011). Currently, more than 4.6 million Canadians age 15 years and older have symptoms of arthritis (The Arthritis Society, February 2015). While in 2010, the prevalence for OA in Canadian men was 10.6% and 15.5% for women, and 13.0% overall (Arthritis Alliance of Canada, 2011). By 2030 OA is estimated to be the greatest cause of disability worldwide (Thomas, Peat, & Croft, 2014) and by 2035, one-in-four Canadians is expected to be living with arthritis (Arthritis Alliance of Canada, 2011). The economic burden for direct and indirect health-care costs related to arthritis in Canada is predicted to reach $233.5 billion in a five year period from 2010 to 2015 (The Arthritis Society, February 2015), while the United States reaches 128$ billion each year (Yelin et al., 2007). Over 80% of hip replacement surgeries and 90% of knee replacement surgeries performed in Canada are attributed to OA (The Arthritis Society, February 2015). This represents a significant economic burden, with disability due to OA rivalling that of cardiovascular disease (Guccione et al., 1994). Individuals diagnosed with arthritis are also more likely to have other chronic health conditions such as diabetes, high blood pressure, heart disease, cancer, or suffer from a stroke or poor mental health (Arthritis Community Research and Evaluation Unit (ACREU), 2013a; The Arthritis Society, February 2015).

In Newfoundland and Labrador, approximately 95,000 individuals aged 15 years or older report having arthritis, this accounts for 22% of the population (Arthritis Community Research and Evaluation Unit (ACREU), 2013b), higher than the national average at 16 % (Public Health Agency of Canada, 2011). This number is expected to

increase from 95,000 to 127,000 by 2036 (Arthritis Community Research and Evaluation Unit (ACREU), 2013b). OA can significantly reduce mobility, 38% of Newfoundlanders report arthritis pain preventing activities, in contrast to only 12% from people with other chronic conditions (individuals aged 45 years and older) (Arthritis Community Research and Evaluation Unit (ACREU), 2013b). The aging population, both in Newfoundland and Canada-wide means these numbers will continue to rise in the coming years, representing a significant increasing socioeconomic burden to society.

1.1.1.4 Aetiology of Osteoarthritis

The primary risk factor for developing OA is aging. As evidenced in a cohort study by Jordan *et al*., the presence of radiographic knee OA rose almost 25% between age groups of 55-64 to 75 + and symptoms of knee OA increased by 15 % between these groups (Jordan et al., 2009). Other risk factors include prior injury, abnormal biomechanics, sex and obesity. Women are more likely to develop OA than men, 20% of women and 13% of men in Canada reported having OA in 2013 (Arthritis Community Research and Evaluation Unit (ACREU), 2013a). However, even when OA patients are excluded, women are at an increased risk of either cartilage loss or defects than men, indicating gender differences in overall cartilage health before the onset of disease (Hanna et al., 2009). Obese individuals are also up to four times more likely to develop knee OA than they are to develop type-2 diabetes or hypertension (Arthritis Research UK, 2011), although the relationship between obesity and risk of OA in other joints is not clear (Grazio & Balen, 2009). Furthermore, although obesity is a risk factor for knee OA, it does not seem to affect its progression (Niu et al., 2009). Despite these known factors affecting its aetiology, the prevalence of OA exceeds what would be expected due to increasing obesity rates or an aging population (Losina, Thornhill, Rome, Wright, & Katz, 2012).

While OA was once considered a necessary condition of aging, it is now understood that not everyone will develop arthritis. Some individuals live into their nineties with no evidence of OA. Over a period of many years, the effect of mechanical loading can have a "wear-and-tear" effect on cartilage, as aging progresses this erosion may or may not develop into OA. Considering the prevailing view for many years (that OA was an inevitable consequence of aging) not a lot of research had been done previously to elucidate the pathogenesis of OA and currently it is still not well understood. Some confusion remains around the initiating factors, mechanisms of progression and the general aetiology of the disease. Multiple tissues and processes have been implicated. As the development of OA is now considered to be a "whole-joint" phenomenon, where bone, cartilage, synovium and ligaments are considered together as a whole organ and all contribute to OA (Hugle, Geurts, Nuesch, Muller-Gerbl, & Valderrabano, 2012). Furthermore, most individuals only seek medical advice during the advanced stages of the disease when significant pain or reduced mobility affects daily activities. This creates some difficulty in determining the mechanisms contributing to OA. More research is necessary to determine these contributing factors and tissues, and how they interact in OA pathology.

1.1.1.5 Management

Treatment options for OA are limited. Initiation of a self-management plan including healthy eating, physical activity and joint-protection exercises is suggested upon diagnosis and throughout disease progression. Analgesic agents such as acetaminophen or non-steroidal anti-inflammatory drugs (NSAIDs) are used to relieve pain. Viscosupplementation can also be effective; it is a procedure where hyaluronic acid is injected into the synovial fluid of the affected joint (used mostly for knee OA). Features of OA include lower levels of hyaluronic acid in the synovial fluid making the fluid unable to protect the joints effectively (Legre-Boyer, 2015). The injection of hyaluron is

used to increase joint fluid volume and lubrication in mild to moderate knee OA (Legre-Boyer, 2015). Following other management strategies, a total joint replacement surgery is typically an eventual requirement for severe OA. However, this surgery is limited mainly to hip and knee replacements. OA in other joints such as the hands or spine cannot be treated effectively with surgery. Unfortunately, as of yet, there are no disease–modifying medications approved to treat OA, even though the incidence of OA is 400-fold higher than rheumatoid arthritis, another common form of arthritis (Sacks, Luo, & Helmick, 2010). Other over-the-counter medicines include oral chondroitin, although the effect is contradictory (J. A. Singh, Noorbaloochi, MacDonald, & Maxwell, 2015). The lack of treatment options for OA is largely due to the limited knowledge regarding its' aetiology, the complexity and variable phenotype of the disease.

1.1.1.6 Genetics

OA is a multifactorial, complex disease influenced by age, genetics and the environment. It has a strong genetic component (Blanco & Rego-Perez, 2014) with heredity ranging from 40% to 65% depending on the affected joint site (Kraus et al., 2007; Spector, Cicuttini, Baker, Loughlin, & Hart, 1996). Heritability is considered higher for hand and hip OA than it is for knee (Kraus et al., 2007; Spector et al., 1996). Familial aggregation studies have estimated the sibling recurrence risk at approximately 5 in the UK (Valdes & Spector, 2011). Genes implicated in OA include those regulating endochondral ossification, embryonic development and postnatal skeletal maintenance (Reynard & Loughlin, 2013). Endochondral ossification is tightly controlled by interactions with a number of hormones such thyroid hormones, and growth factors (*TGF-*β, *BMP*s, *GDF5*) parathyroid-hormone-related-protein (*PTHrP*), transcription factors, *RUNX2* and ECM components secreted by chondrocytes, such as matrix metalloproteinases (*MMP*s) (Mackie, Tatarczuch, & Mirams, 2011). Furthermore, human monogenic skeletal disorders often display mutations in *GDF5*, *RUNX2*, *PTHLH* and

TP63 (Reynard & Loughlin, 2013). These genes, involved in muscoskeletal processes such as joint development, chondrocyte and osteoblast differentiation, are also implicated in OA susceptibility.

To date, genome-wide association studies have identified 22 variants reaching genome-wide significance ($p < 5 \times 10^{-8}$) (Burdett et al., 2015) (http://www.ebi.ac.uk/gwas/search?query=osteoarthritis#association). SNPs associated with the growth differentiation factor 5 (*GDF5*) gene have shown strong, reproducible associations with OA in multiple studies (Miyamoto et al., 2007; Valdes et al., 2011). *GDF5*, also known as bone morphogenic protein 14, is a member of the transforming growth factor (*TGF-*β) superfamily, previously implicated in a multitude of other skeletal disorders in humans (Cornelis, Luyten, & Lories, 2011). Variants in *DVWA* (encoding protein with regions corresponding to the von Willebrand factor type A domain) have been associated with knee OA in Japanese and Chinese cohorts (Miyamoto et al., 2008). Human leucocyte antigen (*HLA*) class II/III has been associated with susceptibility to knee OA in some Japanese and European populations, supporting an immunologic mechanism implicated in the pathogenesis of OA (Nakajima et al., 2010). Furthermore, a SNP located in the *MCF2L* gene (MCF.2 cell line derived transforming sequence-like), has been implicated in skeletal and pain-related outcomes of OA (Day-Williams et al., 2011). Other OA-associated variants are linked to the genes *ASTN2* (astrotactin 2), *FILIP1/SENP6* (filamin A interacting protein 1/ sentrin specific peptidase 6), *KLHDC5/PTHLH* (kelch domain containing 5/ parathyroid hormone-like hormone), *CHST11* (carbohydrate sulfotransferase 11), *TP63* (tumour protein p63), *FTO* (fat mass and obesity associated), *NCOA3* and *DOT1L* (Panoutsopoulou & Zeggini, 2013). The *COL12A1* gene has also been associated with OA in European population (arcOGEN Consortium et al., 2012).

However, these variants and associated genes cannot explain the full susceptibility to the disease, accounting for less than 10% of the genetic factor (Tsezou, 2014).

Additionally, a study by Panoutsopoulou *et al.* found that the effect of the *FTO* variant on OA is due to its effect on body mass index, and no direct association between OA susceptibility and *FTO* was found (Panoutsopoulou et al., 2014). The presence of multiple low penetrance polymorphisms in the general population could explain this discrepancy. Other possible causes for missing heritability include rare variants of strong effect, gene-environment interactions and structural variations, such as copy number variants (CNV's), among others (Manolio et al., 2009a) (see part II, page 54 for more in-depth information on missing heritability). Alternatively, the unexplained heritable component could be due to epigenetic alterations.

Selection of a good candidate gene involves assessing the association between alleles of a gene that may be involved in the disease (a candidate gene) and the disease itself (Kwon & Goate, 2000). Candidate gene studies are generally better suited for detecting genes underlying complex or common diseases where the risk associated with any given candidate gene is small (Collins, Guyer, & Charkravarti, 1997). This study uses a similar approach. The underlying disease pathology is used as a basis for research on the matrix metalloproteinase 13 gene, which is involved in cartilage degradation. However, associations with different alleles are not tested through segregation. Instead, the mechanism of association between OA and *MMP-13* is evaluated through epigenetics.

1.1.1.7 Introduction to Matrix Metalloproteinases

Matrix metalloproteinases (*MMP*s) are endopeptidases that belong to the metzincin superfamily of proteases. They are initially synthesized in latent form and secreted as proenzymes that require extracellular activation. Their structures share common features including a pro–peptide domain, catalytic domain and C–terminal domain. The catalytic domains are specific to the substrate and depend on zinc ions as a co-factor. The collagenases, *MMP-1*, *MMP-8* and *MMP-13* are named because they specifically cleave a single locus in all three collagen chains at a point three-quarters of

the length from the N-terminus. Generally, while all *MMP*s act to degrade ECM proteins they can also process bioactive molecules such as cytokines and chemokines (Van Lint & Libert, 2007). In mammals, the only proteinases that degrade triple helical collagens at neutral pH are from the *MMP* family (Billinghurst et al., 1997). Normal cartilage ECM is in a perpetual state of synthesis and degradation, balancing a state of dynamic equilibrium. *MMP*s are important because they facilitate ECM turnover and breakdown in physiology and pathology (Murphy et al., 2002) and have been found to influence the progression of numerous inflammatory processes (Van Lint & Libert, 2007). Substrates of *MMP-13* (also called collagenase 3) include collagen I-IV, IX, X, XIV and gelatin. Specifically, the gene degrades the resistant triple-helical fibrillar type II collagen (Knauper, Lopez-Otin, Smith, Knight, & Murphy, 1996). In studying OA, type II collagen is particularly important because it gives tensile strength to cartilage and is the most abundant (of only two) cartilage ECM macromolecules. The other abundant protein, aggrecan, resists compression (Eyre, 2002). Therefore, *MMP-13*'s preference for cleaving type II collagen makes it of interest in studying OA pathology.

Murine models of injury-induced OA can be created using surgical destabilization of the medial meniscus (DMM) (Bateman et al., 2013; Glasson, Blanchet, & Morris, 2007) and these models show increased *MMP-13* mRNA expression in osteoarthritic joints when compared to sham-operated controls. However the altered expression does not persist over time (Bateman et al., 2013; Leahy et al., 2015; Loeser et al., 2013). In contrast, using near-infrared fluorescence, a murine model of knee OA showed enhanced *MMP* activity and the total emitted fluorescence intensity increased over the course of OA progression (Leahy et al., 2015).

In humans, aged cartilage has been shown to exhibit increased expression of *MMP-3* and *MMP-13* (Wu et al., 2002), and *MMP-2* and *MMP-13* proteins have been detected in synovial fluid (Ryu et al., 2012). Also increased in synovial fluid of OA patients is interleukin-1 (*IL-1*), a major cytokine involved in arthritic inflammation as well as cartilage destruction (Ahmad et al., 2007). *IL-1* not only represses healthy

chondrocyte ECM genes but it also stimulates *MMP* activity (Ahmad et al., 2007). However, synovial fluid alone does not accurately reflect protein activities such as *MMP*s when using it for OA comparisons due to the "whole-joint" model of OA disease (Poole, 2012) and the potential tissue-specific nature of proteins. Leahy *et al*. speculate that overall *MMP* activity may sensitively reflect metabolic changes occurring as OA develops, although the dynamics of the net *MMP* activity is not clear (Leahy et al., 2015). However, the increased levels of *MMP*s during aging could enhance cartilage matrix degeneration (Dejica et al., 2012). Of the *MMP*s, *MMP-13* appears to be one of the most actively involved in joint damage (Leahy et al., 2015). Rego-Perez *et al*. reported that people in Spain with a mitochondrial haplotype H have higher serum levels of *MMP-13* whereas haplotype J is associated with a lower risk of OA (Rego-Perez et al., 2011). In contrast, another study found *MMP* proteins and mRNA analysis did not indicate a change in *MMP* levels during the course of OA progression (Bateman et al., 2013; Loeser et al., 2013). To date, none of the 22 OA-associated variants identified through GWAS correspond to the *MMP-13* gene and while there is strong evidence to support a role for *MMP*s, specifically *MMP-13* in OA pathogenesis, the mechanism remains controversial.

1.1.2 Epigenetics

1.1.2.1 Introduction to Epigenetics

Epigenetics is defined as stable, mitotically heritable changes that alter gene expression in response to environmental cues (both external and internal), without affecting the underlying DNA sequence. These changes are "heritable", meaning they can persist following cell division and often between generations (Jablonka & Raz, 2009). In mammalian germ cells, which give rise to gametes, the epigenome is reset before reprogramming into a totipotent state (Feng, Jacobsen, & Reik, 2010). Thus, a person's epigenome is established before birth and changes continuously in response to environmental stimuli as he/she ages. Precise control of gene expression is necessary for

maintaining proper development, differentiation, growth and supporting homeostasis. Epigenetic mechanisms are necessary regulators of gene expression and establish stable, long-term expression patterns passed on through cell division (Gibney & Nolan, 2010). Epigenetic mechanisms include DNA methylation, posttranscriptional histone modification and noncoding RNAs (such as micro RNAs), among others. Processes such as development (Kiefer, 2007) and aging (Brunet & Berger, 2014) are now considered to be under significant epigenetic control, as well as almost every biological processes in between (Delcuve, Rastegar, & Davie, 2009). Not only are epigenetics involved in healthy growth and development, they are important mechanisms that respond to, and may even control, certain disease processes (Delcuve et al., 2009). There is strong evidence that epigenetics, specifically DNA methylation, plays a crucial role in many complex, multi-factorial conditions such as cancer and arthritis (Delcuve et al., 2009).

1.1.2.2 DNA Methylation

DNA methylation is a reversible, covalent modification of the 5'– carbon of a cytosine residue to form 5–methyl cytosine (5mC). The methyl group (CH_3) is transferred from a methyl donor S-adenosyl methionine (*SAM*) to the cytosine residue in cytosine-phosphate-guanine dinucleotides. In mammals, the methylation of cytosine is restricted to residues located 5' to a guanosine, known as a CpG site. Areas of the genome where CpG sites occur frequently are known as CpG islands (Vrtacnik, Marc, & Ostanek, 2014). These islands are often found in promoter regions of genes where CpG density tends to be high, and when methylation occurs in these areas it usually results in gene silencing (Blanco & Rego-Perez, 2014). On the other hand, methylation can facilitate protein binding when it occurs in gene bodies, leading to increased gene expression (Hellman & Chess, 2007). Thus, the process of methylation can increase or decrease gene expression depending on environmental cues such as aging or during disease pathology. Ageing and age-related diseases often exhibit genome-wide hypomethylation, but promoter-specific

hypermethylation, resulting in inappropriate silencing and a loss of gene expression regulation (Blanco & Rego-Perez, 2014; A. A. Johnson et al., 2012). The repressed state can be passed on to future cells due to the heritable nature of methylation during mitosis. A limitation to studying methylation is that only some of the CpGs present in the human genome are involved in gene transcription and therefore, it is necessary to experimentally confirm the role of individual CpG sites (Vrtacnik et al., 2014).

1.1.2.3 Chondrocytes

Cartilage is the firm, flexible connective tissue that protects the articulating surfaces of bones. It is susceptible to degeneration. Knee cartilage can thin naturally due to aging, particularly on the femur (Hudelmaier et al., 2001) and patellae (Ding, Cicuttini, Scott, Cooley, & Jones, 2005). Chondrocytes are the only cell type found in healthy cartilage and as such they are of particular interest in studying OA. Chondrocytes are essential to maintaining healthy cartilage homeostasis. They produce extracellular matrix proteins, integral for structure and biochemical support to the cartilage tissue. Normal adult articular chondrocytes do not typically undergo apoptosis or cell division (Aigner et al., 2001). They are vulnerable to the effects of aging, for example through telomere shortening, and aged chondrocytes collected from older adults are prone to cell death when exposed to reactive oxygen species (ROS), which increase naturally in cartilage as aging occurs (Carlo & Loeser, 2003; J. A. Martin & Buckwalter, 2003). Chondrocyte development is mainly regulated by anabolic growth factors (Fortier, Barker, Strauss, McCarrel, & Cole, 2011). For example, aged chondrocytes become less responsive to growth factors such as bone morphogenic protein-7 (*BMP-7*) (Chubinskaya et al., 2002) and transforming growth-factor β (*TGF-*β) (Davidson, Scharstuhl, Vitters, van der Kraan, & van den Berg, 2005). *TGF-*β also stimulates articular chondrocyte proteoglycan synthesis in young mice, but the ability is diminished in older ones (van Beuningen, van der Kraan, Arntz, & van den Berg, 1994). Providing evidence that the

aging process could alter the epigenetic landscape in cartilage, resulting in decreased synthesis of healthy maintenance proteins.

1.1.2.4 Epigenetics in Osteoarthritis

Several rheumatic diseases have been linked to aberrant epigenetic regulation causing either gene silencing or increased expression. Epigenetics and the aetiology of systemic lupus erythematosus have been well established (Jeffries & Sawalha, 2011) and epigenetic abnormalities are thought to play a key role in pathogenicity of rheumatoid arthritis (Bottini & Firestein, 2013; Liu et al., 2013). Epigenetic profiles are often disease specific, for example osteoporotic and osteoarthritic bone tissue samples display hypomethylation in members of the tumour necrosis (*TNF*) superfamily that are unique to the underlying pathology (Delgado-Calle et al., 2012) and chondrocytes from OA patients have distinct methylation profiles when compared to healthy controls (Rushton et al., 2014). Epigenomic landscapes can also be tissue or joint-specific. It has been noted that the majority of tissue-specific methylation correlating with transcriptional suppression of genes occurs in regions of low-density CpGs close to CpG islands such as in promoter regions (Barter, Bui, & Young, 2012). Rushton *et al*. found that epigenomic profiles of hip and knee OA segregate into clusters, and that a proportion of differentially methylated loci did not overlap between the two joints, indicating different genes or pathways involved in the disease (Rushton et al., 2014). Also, epigenetic profiles from hip OA alone can segregate into distinct clusters, indicating potential disease heterogeneity (Rushton et al., 2014). However, most genes were involved in anabolic/catabolic pathways of cartilage homeostasis such as the *TGF-* β family and other genes involved in the degradation of the extra cellular matrix (ECM) including matrix metalloproteinases (*MMP*s) (Rushton et al., 2014). There is even evidence to suggest a correlation between methylation of CpG sites and histologic severity of OA (Jeffries et al., 2014). In sum,

evidence suggests that OA not only has a distinct epigenetic profile but that it is also tissue and possibly even joint-specific.

Cartilage and bone homeostasis is an important factor in OA pathology and can also be influenced by epigenetics. Genome-wide methylation profiling reveals samples taken from bone tissue of osteoarthritic patients have differences in methylation of genes involved in cell-matrix interactions during skeletal development (Vrtacnik et al., 2014). DNA methylation has been shown to regulate multiple genes involved in bone remodelling and homeostasis (Vrtacnik et al., 2014). One essential method of maintaining bone homeostasis, that has also been shown to affect OA pathogenesis, is the mechanical stimulation of joints (Arnsdorf, Tummala, Castillo, Zhang, & Jacobs, 2010). The process increases expression of osteopontin (an extracellular structural protein in bone) by reducing DNA methylation in the promoter region (Arnsdorf et al., 2010). Thereby providing evidence that knee and hip articular cartilage can have distinct cellular phenotypes due to differing epigenomic influences and that gene expression can also be influenced by joint-specific mechanical signals, through epigenetic regulation.

OA has a distinct methylation profile from that found in other rheumatic diseases. Other epigenetic modifications are most likely distinct as well, such as histone modifications or microRNAs. Regardless, studying DNA methylation holds promise for elucidation of complex disease mechanisms with the potential to discover novel gene associations, biomarkers and develop medications. In fact, the field of oncology has already begun using DNA methylation in clinical applications. For example, as biomarkers that can potentially predict recurrence risk in lung cancer (Harada et al., 2015) and renal cell cancer (Luo et al., 2015). Furthermore, DNA methyltransferase (*DNMT*) inhibitors such as zebularine have been approved by the FDA to treat breast cancer (Pouliot, Labrie, Diorio, & Durocher, 2015). This makes the use of epigenetics for diagnostics or treatments of osteoarthritis a tangible goal.

1.1.2.5 *MMP-13* and Epigenetics

MMP-13's unique ability to cleave type II collagen, the most abundant protein in healthy cartilage makes it important for OA studies. The emergence of epigenetics has brought new methods for studying the relationship between genetics and environment in complex diseases. A study by da Silva and colleagues found that OA cartilage exhibited increased expression of *MMP*s including *MMP-13* and that it correlated with reduced DNA methylation (da Silva, Yamada, Clarke, & Roach, 2009). However, it is difficult to draw conclusions from this study based on the fact that cartilage from only one control was used and most samples were compared to secondary OA patients. An article by Cheung *et al*. also identified hypomethylated CpG sites in the promoter region of *MMP-13* in OA chondrocytes. However, no gene expression analysis or correlation was done (Cheung, Hashimoto, Yamada, & Roach, 2009). There is further evidence that a hypomethylated region within the promoter of *MMP-13* could lead to increased gene expression. Bui *et al*. found a CpG site located at –104 relative to the TSS within the promoter region of *MMP-13* to be hypomethylated in OA chondrocytes where they also found increased *MMP-13* expression (Bui et al., 2012). They also found that the transcriptional binding factor cAMP-response-element binding protein (*CREB*) interacts with the promoter at site –104. This interaction relies on the methylation status and affects binding affinity (Bui et al., 2012). *MMP-13*'s transcription may be regulated by *CREB*, indicating other factors involved in its regulation. Therefore, some studies suggest an association between reduced methylation of certain CpG sites within the promoter of *MMP-13* and increased gene expression (Tsezou, 2014). Whether or not there is a correlation between the gene expression and methylation status is still under debate.

Mechanisms of methylation are well known, such as the addition of a methyl group from *SAM* to a CpG dinucleotide. Just recently, a pathway of demethylation has been described. Ten eleven translocation protein (*TET*) has been identified as an enzyme that converts 5-methyl-cytosine to 5-hydroxymethyl-cytosine (5hmC), a de-methylation pathway (Taylor, Smeriglio, Dhulipala, Rath, & Bhutani, 2014). Taylor *et al*. also

identified an increase in *TET* expression in OA chondrocytes as well as an enrichment of 5hmC in the promoter region of OA-related genes *MMP-1*, *MMP-3* and *MMP-13*. Leading to an increase in the metalloproteinases expression in OA, supporting the hypothesis that up-regulation of *MMP*s in OA can occur via a de-methylation pathway (Taylor et al., 2014). In addition, chondrocytes treated with the DNA methyltransferase inhibitor Aza resulted in specific demethylation of sites within the *MMP-13* promoter, concomitant with increased gene expression (Bui et al., 2012). The effect of methylation on gene expression is not certain. Alternative theories of regulatory methods exist. Other possible mechanisms of *MMP-13* regulation include microRNAs such as miR-140, which has been shown in vitro to directly mediate *MMP-13* expression in articular chondrocytes (Liang et al., 2012). Additional studies have shown other microRNAs capable of regulating *MMP-13* expression in human OA chondrocytes as well (Akhtar et al., 2010). It is important to determine the mechanisms of regulation in genes associated with OA such as *MMP-13*, to better understand the pathogenesis of disease. Studying these mechanisms, including DNA methylation of genes gives us the ability to examine more closely how environmental cues alter gene expression contributing to an OA phenotype.

1.1.3 Summary of Current Hypothesis

While the mechanisms of development and progression of OA remain unclear, several theories exist and the development of OA is likely due to a combination of several factors. For example, mitochondrial dysfunction and oxidative stress (Blanco, Rego, & Ruiz-Romero, 2011) may play a role, as well as an abnormal increase in advance glycation reaction (or non-enzymatic glycosylation) end products (AGEs) (Huang et al., 2011). A delicate balance between catabolic and anabolic processes most likely exists in joints and perturbations to the balance can encourage the development or progression of OA (Henrotin, Kurz, & Aigner, 2005). In fact, an age driven decline in anabolic activity could tip the balance and result in up-regulation of catabolic genes that drive OA

susceptibility (van der Kraan, 2014). Several lines of evidence support this theory. With OA progression, the disproportionate increase in catabolic activity results in further destruction of the cartilage matrix and aberrant cartilage homeostasis (Li, Wei, Zhou, & Wei, 2013). Epigenetics are most likely involved in the phenotypic modulation that causes healthy articular chondrocytes in anabolic homeostasis to become "altered" destructive chondrocytes, over-expressing cartilage-degrading proteases, cytokines and other inflammatory mediators which can alter a healthy homeostasis towards catabolism and matrix degradation (Blanco & Rego-Perez, 2014; Roach et al., 2005). The catabolic activities occurring during cartilage degradation are regulated by pro-inflammatory cytokines (for example *IL-1*) and mediators such as the matrix metalloproteinases (*MMP*s) and disintegrin and metalloproteinase and thrombospondin motifs (*ADAMTS*) (Sun, 2010). These pro-inflammatory cytokines and matrix metalloproteinases can be up-regulated during times of joint stress, for example in the presence of increased ROS levels (Nelson & Melendez, 2004). ROS also activate *ERK/ MAP* kinase, suppressing the expression of type II collagen and chondrocyte aggrecan (Goodwin et al., 2010). The association between *ERK/MAP* kinase and *TGF-β* pathways influencing *MMP*s is well established in cancer research. In a cancer study, *TGF-β* was shown to activate the *ERK* pathway and affect expression of *MMP-9* leading to glioblastoma progression (W. Chen et al., 2015). Also, genistein, a candidate chemopreventative drug in clinical trials, has been shown to block prostate cancer cell invasion by preventing *TGF-β* from activating *MAP* kinase, a pathway that regulates matrix metalloproteinase activity and cell invasion (Xu & Bergan, 2006).

Van der Kraan and colleagues hypothesize that changes in the *TGF-β* signalling pathway, driven by age, can trigger chondrocytes to a destructive phenotype thereby promoting the degradation of cartilage and the surrounding matrix (van der Kraan, 2014). *MMP-13* can be up regulated by *TGF-β*, and has been observed to localize to deep tissue in arthritic cartilage (Moldovan, Pelletier, Hambor, Cloutier, & Martel-Pelletier, 1997).

IL-1 and *TGF-β* were both found to increase the number of chondrocytes producing *MMP-13* (Moldovan et al., 1997). *SMAD-3* is an intracellular mediator of *TGF-β* signalling and has been shown to regulate matrix mineralization and chondrocyte proliferation. *SMAD3* can also regulate cartilage homeostasis by inducing type II collagen and *MMP-13* expression (C. G. Chen, Thuillier, Chin, & Alliston, 2012). In articular cartilage, *TGF-β* signalling through a *SMAD-3* dependent mechanism maintains homeostasis by preventing improper expression of *RUNX2*-inducible genes such as *MMP-13* (Reynard & Loughlin, 2013). Interestingly, a *TGF-β1* gene polymorphism has also been shown to interact with obesity in influencing the risk of OA in large joints (Muthuri et al., 2013).

In summary, in an aged individual, joint stress could lead to increased reactive oxygen species in the joints that can up-regulate catabolic pathways such as *TGF-β* and *ERK/ MAP* kinase causing an up-regulation of matrix metalloproteinases such as *MMP-13* and concomitant suppression of healthy anabolic type II collagen expressed by chondrocytes, shifting homeostasis to a destructive phenotype.

MMP-13, a catabolic enzyme essential for cartilage homeostasis, is suspected to be up regulated by epigenetic means in OA cartilage. While previous studies have looked into this question, the mechanisms by which it occurs remains unclear.

I hypothesize that the promoter region of *MMP-13* is hypomethylated in OA cartilage when compared to healthy cartilage, and that this aberrant methylation up-regulates the *MMP-13* gene, thereby stimulating cartilage degradation.

1.1.4 Objectives

The overall objective of this study is to compare the role of *MMP-13* in OA patients versus healthy controls. This will indicate whether epigenetic phenomena, specifically gene methylation, regulate the expression of *MMP-13* and modulate the effect it has on the pathogenesis of OA.

Specific objectives:

1. Assess the level of *MMP-13* expression to determine if there is a difference between osteoarthritic cartilage and healthy cartilage.
2. Determine the methylation level at several CpG sites within the *MMP-13* promoter region.
3. Use statistical methods to determine if a significant correlation exists between methylation status and *MMP-13* mRNA levels.
4. Identify and adjust for potential confounders such as age, sex and body mass index (BMI).

1.2 CHAPTER 2. Materials and Methods

1.2.1 Study Participant Recruitment

Study participants are part of the ongoing Newfoundland Osteoarthritis Study (NFOAS). NFOAS is a hospital-based case-control study, aimed at identifying novel biochemical, genetic and epigenetic factors for OA (Zhai et al., 2013; Zhang et al., 2014; Zhang et al., 2015). Recruitment began in November 2011, with the goal of obtaining 1000 individuals with OA and 1000 healthy controls. Patients were recruited from individuals undergoing total knee or hip joint replacement due to primary OA at St. Clare's Mercy Hospital and Health Sciences Centre General Hospital in St. John's, Newfoundland Labrador, Canada between November 2011 and December 2013. Healthy controls with no evidence of OA undergoing emergency hemiarthroplasty due to hip fracture were recruited from the same hospitals. Knee or hip OA was confirmed by pathology reports on the removed cartilage and orthopaedic surgeon's assessment based on the American College of Rheumatology criteria (R. Altman et al., 1991). Consent was obtained along with a medical health questionnaire and Western Ontario and McMaster Universities Osteoarthritis index (WOMAC) questionnaire, used to assess pain, stiffness, and physical function in patients with hip and/or knee OA (see appendix B-E for consent form, health questionnaire and WOMAC).

Demographic information obtained by the health questionnaires included sex, age, height, weight, relevant medical history and medications, and was verified where necessary by research staff using hospital admission and medical records. BMI was also calculated by weight (kilograms) / height (metres^2) (for sample calculation see appendix A).

1.2.2 Ethical Considerations

Ethical approval was obtained from Health Research Ethics Authority (HREA) of Newfoundland and Labrador. All study participants were entered into a password-

protected database and given study ID, DNA ID and RNA ID numbers to obscure identity. All files, including consent forms, questionnaires and WOMAC assessments were kept in a locked file cabinet. Informed consent was obtained while study participants were in hospital for joint replacement surgery. Participants either filled out the consent forms and questionnaires while in the hospital with help from a research team member when necessary or took the forms home to be completed at their leisure and mailed back upon completion.

1.2.3 Cartilage Preparation

Up to 5 cartilage samples, weighing on average 200 mg per sample, were retained during knee or hip replacement surgery from patients and controls. Cartilage was procured from the tibial plateau during knee replacements or from the femoral head during hip replacement. Cartilage samples were flash frozen within thirty minutes of surgery and stored in liquid nitrogen (LN_2) at -140 °C.

1.2.4 RNA/DNA Extraction from Cartilage

RNA and DNA extraction was performed on the same cartilage sample to avoid sampling bias. RNA and DNA were extracted as follows, a frozen cartilage sample of approximately 200 mg was transferred to a homogenizing cylinder with 1 mL TRIzol reagent and 200 uL guanidine thiocyanate, then ground in a cryogenic freezer mill (Spex CertiPrep 6770; Spex, Metuchen, NJ, USA) with the following protocol; initial pre-cool of ten minutes followed by two rounds of two minutes of grinding followed by a ten minute cool-down between grinding cycles. After grinding, the sample was transferred to a 50 mL tube until thawed. 200 µL chloroform was added and mixed gently then transferred to a 2 mL tube and mixed vigorously for thirty seconds then centrifuged for 15 minutes at 11,500 *g* at – 4 °C. Following centrifugation, the mixture separates into a lower red organic phase, an interphase and a colourless, upper aqueous phase. RNA

remains exclusively in the upper aqueous phase while DNA remains in the lower and interphase. RNA was removed and placed in a new 2 mL tube then purified using RNeasy Lipid Tissue Mini Kit according to the Manufacturer's protocol from (74804, Qiagen, Venlo, Netherlands). DNA extraction protocol was adapted from TRIzol and Qiagen protocols as follows. After removal of upper RNA phase, DNA remaining in lower and interphase was treated with 0.3 mL's 100% ethanol (per 1 mL TRIzol reagent used for initial homogenization). Samples were mixed by vortex and stored at 15-30 °C for 3 minutes followed by centrifugation at 8,000 *g* for 5 minutes at 4 °C. The supernatant was discarded and an additional wash was completed using the same conditions. Following the second wash and removal of supernatant, 900 uL ATL buffer and 120 uL proteinase K was added and samples were mixed thoroughly by vortex then incubated in a 56 °C water bath overnight. The following morning, samples were centrifuged and the supernatant was pipetted into a new 2 mL tube. One volume of phenol:chloroform:isoamyl alcohol (25:24:1, Invitrogen, 15593049, Waltham, Massachusetts, USA) was added to the sample and shaken thoroughly for approximately 20 seconds. The sample was then centrifuged at 16,000 *g* for five minutes at room temperature (22 °C). The lower organic phase was carefully removed and discarded and the sample was centrifuged for another five minutes at 16,000 *g* at room temperature. The upper phase containing DNA was carefully transferred to a new 1.5 mL tube. One volume of chloroform:isoamyl alcohol (24:1) (Sigma-Aldrich, St. Louis, MO, USA) was added to the sample, capped and shaken thoroughly for 20 seconds. The sample was centrifuged at room temperature for five minutes at 16,000 *g*. The bottom organic phase was carefully removed and sample was centrifuged again at 16,000 *g* for 3 minutes. The sample separated into two layers with an upper aqueous phase that was carefully transferred to a new 2 mL tube. 3 volumes of 100 % ethanol were added and mixed by inversion. Samples were then placed in a – 80 °C freezer for at least 30 minutes to precipitate the DNA. Following precipitation, the sample was centrifuged at 20,000 *g* for 30 minutes at room temperature. The supernatant was then discarded, leaving a small DNA pellet. The pellet was rinsed with 70% ethanol and

centrifuged again for 15 minutes followed by removal of supernatant (discarded). Pellet was washed 3 more times following the same procedure. Following removal of supernatant after the last wash, DNA pellet was left to dry for ten minutes then dissolved in 50 uL Tris-EDTA (TE) buffer. DNA concentration, A260/280 and A260/230 ratio were measured on NanoDrop 1000 Spectrophotometer and software v 3.6 (ThermoFisher). DNA was further purified when necessary using DNeasy Blood & Tissue Handbook according to the manufacturer's protocol (69504, Qiagen, Venlo, Netherlands).

1.2.5 *MMP-13* Promoter Methylation Analysis

Methylation analysis was completed using the Sequenom MassARRAY EpiTYPER protocol. The Sequenom MassARRAY is used for quantitative analysis of DNA methylation by mass spectrometry, displaying methylated and non-methylated DNA as distinct signals. Amplicon primers were designed using Epidesigner website (http://www.epidesigner.com/). Two different sets of primers were designed due to complications that arose and to cover more CpG sites. Originally a sequence of 600 base pairs from the promoter region was utilized to generate a primer pair covering 4 CpG sites starting at 230 base pairs upstream and ending approximately 40 base pairs upstream of the first exon (see Table 1.2).

Table 1.2 1st Primer Pair Sequence: Target 224 – Reverse Primer

Primer
Left: aggaagagagTTTTTATAGGTTTGTAATGGTGAGTT
Right: cagtaatacgactcactatagggagaaggctCCCACAATATCCATAAATATACTAAAACC

Table 1.3 Description of CpG sites analyzed with primer pair # 1

CpG Name	CpG Unit	Position	Sequence	CpGs	Base Mass (kDA)
CpG_Unit_1	CpG_Unit_1	156	*taaac	1	1945.244
CpG_Unit_2	CpG_Unit_1	151	*c	1	653.417
CpG_Unit_3	CpG_Unit_2	131	*aaaaaaaaaaaaaaaaatc	1	6554.181
CpG_Unit_4	CpG_Unit_3	41	*tttattttac	1	3441.214

Position refers to number of base pairs from forward primer, not position from TSS. CpG sites 1 and 2 are located so close together that the signals overlap, therefore CpG sites 1 and 2 are described as unit 1 as they are indistinguishable using this method.

A second primer pair was developed from –550 bp to –200 bp upstream of the first exon of *MMP-13*. The second primer pair overlapped with one CpG site (located –218 bp upstream from the first exon) from the first primer pair (see Table 1.4).

Table 1.4 2nd Primer Pair Sequence: Target 350 – Forward Primer

Primer

Left: aggaagagagTAAGTTATTAAGTTTGGTTTTGGTTT
Right: cagtaatacgactcactatagggagaaggctAACATCTCTTATTTCAACAAAATCTC

Table 1.5 Description of CpG sites within primer pair #2

CpG Name	Position	Sequence	CpGs	Base Mass (kDA)
CpG_Unit_1	327	cac*at	1	1891.192
CpG_Unit_2	179	c*aat	1	1602.007
CpG_Unit_3	154	aaaaac*t	1	2589.636
CpG_Unit_4	57	aaac*acct	1	2838.797

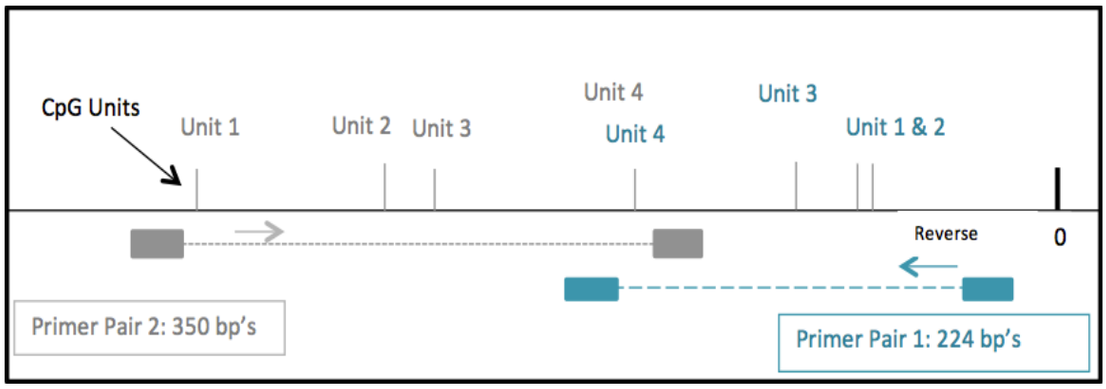


Figure 1.1 Illustration of overlap between primer pairs #1 and #2

Figure showing CpG sites analyzed and overlap of two primer pairs from experiment 1 (primer pair 1) and experiment 2 (primer pair 2). CpG unit 4 was included in both experiments due to overlap of primer pairs. CpG units 1 and 2 from experiment one were combined, the mass spectrometer could not measure them individually due to their proximity.

Following primer design a 96 well Plate (PCR MicroPlate PCR-96-FLT-C, Axygen, ThermoFisher) was arranged for bisulphite conversion based on the manufacturer's protocol for Sodium Bisulphite Conversion of Unmethylated Cytosines in DNA (from Low-Concentration DNA Solutions) (Epitect Bisulfite Kit, 59104, Qiagen, Venlo, Netherlands). Standards of 0% (Human Controls: Low Methylated DNA Control: 80-8062-HGUM5, EpigenDx, Hopkinton, MA, USA) 50% and 100% methylation (Human Controls: High Methylated DNA Control, 80-8061-HGHM5, EpigenDx, Hopkinton, MA, USA) were included as a reference. The 50% methylation standard was created by mixing a 1:1 volume of 0% methylation standard and 100% methylation standard.

Following conversion, PCR was performed to amplify converted DNA using Platinum Taq DNA Polymerase (600 Rx, Invitrogen, 10966-034, Waltham, Massachusetts, USA). PCR experiment was arranged according to the following parameters:

Table 1.6 PCR to amplify Bisulphite Converted DNA

Reagent	1X (µl)
H_2O	1.32
10X buffer	0.50
dNTPs Mix (10 mM)	0.04

$MgCl_2$ (50 mM)	0.1
Taq (Platinum)	0.04
Total	2

The Eppendorf Mastercycler Gradient (Hamburg, Germany) was utilized for the PCR reaction under the following conditions (Table 1.7).

Table 1.7 PCR conditions for amplification of bisulphite converted DNA

95 ˚C	2 min	
95 ˚C	30 sec	45 cycles
56 ˚C*	30 sec	
72 ˚C	1 min	
72 ˚C	5 min	
4 ˚C	∞	

Following the PCR reaction, 1 µL of several random samples was run on a 1% agarose gel (ultrapure agarose, Invitrogen, Waltham, MA, USA) in Tris-borate-EDTA (TBE) buffer at 100 V for approximately fifty minutes on BioRad Power Pac 3000 (Hercules, CA, USA) to verify presence of converted DNA.

Following successful bisulphite conversion of DNA, Shrimp alkaline phosphatase (SAP) treatment and Transcription/Cleavage reactions were performed using Epityper reagent kit (10239, Sequenom, San Diego, CA, USA) according to manufacturer's protocol and resin addition (SpectroCHIP II Resin Kit 2 x 96, 10188, Sequenome, San Diego, CA, USA), also according to manufacturer's protocol. SpectroCHIPs were spotted using Sequenom MassARRAY Nanodispenser Model RS1000 and then analyzed by Sequenom MassARRAY Analyzer 4 TYP: PHX-1 and EpiTYPER 1.2 software. The second experiment was completed using the same protocol but using different DNA samples and the second set of primers developed (see Table 1.4).

1.2.6 *MMP-13* Gene Expression Analysis

RNA was converted to complementary DNA (cDNA) using Maxima H Minus First Strand cDNA Synthesis Kit W/DsDnase (K1682, Thermo Fisher Scientific,

Waltham, MA, USA) following manufacturer's protocol. Presence of target sequence was verified by PCR using Platinum Taq DNA polymerase kit (Invitrogen, Waltham, MA, USA 18038042) and dNTPs (10297-018, Invitrogen, Waltham, MA, USA). PCR was carried out using Eppendorf Mastercycler Gradient (Hamburg, Germany) program using the following parameters; initialization phase of five minutes at 94 °C, then 28 repeating cycles of one minute at 94 °C for denaturation, one minute at 61 °C for annealing, and one minute at 72 °C for elongation followed by ten minutes at 72 °C for final elongation and a final hold at 4 °C (see Table 1.8).

Table 1.8 Conditions for PCR for verification of target sequences

94 ˚C	5 min	
94 ˚C	1 min	28 cycles
61 ˚C*	1 min	
72 ˚C	1 min	
72 ˚C	10 min	
4 ˚C	∞	

Presence of target product was verified by running eight µL of PCR products on a 1% agarose gel in Tris-borate-EDTA (TBE) buffer at 100 V for forty-five minutes using BioRad Power Pac 3000 (Hercules, CA, USA).

Primers for *MMP-13* were designed using Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Table 1.9).

Table 1.9 Primer Pair for *MMP-13* Gene Expression Analysis using rt-PCR

Primer Name	Sequence
MMP13-Q92L	AGCTGGACTCATTGTCGGGC
MMP13-Q92R	AGGTAGCGCTCTGCAAACTGG

Gene expression analysis by relative quantification was determined on ABI 7900HT Fast Real-Time PCR System (LifeTechnologies, Carlsbad, CA, USA) using *GAPDH* as a reference gene. A four point standard curve was established as a reference using cDNA dilutions from 0.25 µL, 0.12 µL, 0.06 µL, and 0.03 µL. Reaction was carried

out using 10 µL Power SYBR Green PCR Master Mix (4367659, ThermoFisher, Waltham, MA, USA), 0.06 µL cDNA, 0.4 µL each forward and reverse primers and water was added to a final volume of 20 µL. Standard mode and conditions were set to two minutes at 50 °C followed by initial denaturation phase for 10 minutes at 95 °C and forty-five repeating cycles of fifteen seconds at 95 °C and one minute at 60 °C for denaturation and annealing respectively.

Table 1.10 Settings for rt-PCR

50 ˚C	2 min	
95 ˚C	10 min	
95 ˚C*	15 sec	45 cycles
60 ˚C	1 min	

1.2.7 Statistics

Statistical analysis was performed using STATA/SE 11.2 (Stata Corp, College Station, Texas USA). Mean comparisons of gene expression and methylation levels between OA cases and controls were assessed using Student's t-test and the non-parametric Mann-Whitney rank-sum test with a significance level defined at alpha = 0.05. Spearman's rank correlation coefficient and significance was calculated to determine the relationship between methylation levels in the promoter region and gene expression.

1.3 CHAPTER 3: Results

1.3.1 Demographic Information

A total of 78 individuals were included in the analysis: 38 hip OA patients, 10 knee OA patients and 30 controls. These individuals were chosen randomly, the only criteria being that we had obtained multiple cartilage samples during surgery, the DNA and RNA extraction experiment was successful and a high concentration and volume was obtained. The mean age was 64.2 ±1.36 years for cases and 78.4 ±1.92 for controls. The age difference between our cases and control was significant at $p = 3.27 \times 10^{-10}$. The body mass index (BMI) was 32.19 ±0.99 kg/m^2 for cases and 23.83 ±0.85 kg/m^2 for controls. This was significantly different at $p = 2.11 \times 10^{-9}$ (see appendix for BMI calculations). A total of 53 females and 25 males were included. Of the cases 62.5% were female and 37.5% were male. From the control group 76.7% were female and 23.3% were male. The gender difference between cases and controls was not significant, p = 0.20. OA status was confirmed by a surgeon's assessment and pathological report following surgery. Controls were classified as true controls if no signs of arthritis were present in the joint at the time of surgery. An additional classification, controls with minor degenerative changes in the removed cartilage, was included following interpretation of pathological report. This classification was used to indicate individuals with some degeneration of their cartilage. This degeneration was determined to be normal age-related erosion and not due to (OA). Our analysis indicated that both gene expression and methylation levels demonstrated no significant differences between true controls and those with minor degenerative changes. Therefore, they were both included in the control category for the overall analysis. Gene expression analysis included 53 individuals: 25 hip OA, 7 knee OA and 21 controls. The initial methylation analysis was done using DNA from 6 hip OA patients, 4 knee OA patients and 5 controls. The second round of methylation analysis included DNA from 62 individuals: 30 hip OA, 6 knee OA, 26 controls. Of the 62 individuals, 2 controls and 1

hip OA patient had been previously included in the first methylation analysis, therefore only the measurement from the second run was used for the final analysis.

Table 1.11 Demographic information of cases and controls

Characteristics	Controls	Cases	p-value
Number of participants	30	48	-
Sex, *n* (%)			
Male	7 (23.3)	18 (37.5)	0.2
Female	23 (76.7)	30 (62.5)	
Age at surgery (years), mean ± SD	78.4 ± 1.92	64.2 ± 1.36	3.27×10^{-10}
BMI (kg/m^2), mean ± SD	23.83 ± 0.85	32.19 ± 0.99	2.11×10^{-9}

Age calculated in years at time of surgery ± standard deviation
BMI calculated in kg/m^2 at time of surgery ± standard deviation
Sex frequency calculated for cases and controls
P-value calculated by student's t-test at significance $p < 0.05$

Odds ratio calculations were also performed for obesity (OR 9.1, 95% CI: 2.7-30.18, p= 0.0003), hypertension (OR 0.8, 95% CI: 0.3-1.9, p= 0.6) and type 2 diabetes (OR 0.4, 95% CI: 0.1-1.0, p= 0.04).

1.3.2 Gene Expression

To establish the expression level of *MMP-13* we carried out rt-PCR on RNA samples extracted from cartilage from hip and knee OA cases and healthy controls. To establish our control group the average RQ value from (14) true controls was compared to those with minor age-related changes (7), the mean RQ ± standard error was 0.313 ±0.14 and 0.405 ±0.303 respectively (for sample calculation see Appendix A). The means were not significantly different at p = 0.8. Therefore, our two control categories were combined as one group. The difference between the mean RQ of knee and hip OA were also examined, at 0.903 ±0.28 and 1.75 ±0.44 respectively. This difference was also found to be not statistically significant at p =0.3. We therefore combined knee and hip OA patients into one group of "cases" in order to increase our sample size and study power. The average fold-change for cases was 1.56 ±0.36 and for controls was 0.34 ±0.13. Osteoarthritic cartilage was found to have greatly increased expression of *MMP-13* by

359%, this was a significant difference, p =0.01 based on Student's T-test (power = 1.00). Due to the non-parametric distribution of the gene expression data, the Mann-Whitney rank-sum calculation confirmed the significance at p = 0.006 (see Figure 1.3.1).

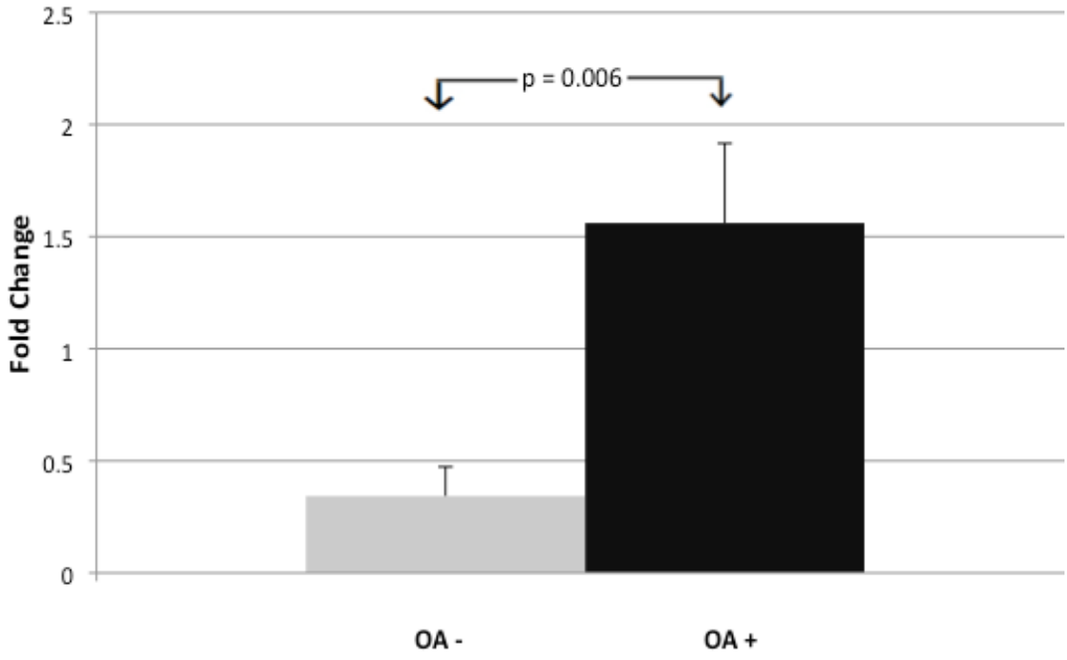


Figure 1.2 *MMP-13* expression results from rt-PCR of cartilage from OA patients (hip and knee) compared with healthy controls.

1.3.3 Methylation analysis of *MMP-13*

The level of methylation in the promoter region of seven CpG sites was investigated. The initial experiment included 6 hip OA cases, 4 knee OA cases and 5 controls. The Ensembl genome browser was used to retrieve the promoter sequence of *MMP-13*. Using the browser we identified an ideal sequence covering 600 base pairs upstream of the first exon in the promoter region. Using the identified sequence, a primer pair spanning 224 base pairs (the closest primer pair available to the first exon) was chosen for the experiment. The primer pair covered 4 CpG units, 2 of which were analyzed successfully. Two of the CpG units (designated 1 and 2) were located very close together at –103 and –108 upstream of the transcription start site (TSS). Due to their proximity the mass spectrometer could not distinguish between the two sites and the results reported are based on overlapping signals that cannot be interpreted separately.

The other sites were CpG site 3 located at –128 and CpG unit 4 at –218 bp. The mean β value of CpG units 1 and 2 in cases was 0.92 ±0.015 and in controls was 0.93 ±0.031. Median for patients and controls was 0.94. The difference between the cases and controls was not significant at p = 0.78. The mean of CpG unit 3 in cases was 0.35 ±0.11 and in controls was 0.38. The median for patients was 0.27 and 0.38 for controls. The standard error could not be calculated for controls, only one control for this CpG site reached an acceptable threshold for measuring β-value and subsequently a Student's t-test could not be applied. At CpG site 4 the mean for controls was 0.79 ±0.08 and for cases was 0.49 ±0.05, the median was 0.75 for controls and 0.43 for patients. This difference was significant at p = 0.005. CpG site 4 was therefore the only site found to be statistically significant in the first experiment. Therefore an additional experiment was performed to confirm this result and to measure several other CpG sites.

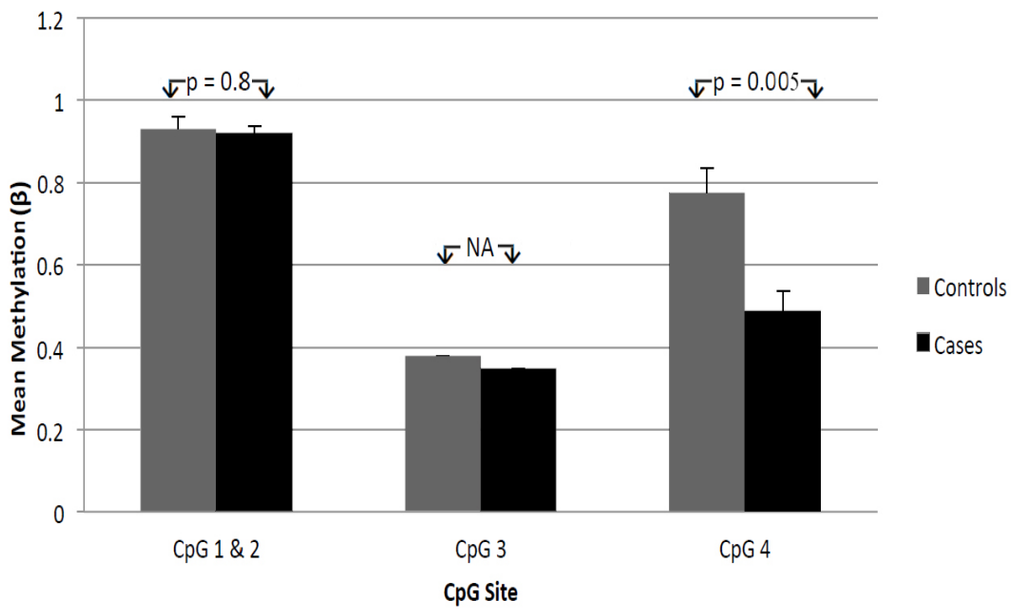


Figure 1.3 Mean β values from initial methylation experiment of CpG sites within the promoter region of *MMP-13* cases versus controls.

1.3.4 Methylation of *MMP-13* Experiment #2

For the subsequent experiment a new primer pair was designed. The development procedure was the same. Using Ensembl genome browser a primer pair was designed from a 600 base pair promoter region upstream of the first exon. The 350 base pair

primers chosen overlapped CpG unit 4 from the previous experiment and also covered 3 additional CpG sites not previously analyzed. CpG site 1 was located at –487 bp, CpG 2 at –337 bp, CpG 3 at –317 bp and CpG 4 at –218 bp. The average β value of CpG unit 1 in cases was 0.80 ±0.037 and controls was 0.81 ±0.032 at p = 0.9 (power = 0.18). CpG unit 2 cases had a mean of 0.65 ±0.064 and 0.71 ±0.083 in controls (p = 0.6) (power = 0.76). CpG unit 3 cases were 0.84 ±0.036 and controls 0.92 ±0.037 (p = 0.1) (power = 1.00). For CpG site 4 the mean β values for cases was 0.42 ±0.026 and controls 0.52 ±0.023. Of the sites tested only the fourth was found to be statistically different between OA cases and healthy controls at p = 0.008 based on Student's T-test. This CpG site is the same as CpG site 4 from the first experiment.

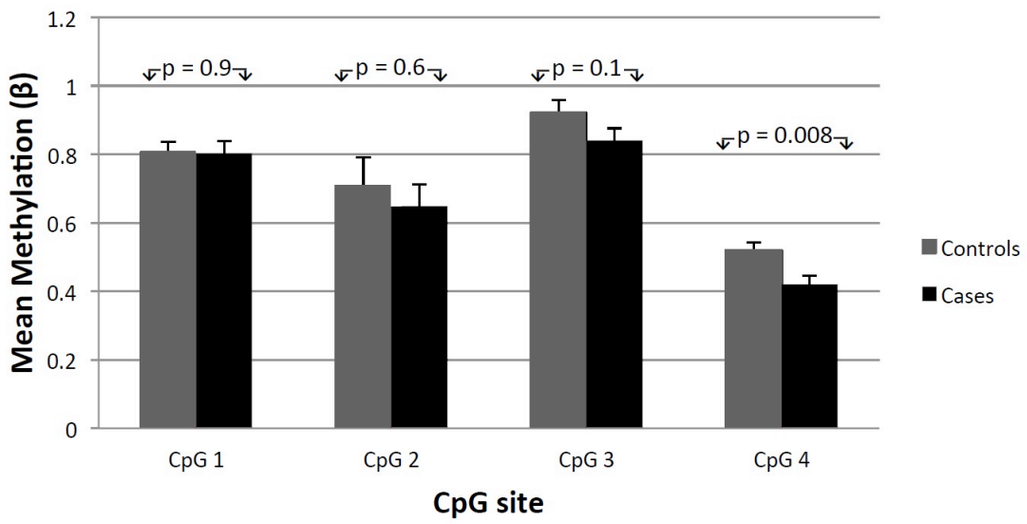


Figure 1.4 Mean β values from second methylation experiment of 4 CpG sites within the promoter region of *MMP-13* cases versus controls.

1.3.5 CpG site 4 combined data from experiment #1 and #2

Based on the analysis, there was a significant difference between mean β values from the same CpG site for both runs. In order to obtain a more robust data set we combined the methylation data from CpG site 4 from the two experiments. A total of 66 patients, 33 hip, 6 knee and 26 controls were included. Again, hip and knee OA patients were combined as cases, and controls with minor changes were included in the control category. The mean β values of the combined data for CpG site 4 for cases was 0.44

±0.023 and controls was 0.54 ±0.028. This difference was statistically significant based on Student's t-test at p = 0.006 (power = 1.00).

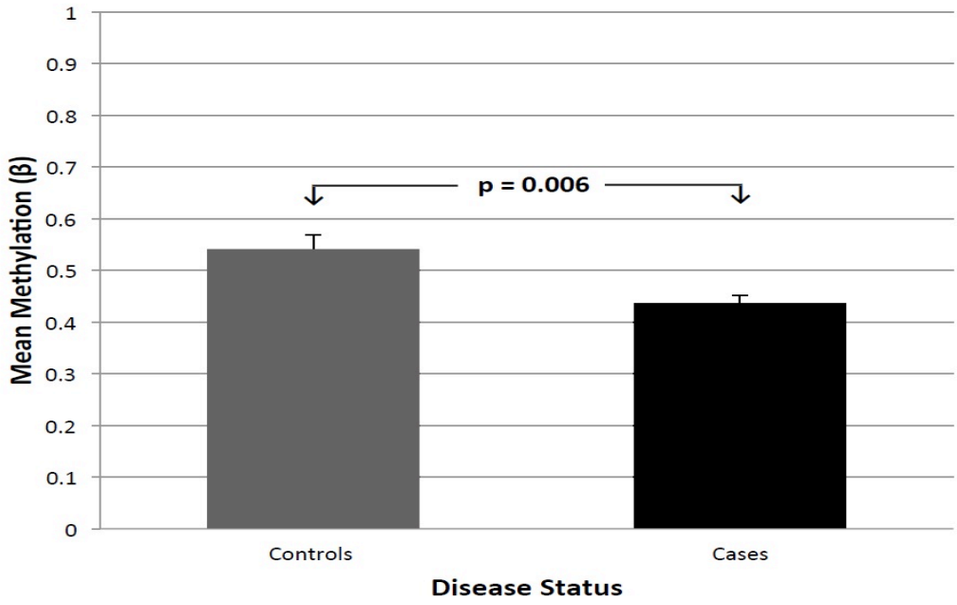


Figure 1.5 Combined mean β values from CpG site 4 from promoter region of *MMP-13* in cases versus controls.

1.3.6 Correlation between gene expression and methylation

To determine if any correlation exists between the expression of *MMP-13* and methylation in the promoter region we utilized Spearman's rank correlation analysis (see Table 1.12). Of the seven CpG sites examined, CpG site 2 from the second experiment exhibited a statistical significance. This CpG site is located at –337 bp from the TSS. However this site did not show a significant difference between cases and controls in the methylation analysis. Additionally, CpG 4 was not significantly associated with gene expression.

Table 1.12 Spearman's rank correlation for all CpG Sites

CpG site	Number of Observations	rho	p-value
Exp 1, CpG 1 & 2	8	0.48	0.23
Exp 1, CpG 3	4	0	1
Exp 2, CpG 1	34	0.13	0.47
Exp 2, CpG 2	30	- 0.39	0.035
Exp 2, CpG 3	41	- 0.30	0.06
Exp 1 & 2, CpG 4	48	- 0.24	0.1

Spearman's rho and statistical significance calculated for each CpG site to determine the correlation between gene expression and methylation levels.

1.3.7 Confounding Factors

Several potential confounding factors were examined based on our study population. OA prevalence is higher in women, older individuals and those who are overweight. Therefore, age, BMI and sex were examined as the primary confounding variables. Statistical analysis was performed using Spearman's rank correlation. Briefly, it measures the strength of the association between two ranked variables that display a monotonic relationship. Scores are ranked and given values between +1 to -1, where +1 indicates a perfect positive relationship, 0 indicates no association and -1 indicates a perfect negative relationship. The closer the r (or rho) value is to zero, the weaker the association between ranks. Based on an α value of 0.05 none of these potential confounders significantly influenced the results of this study (Table 1.13, Table 1.14). Therefore, the methylation results were not adjusted based on any of these confounders.

Table 1.13 Spearman's rank correlation analysis of confounding factors for methylation analysis of CpG site 4

Confounding Factors	Cases		Controls	
	Rho	**p-value**	**Rho**	**p-value**
Age	0.09	0.57	0.29	0.13
Sex	0.12	0.45	0.11	0.57
BMI	0.15	0.32	0.25	0.20

Table 1.14 Spearman's rank correlation analysis of confounding factors for gene expression

Confounding Factors	Cases		Controls	
	Rho	**p-value**	**Rho**	**p-value**
Age	0.23	0.20	0.16	0.50
Sex	0.15	0.40	0.13	0.58
BMI	0.17	0.35	0.03	0.90

1.4 CHAPTER 4: Discussion

1.4.1 Gene Expression of *MMP-13*

Matrix metalloproteinases 1, 8 and 13 are the only mammalian enzymes capable of initiating cleavage of type II collagen, a major component of cartilage, at neutral pH (Billinghurst et al., 1997). This makes the *MMP* family essential for normal development and turnover of the ECM and has implications for cartilage pathologies (Murphy et al., 2002). Specifically, *MMP-13* is a downstream target gene in the *TGF-*β signalling pathway, suspected to be involved in OA pathogenesis (van der Kraan, 2014). In murine models of OA, *MMP-13* is over expressed in OA joints when compared to sham-operated controls (Bateman et al., 2013). Also, previous studies on humans have demonstrated an over-expression of *MMP-13* in aged (Wu et al., 2002) and osteoarthritic cartilage (Sakata et al., 2014), chondrocytes (Bui et al., 2012) and synovial fluid (Ryu et al., 2012) from OA patients.

To quantify the level of *MMP-13* expression in OA cartilage when compared to controls we performed rt-PCR. We found the gene expression levels between knees and hips to be comparable, with minor variations that were not significant. Although, some studies indicate differences in the genes or pathways involved in OA between joints (Rushton et al., 2014). *MMP*s have previously been associated with OA progression in both joints (Gonzalez, 2013; Rushton et al., 2014). The present study demonstrates that *MMP-13* is likely involved in OA pathogenesis in both joints. However, the study has only a small number of knee OA patients and verification with a more robust sample size is necessary. Gene expression was also similar in healthy controls and controls with minor degenerative cartilaginous changes. Those with minor changes had a slightly higher expression but it was not significant. These findings suggest that age-related degenerative changes in cartilage do not necessarily exhibit an over-expression of *MMP-13* such as that seen in osteoarthritic cartilage, and that the over-expression of the gene is specific to the pathology. The overall expression from cases was 359% fold higher than

controls. This demonstrates a significant increase in *MMP-13* expression in OA patients. This finding was verified using the Mann-Whitney rank-sum statistical analysis due to the nonparametric distribution of the gene expression data. The expression was not dependent on either BMI or sex, indicating that *MMP-13* expression is equally up regulated in men and women irrespective of their BMI. While there are genes and pathways associated with OA specific to women (Lian et al., 2007) and obesity, such as adipokines (Poonpet & Honsawek, 2014), these do not appear to be a factor in *MMP-13*'s association with OA. This study supports previous observations that *MMP-13* mRNA is present at higher levels in cartilage tissue from OA patients when compared to healthy controls. The high level of *MMP-13* indicates that OA patients are undergoing more significant cartilage degeneration and ECM turnover, supporting a potential role for *MMP-13* in OA pathogenesis. However, the nature of this relationship (whether the up-regulation of *MMP-13* is a cause or consequence of OA) is unclear. The catabolic erosion of cartilage is a dynamic process and *MMP-13* plays a key role in its degeneration. This experiment provided evidence of altered *MMP-13* expression in RNA from osteoarthritic cartilage as a necessary step to evaluate the effect DNA methylation could have on the expression of this gene.

1.4.2 Methylation Analysis

Methylation is the addition of a methyl group to 5' cytosine-phosphate-guanidine, also called a CpG site. It is a method of epigenetic modification, whereby genes are able to respond to environmental stimuli. Previous studies have examined areas of hypomethylation in *MMP-13* that may correspond with gene expression (Bui et al., 2012). By measuring the methylation of seven CpG sites within the promoter region of *MMP-13* we are able to gain a better understanding of its regulation and role in OA. In our study, while there were minor changes between methylation levels in knee and hip articular cartilage in *MMP-13*, they were not significantly different. Although not a robust knee

sample size, at least for *MMP-13* the differences in methylation profiles at these seven CpG sites was not significantly different between hip and knee OA. However, a study by den Hollander and colleagues found that epigenetic profiles of knees and hips are distinct (den Hollander et al., 2014). They measured the overall epigenetic landscape, and not individual CpG sites. It is possible that the disease pathology has made the methylation in the *MMP-13* promoter comparable in the two joints. The similar methylation level in osteoarthritic cartilage from knees and hips allowed us to combine these cases into OA patients and perform a larger comparison with the control group.

We performed two experiments, covering a total of seven CpG sites located in the promoter region of *MMP-13*. Of the seven CpG sites analyzed, only one was found to have a significant difference between cases and controls. The site located at –218 from the TSS was significantly demethylated in cartilage from OA patients (both hip and knee) as compared to controls. This CpG site was included in the experiments from both sets of primer pairs, allowing a more robust statistical evaluation of this site. This result supports previous experiments that had found hypomethylated CpG sites within the *MMP-13* promoter (Bui et al., 2012; Taylor et al., 2014), and a study by Rushton *et al*. that demonstrated that chondrocytes from cartilage of healthy controls and OA patients had different methylation profiles (Rushton et al., 2014). After examining potential confounders we determined that age, sex and BMI were not significantly associated with methylation at CpG site 4 at $p < 0.05$.

Many genes, particularly those associated with transcription regulation, development and processes of anatomical structure, tend to become hypermethylated in the promoter region with advancing age (Johansson, Enroth, & Gyllensten, 2013). In osteoarthritic cartilage this process can become disrupted, and hypomethylation of some CpG sites occurs instead, potentially contributing to an up-regulation of *MMP-13*. One theory of the mechanism of OA is that the balance of catabolic and anabolic processes becomes disrupted, causing an up-regulation of some catabolic genes while anabolic genes are suppressed (Blanco & Rego-Perez, 2014; Roach et al., 2005). Our study

supports the up-regulation of the catabolic *MMP-13* gene. Pathways known to be involved in the regulation of *MMP*s are *TGF-* β and *ERK/MAP* kinase signalling pathways, which also suppress the healthy, anabolic expression of type II collagen and chondrocyte aggrecan, potentially leading to an OA phenotype (C. G. Chen et al., 2012; W. Chen et al., 2015; Goodwin et al., 2010).

1.4.3 Gene expression and methylation

When methylation occurs in the promoter regions of genes it typically represses the transcription of the gene thereby silencing it. Most tissue-specific methylation occurs in areas of low CpG density, called "shores", close to CpG islands, and correlates with transcriptional repression of the gene (Barter et al., 2012). In order to elucidate the role of methylation in regulating *MMP-13* expression, we used statistical methods to evaluate the correlation between gene expression and methylation at CpG sites in the promoter region of *MMP-13*. Using Spearman's rank correlation (due to the nonparametric of the gene expression data) we tested the correlation between the two and found that none of our groups (controls, hip or knees) demonstrated a correlation between methylation at CpG site 4 in the promoter region and gene expression. Specifically, while the methylation analysis indicated hypomethylation of the CpG site at –218 bp upstream of the first exon, this site did not correlate with the expression of *MMP-13*. The hypomethylation of this CpG site cannot, by itself, explain the over-expression of *MMP-13*. Interestingly, expression did correlate to a different site, CpG 2 from the second experiment but the methylation difference between cases and controls was not significant at $p > 0.05$. It seems methylation of these CpG sites alone cannot explain *MMP-13*'s over-expression at the level of 359% measured in this study. Several possible explanations exist. Either we did not test the site(s) responsible for the change observed in gene expression in osteoarthritic cartilage, or perhaps gene expression corresponds to epigenetic regulation from multiple sites. Alternatively, control of gene expression could be achieved through

methods other than DNA methylation. Other possible epigenetic mechanisms of regulation include miRNAs or histone modification. There is evidence to suggest miRNAs could play in a role in *MMP-13*'s expression (Akhtar et al., 2010; Liang et al., 2012). Also, regulation of genes could occur through multiple epigenetic mechanisms in conjunction.

1.4.4 Study Strengths

The strengths of this study include the use of human OA cartilage. Due to the resilient nature of cartilage, it can be difficult to manipulate or utilize for experiments. As such, many studies use murine models or cultured OA chondrocytes. The use of cartilage taken directly from osteoarthritic joints can more accurately reflect the environment within the joint and the genes involved in a way that tissue culture cannot. Furthermore, epigenetic landscapes are different depending on the tissue used and could be altered in a tissue culture in unknown ways. Therefore, it was advantageous to use one tissue type, taken directly from a human joint for the analysis.

Another strength is the use of DNA and RNA extracted from the same cartilage sample. As far as I can tell this is the first study to examine *MMP-13* using DNA and RNA from the same cartilage sample. There are indications that even within the same joint, cartilage taken from different sections can reflect differing gene expression or methylation patterns (Aicher & Rolauffs, 2014; Moldovan et al., 1997). Therefore, using DNA and RNA extracted from the same tissue specimen is especially useful for correlation analysis and further insight into the genes and epigenetic mechanisms involved in pathogenesis.

While the role of *MMP-13* in OA has been previously examined in other populations, this is the first study in the Newfoundland population. The unique population genetics and high prevalence of OA in NL make it an advantageous study population to use.

1.4.5 Study Limitations

While cartilage is often used to study the joint environment, it does not necessarily reflect what is occurring in all tissues involved in OA. Bone, cartilage, synovium and ligaments all contribute to disease progression as a "whole-joint" (Hugle et al., 2012). However, it is currently unknown which tissues commence OA progression or the extent to which they all contribute (Cox, van Donkelaar, van Rietbergen, Emans, & Ito, 2012; Pitsillides & Beier, 2011). However, analysis of OA progression using a whole joint from human patients is difficult with current methods. Therefore, cartilage tissue, considered to be actively involved in the disease process, is frequently used for OA studies.

Furthermore, we used mRNA and DNA to study gene expression and methylation. However, protein levels were not examined in this analysis. Studying DNA and mRNA expression with corresponding protein levels would have improved our analysis and knowledge of the regulation of *MMP-13*.

The number of cases versus controls utilized was not equal, this may bias the results as the methylation levels in cases may have been more accurately calculated. As well, a more robust sample size could have been used. Although most CpG sites were sufficiently powered, CpG unit 1 in this analysis was underpowered.

A complication of studying methylation profiles is that they do not always correlate directly to gene expression (den Hollander et al., 2014). Epigenome-wide experiments can inform us as to what areas may correspond to expression (den Hollander *et al*., 2014), and verification by studying the effects of single CpG sites is important (Vrtacnik et al., 2014). Both methods can provide novel information about genes or pathways and their regulation. This study measured seven CpG sites within the promoter region, up to 600 base pairs upstream from the TSS, it is possible that studying methylation throughout the gene or from related genes and pathways might provide a more complete picture of *MMP-13*'s involvement in OA.

1.5 CHAPTER 5: Conclusions

1.5.1 Significance of this Study

Osteoarthritis is a leading cause of pain and disability worldwide. As the most common form of arthritis, it currently affects approximately one-in-five Canadians (Arthritis Alliance of Canada, 2011). Despite this high prevalence and increasing incidence, the aetiology and pathogenesis are poorly understood. Also, while over-the-counter supplements exist, their effects are controversial, and as of yet no disease modifying osteoarthritis medications have been approved by the FDA. Furthermore, no biomarkers or methods for early detection are available. Due to the significant socioeconomic burden and increasing prevalence of OA, in excess of what would be expected from the obesity epidemic and an aging population, research into the genetics and epigenetics of OA is necessary. Enhancing our knowledge and subsequently developing therapies or diagnostics requires further research into genes such as *MMP-13*, and their regulation.

1.5.2 Future Directions

Next steps could include testing other genes suspected to play a catabolic role in ECM turnover and cartilage degeneration in OA, such as other *MMP*s, for a relationship between methylation and gene expression.

Further investigation of CpG site 2, located – 337 bp from TSS is required to verify its correlation with *MMP-13* gene expression. Although the methylation level at this site was not significantly different between cases and controls, it was lower in cases. Indicating this site could be involved in regulating gene expression. Validation by testing with a larger sample pool could be useful. As well, the age of our controls was much higher than cases, the study could benefit from a matching case control study.

Using a grading scale such as the K-L scale to measure patient and control samples and comparing them to gene expression and methylation may yield a pattern

reflective of histological grading. There is evidence from previous studies to suggest there may be correlation between methylation of CpG sites and histologic severity (Jeffries et al., 2014). It would be interesting to experimentally verify if this correlation occurs with gene expression or methylation of *MMP-13*. Furthermore, this experiment could shed light onto the cause/consequence relationship between *MMP-13* and OA.

2 PART II: BURDEN OF RISK ALLELES FOR COMPLEX TRAITS IN THE NEWFOUNDLAND AND LABRADOR POPULATION

2.1 CHAPTER 1. Introduction

2.1.1 Genetics in Founder Populations

A small proportion of individuals from one population migrating to a new location and establishing a colony usually experience a loss of genetic diversity, this is known as a "founder effect". A random sample of alleles is removed from the original population with the migrants and ends up in the new population. As a result of this effect, the new population often has a distinct genetic profile, with less genetic variation than the original population. Allele frequencies at the expansion front will have higher variance, and alleles are more likely to become fixed or lost (Peter & Slatkin, 2015). The founder population may also experience a population bottleneck, increased sensitivity to genetic drift, low genetic variation and an increase of inbreeding. In human populations, numerous factors, most notably geographic isolation, can result in a founder effect. Other factors that can influence isolation are social status and religion, among others.

Consequences of a founder effect can include higher prevalences of certain genetic conditions and lower prevalences of others due to the loss of genetic variation. There is an ongoing debate as to whether founder populations truly exhibit an excess burden of disease due to deleterious variant alleles. Lohmueller *et al*. suggested that European populations have more deleterious variants as compared to African American individuals (Lohmueller et al., 2008). Utilizing a computer simulation they found bottlenecked populations (European Americans in this case) were found to have about 26% more homozygous damaging genotypes and an excess of non-synonymous SNPs (Lohmueller et al., 2008). The findings support the "Out-of-Africa" bottleneck theory, suggesting that bottlenecked populations (European populations in this case) most likely accumulate more deleterious variants following the event (Lohmueller et al., 2008). In contrast, a recent study suggests that Europeans and West Africans have comparable burdens of deleterious alleles (Simons, Turchin, Pritchard, & Sella, 2014). While

evidence suggests bottleneck populations could harbour more deleterious variants, this theory is disputed.

Many current examples of founder population have been identified. Perhaps the best-known example of genetic isolation driving a high incidence of genetic disorders is in the Ashkenazi Jew population, which has been extensively studied. Ashkenazi Jews are known to have a higher prevalence of Tay-Sachs and cancer susceptibility (particularly *BRCA* 1 and 2 mutations) (Carmeli, 2004). Approximately 1 in 4 people of Ashkenazi Jew ethnicity is a carrier of either: Bloom Syndrome, Canavan disease, Cystic Fibrosis, Familial Dysautonomia, Fancomi Anemia, Gaucher disease, mucolipidosis IV, Niemann-Pick disease, Tay-Sachs or Torsion dystonia (National Cancer Institute (NCI), 2014). The Finnish population is also often used as a classic example of a founder effect with higher rates of recessive Mendelian conditions such as megaloblastic anaemic 1 (Aminoff et al., 1999; Savukoski et al., 1998). These high rates of disease are due to higher incidences of specific deleterious mutations within the population. At least 35 disease genes have been identified as enriched in the Finnish population (Peltonen, Jalanko, & Varilo, 1999). A cluster of families with progressive epilepsy with mental retardation (*EPMR*) (Tahvanainen et al., 1994) live on the Eastern border and a similar cluster of the variant form of late infantile neuronal ceroid lipofuscinosis (*vLINCL*) cases on the Western coastline have been described (Peltonen et al., 1999). On the other hand, some disease alleles have become almost non-existent in Finnish populations, for example phenylketonuria, galactosemia, maple syrup disease and cystic fibrosis (Peltonen et al., 1999).

Another example, a small community in the Netherlands with common ancestry, has been found with several genetic disorders that are more common than in surrounding areas. Due to social and religious separation, the community is isolated and has been described as genetically homogeneous. Four rare disorders, pontocerebellar hypoplasia type 2 (*PCH2*), fetal akinesia deformation sequence (*FADS*), rhizomelic chondrodysplasia punctata type 1 (*RCDP1*), and osteogenesis imperfecta (*OI*) type IIB/

III have recently been found in multiple patients from this village, driven to higher prevalence by a founder effect (Mathijssen et al., 2015). Marchi *et al*. recently identified the first proof of a founder effect in Northern Europe and were able to narrow down the timing of the occurrence to an event that had taken place nine generations previously (Marchi et al., 2014). They utilized microsatellite markers of β –globin to identify variant carriers in North England sharing a common haplotype for β –Thalassemia (Marchi et al., 2014).

These examples of high disease prevalence in founder populations are not observed in out-bred populations, where these conditions have a lower incidence and a wider array of mutated regions, indicating no single common ancestor (Marchi et al., 2014). In isolated populations the prevalence of genetic disorders can be relatively high. The high prevalences are likely driven by one or a few specific founder mutations in a gene that has spread throughout the population. Many successful studies have discovered novel rare variants that have become enriched in such populations, making founder populations a good resource for studying genetic conditions. These elevated disease prevalences have been found to mostly be consequences of a historic founder effect, population bottlenecks, genetic drift or an increase in inbreeding.

2.1.2 Settlement History of Newfoundland & Labrador

Newfoundland's unique settlement history and population structure make it a valuable location for studying population genetics. Of the approximate 520,000 current Newfoundland residents, approximately 80-90% can trace their ancestry to a group of around 20,000 – 30,000 settlers from South West England and South East Ireland in the 1700 and 1800's (Mannion, 1977). Although a small proportion of French and Basque descendants remain as well (Mannion, 1977). Early native groups, the Beothuk Native Americans, Maritime Archaic Indians, Paleo Eskimos, as well as Norse settlers inhabited Newfoundland in the past but no permanent communities remain (Rowe, 1980), though

there are considerable numbers of Mi'kMaq residing in NL today. After John Cabot's expedition to Newfoundland in 1497, overwintering establishments were constructed to support the fishing trade. These settlements were seasonal due to the harsh winters but are recorded as early as 1527 (Pope, Carr, Smith, & Marshall, 2011). At that time, Newfoundland had an abundance of natural resources and colonization was strictly controlled by England. However, evidence of permanent settlements began in 1621 from English immigrants and by 1675, Irish Catholics immigrants were documented in St. John's (Pope et al., 2011). At that time, English Protestants inhabited over 30 settlements along the Avalon Peninsula of Eastern Newfoundland. The settlements were primarily founded by less than five families originally from localized regions of southwest England (Mannion, 1977). By 1753 all major communities along the Avalon Peninsula also had large Irish communities (O'Neill, O'Neill, & O'Neill, 2003; Rowe, 1980). As more people immigrated, settlements and out ports spread further inland along waterways and bays, supported by the healthy fishing trade. Throughout the 17^{th}-20^{th} centuries, there was limited contact between settlements due to religious, social and geographic barriers (Mannion, 1977). Two main religions were present in Newfoundland at the time, the Anglicans and Catholics. The animosity between them made relationships between individuals from different religious communities rare. Historically, out-ports were small and isolated and most offspring remained near the original settlements (Bear et al., 1987). Of the outport populations, only 1 – 8% of the population were not originally from the area, and the majority of births were to parents from the same community (Bear et al., 1987). As a result, Newfoundland has had large family sizes, small settlement populations and minimal dispersion between out-ports, resulting in a high degree of genetic homogeneity and an abundance of several monogenic disorders (Crawford et al., 1995). NL has been utilized as a resource for identifying genes associated with monogenic disorders and is now of interest for studying genes implicated in common and complex diseases (Rahman et al., 2003).

2.1.2.1 Genetic conditions in Newfoundland

Newfoundland's unique settlement history has driven research into its genetic architecture and associated medical conditions. At first, this research in Newfoundland concentrated on rare medical conditions and later on genetic variation (L. J. Martin et al., 2000). For example, a study by Crawford and colleagues in 1995 examined red cell antigens and calculated gene frequencies and describe Burgeo, South Ramea and Fogo Island as being populations that had differing genetic structures (Crawford et al., 1995) and had ultimately experienced a loss of heterogeneity with respect to the source population (Pope et al., 2011). In Newfoundland, a number of diseases are considered more prevalent when compared to out-bred populations, type 1 diabetes (40/100,000 compared to the general population at 7-15/100,000), Lynch syndrome (Stuckless et al., 2007), hereditary diffuse gastric cancer syndrome (Kaurah et al., 2007), psoriasis and some cardiomyopathies, are all measured at higher rates in the Newfoundland population (Kosseim et al., 2013). Evidence for other diseases enriched in the Newfoundland population due to the founder effect include several autosomal dominant disorders such as arrhythmogenic right ventricular dysplasis type 5 (Merner et al., 2008) and familial forms of pulmonary fibrosis (Fernandez et al., 2012). Xie *et al*. describe a very high prevalence of mild haemophilia type A at a rate of 44 in 3,300 males in isolated populations in rural Newfoundland. This is an extremely high rate as compared to out-bred populations (Xie, Zheng, Leggo, Scully, & Lillicrap, 2002). The same mutation, a missense mutation (Val2016Ala) had previously been described in five other patients, all from areas around London UK. The mutation most likely arose independently in the two populations and expanded in Newfoundland through a founder effect (Xie et al., 2002). Further evidence, Doucette *et al*. describe a rare, severe retinal disorder called achromatopsia in Newfoundlanders that had previously been described at a high rate (4-10%) in another founder population from the Pacific Islands. They determined that the large family size and isolation of communities along the coast of Newfoundland had resulted in higher than expected inbreeding coefficient and a high rate of disease. Furthermore, unrelated patients

shared identical alleles suggesting a likely common ancestor. By reconstructing the haplotype they determined that 80% of achromatopsia alleles identified were the result of a founder effect in the Newfoundland population (Doucette et al., 2013). These examples of rare genetic conditions more prevalent in Newfoundland have been identified to be a result of the founder effect during the settling of NL. More recently, a study to measure the loss of genetic variation due to a founder effect, drift and inbreeding in Newfoundland, was undertaken by Pope and colleagues. They studied microsatellite loci and complete mitochondrial DNA (mtDNA) genomes from 27 individuals of known matrilineal ancestry (Pope et al., 2011). They found a high degree of microsatellite diversity (H_E = 0.763) and a significant degree of inbreeding (F_{IS} = 0.0174) (Pope et al., 2011). Thus, the observed increase in disease prevalence in NL may also be highly influenced by inbreeding and large family sizes.

2.1.3 Genome-Wide Association Studies

Genome-wide association studies (GWAS) have become increasingly useful for identifying allele variants associated with given traits. These traits can include eye colour or hair colour, or they could be associated with an increased risk for disease such as cancer or arthritis. The introduction of population-scale whole-genome data sets has enhanced our capabilities to identify or interpret genetic variants in the human population. Genome-wide association studies (GWAS) examine the genome for common genetic polymorphisms to determine potential associations with given phenotypic characteristic. The principle idea behind GWAS is that a mutation in a common ancestor entered the population and spread. This process has resulted in many, common genetic variants in the population. These alleles, now prevalent, generally have a small effect or low penetrance and may not have equal impacts on gene function. GWA studies have successfully identified thousands of single nucleotide polymorphism (SNPs) associated with a variety of complex diseases (Welter et al., 2014). In addition, these studies provide insight into

common pathways or genes not previously implicated in disease processes by other methods of analysis. The presence of these multiple low penetrance polymorphisms in the general population could explain some of the missing heritability of certain complex diseases.

Other theories of missing heritability include rare variants of large effects, rare combinations of common variants, parent-of-origin effects, or epigenetics (Eichler et al., 2010). Parent of origin effects have been described in cancer and for type 1 and 2 diabetes, where variants have been discovered that can either confer or reduce risk depending on the parent of origin (Kong et al., 2009). These effects are difficult to evaluate, and could be underestimated in models that do not include parental origin (Eichler et al., 2010). Epigenetics, the heritable changes that alter gene expression without affecting the underlying DNA sequence, can also contribute to heritability. The environmentally induced epigenetic change can be transmitted to one of more generations, as an "epigenetic persistence of genetic memory" (Eichler et al., 2010). For example, a paternal grandmother who experienced sharp changes in food supply during childhood confers an increased risk of cardiovascular mortality to her female grandchildren (Bygren et al., 2014). Another possibility is that large variants, individually rare but common in the general population exist. In fact, these large deletions and duplications result in many of the genetic differences observed in humans (Sebat et al., 2004). A study conducted by Itsara *et al*. identified large variants in many individuals, variants larger than 500 kb in 5% – 10% of individuals and variants greater than 1 Mb in 2% (Itsara et al., 2009). Up to 8% of people are estimated to carry a large duplication or deletion greater than 500 kb and occurring at a frequency of < 0.05% (Eichler et al., 2010). While these rare variants are common, 1% carry >1 Mb, they are generally considered deleterious and have been implicated in several conditions such as neurological disease, schizophrenia, mental delays (Itsara et al., 2009) and autism (Sebat et al., 2007).

Most variants discovered through GWAS are associated with only a small increased risk of disease and have a small predictive value but the magnitude of their attributable risk may be large due to their frequency (Manolio et al., 2009b). In most cases of common complex disorders the combined contribution of loci to disease variation is often <10% (Bjorkegren, Kovacic, Dudley, & Schadt, 2015). For example, the 153 known coronary artery disease (CAD)-associated variants explain less than 10.6% of the genetic variation observed across the population (Bjorkegren et al., 2015). Another way to assess heritability is to consider all measured SNPs together. A study by Yang *et al*. utilized a linear model to explain the complex trait of height by estimating the variance explained by all height-associated SNPs together (Yang et al., 2010). When applying this "polygenic" model to height 294, 831 SNPs explained as much as 45% of height variance, a significant increase relative to the 5% explained by published individual SNPs (Yang et al., 2010). Currently, there is a wide spectrum of identified SNPs corresponding to various complex diseases. The approach of combining individual SNPs associated with complex diseases could increase the predictive value of risk alleles discovered through GWAS.

Additionally, these association studies can benefit from populations in tight linkage disequilibrium (LD). LD is the non-random association of alleles that likely arose in a single, common ancestor. In association studies, studying tightly linked variants can conserve resources. Genetically isolated populations, particularly those founded recently could have longer stretches of LD than other populations. Populations with descendants from a small number of founders, occurring relatively recently exhibit extensive LD with few gaps (Service et al., 2006). In theory, these isolated populations could achieve better genome coverage or require fewer markers for association studies (Service et al., 2006). It is suggested that association studies from these populations could be more useful than out-bred populations (Service et al., 2006). Assessing the excess disease burden of

deleterious variants in a bottleneck population can offer many advantages to elucidating the burden of disease in these populations and others.

2.1.4 Founder Populations as Models for Studying Genetic Disease

Most complex diseases have a strong heritable component, but identifying all potential genetic variants has been difficult. An important question is why these common disease-associated variants are so prevalent in the human genome. A current theory suggests that such diseases arose due to the accumulation of mildly detrimental variant alleles in the human genome. In theory, genetically isolated populations, having undergone a recent bottleneck, could offer advantages to studying complex disease due to the higher frequency of certain risk variants present in such populations. A recent study by Lim and colleagues used exome sequencing of a Finnish population to compare variant alleles to other similar European populations. They reported more unusually strong acting alleles and variants at higher frequencies in the Finnish population, up to 1% for highly penetrant and lethal conditions (Lim et al., 2014; Polvi et al., 2013). They proposed that while the bottleneck removed most rare variants, the ones that were still present had become substantially elevated in frequency and that the Finns have an enrichment of low-frequency (0.5% – 5%) loss-of-function (LoF) variants (de la Chapelle & Wright, 1998; Lim et al., 2014). This observation led them to pursue potential health-related consequences of the deleterious low frequency alleles with significant results. For example, a nonsense variant in the *Translation Elongation Factor, Mitochondrial gene* (*TSFM*) was present at 1.2% allele frequency in the Finn population, but was absent in the out-bred European population. Furthermore, this variant was not found in a homozygous state in over 36,000 Fins, and two families with Finnish mitochondrial disease were identified with compound heterozygosity of this nonsense variant (each family had a different second hit in the *TSFM* gene) (Lim et al., 2014). This suggests that a loss of this gene is severely detrimental in humans, and that a complete loss could result in

embryonic lethality, or severe early childhood disease (Lim et al., 2014). The high prevalence of this allele in the Finnish population is likely a consequence of the population bottleneck, where alleles were drawn at random from rare variants from the original population and spread at a higher frequency in the new population. While the out-bred population had higher proportions of LoF variants, they were either selected against or arose recently, keeping the frequencies very low.

Founder populations such as the Finnish population also provide good resources for GWA studies of complex traits due to synthetic associations. Synthetic associations can be defined as multiple rare variants that occur more frequently in association with one allele at a common SNP than with the other (Goldstein, 2009). Dickson *et al.* postulate that multiple low-frequency variants spanning large regions of the genome may account for some observed disease-association (Dickson, Wang, Krantz, Hakonarson, & Goldstein, 2010). The causal variants underlying GWAS associations are not necessarily frequent as GWA studies can find associations resulting from SNPs in partial linkage disequilibrium to rare mutations with strong effect. Thus, founder populations are useful study populations due to an enrichment of mutations with a strong effect that are otherwise rare in outbred populations.

A recent study using whole-exome sequencing of a founder population of French-Canadians that had experienced a bottleneck less than 20 generations ago was done by Casals *et al*. They discovered reduced levels of diversity, higher homozygosity, an excess of rare variants and a higher proportion of putatively damaging functional variants in the founder population (Casals et al., 2013). The authors hypothesize this finding could partially explain the increased incidence of genetic disease in the group (Casals et al., 2013).

Another recent study of European Romanis, who recently underwent a population bottleneck resulting in a higher prevalence of several diseases, was conducted by

Mendizabal *et al*. Researchers tested SNPs associated with several metabolic and cardiovascular disorders previously reported at a high prevalence in the Romani and compared the average risk allele frequencies from the Romani population to an outbred European population (Mendizabal, Lao, Marigorta, Kayser, & Comas, 2013). The Romani population exhibited excess homozygosity and had higher than average risk alleles frequencies when compared to other European populations for four of the five cardiovascular/metabolic conditions tested, with one reaching a significant difference (Mendizabal et al., 2013). The results also indicated the genetic risk for these conditions matched the known pattern of morbidity, suggesting common risk alleles discovered by GWAS (at a MAF of greater than 5%) could somewhat explain the higher disease prevalence in the founder Romani population (Mendizabal et al., 2013).

A study by Lohmueller *et al*. found that each person likely carries several hundred potentially damaging SNPs (Lohmueller et al., 2008). Apparently healthy individuals can have many disadvantageous variants in their genome without showing any obvious ill effects, possibly due to the late onset of the disease or mild clinical phenotype. It is now suspected that up to 10% of functional genes in the human genome varies between individuals, with only minor health consequences (Xue et al., 2012). Xue *et al*. found that the CEU population from the 1000 Genomes Project website, genotyped from healthy adult residents of Utah with Northern and Western European ancestors, had ~12,000 derived missense-alleles per individual. By assessing the number of potentially deleterious variants in healthy humans populations they found an average individual might eventually be found to carry upwards of 400 damaging variants and more than 2 disease-causing ones. Each individual likely has up to 515 missense substitutions, approximately 60 of which were homozygous and predicted to be highly damaging, and many more disease-associated SNPs (Xue et al., 2012). The high numbers of disease associated SNPs in apparently healthy individuals gives credence to the common variant, common ancestor hypothesis of complex disease. Where each individual harbours many

disease-associated SNPs but due to low-penetrance or late onset, do not have the disease.

Kryukov *et al*. propose that the high incidence rate of complex diseases suggests high cumulative frequencies of medically deleterious variants are probably present in the human population (Kryukov, Pennacchio, & Sunyaev, 2007). Up to 70% of low-frequency missense alleles are mildly deleterious and associated with a small heterozygous fitness loss (Kryukov et al., 2007). Why such high numbers of detrimental SNPs occur is potentially due to several factors; the complex-disease associated polymorphisms present now were potentially evolutionarily advantageous. This is either because of late-onset disease, or a changing direction of selection, the "thrifty-gene hypothesis". For example, genes that predispose modern humans to obesity possibly rose to high frequency in early human history due to selective advantages during times of food scarcity. Another theory is that balancing selection, maintaining a population of heterozygotes, is advantageous. For example, the haemoglobin mutation causing sickle-cell anemia in homozygous individuals is protective against malaria in heterozygotes. Another possibility is that the negative effect of a mutation on a trait is balanced by the positive effect on another, so called "antagonistic pleitropy". There are many polymorphisms in the GWAS database that offer a protective effect on one hand but a detrimental effect on a different disease. A mutation-selection hypothesis states a balance between a high rate of deleterious mutations is balanced by purifying selection (Kimura & King, 1979). Among *de novo* missense mutations in human proteins, it is estimated that approximately 20% are strongly detrimental, while 53% are mildly deleterious and the 27% are effectively neutral (Kryukov et al., 2007).

A current hypothesis is that many, low-penetrance polymorphisms exist in the human genome and exert a mild, increased predisposition to complex disease. Studying the proportion of such common variants in certain populations, particularly founder populations with known increased prevalence of disease, can provide insight into the history or burden of such diseases and may help inform public health policies for the

population.

2.1.5 Hypothesis

Founder populations have inherent genetic consequences such as increased sensitivity to genetic drift, population bottlenecks, inbreeding and increased prevalence of founder mutations. Newfoundland, a known founder population, has experienced some of these genetic effects. They are predicted to have modern-day genetic consequences for Newfoundlanders in terms of increased burden of disease. This theory is supported by numerous studies of diseases with known mutations present at higher frequencies in the Newfoundland population. **Based on the hypothesis that complex-diseases are attributed to many low-penetrance SNPs, I hypothesize that a founder effect contributed to a high rate of some risk variants in the Newfoundland population and this may contribute to increased rates of certain complex diseases.**

2.1.6 Objectives

The objectives for this study include:

1. Predicting the prevalence of complex diseases in the Newfoundland and Labrador founder population based on the mean frequencies of disease-associated SNPs as compared to an out-bred population of European ancestry.
2. Comparing the prevalence of disease in NL with the national statistics from the rest of Canada in an effort to support the disease burden predictions from mean frequencies of risk variants.

2.2 CHAPTER 2. Materials and Methods

2.2.1 Study Population

A total of 494 individuals were included in the study, recruited by random digital dialling across Newfoundland and Labrador as control participants of a colorectal cancer (CRC) research study (Woods et al., 2005). The only inclusion criterion was that the individual had not been diagnosed with colorectal cancer at any point in their life. The study was conducted from 2001-2003; blood samples and health information were collected from participants. DNA was extracted from the participants and genotyped using an Affymetrix Axiom Genome-Wide Array (Santa Clara, CA, USA) at the USC Norris Comprehensive Cancer Centre (LA, USA). The array covered 1.3 million SNPs and samples were then subject to quality control as follows. Unintended replicate samples were removed. Improper genotype to reported sex and individuals with missing genotype at ≥ 3 % were removed. Two "wells" were used for each genotype sample, labeled "peg A" and "peg B" to increase the number of SNPs tested, because a limited number of SNPs can be tested in each well. Some overlap of SNPs existed between the wells, which was removed later in the QC process. SNPs with poor quality were removed as per Affymetrix recommended guidelines as follows; SNP call rate <95 %, Fisher's Linear Discriminant and SNPs out of Hardy-Weinberg Equilibrium at < 0.0001 and duplicates. Overlapping SNPs between peg A and B with concordance < 95 % and >1 mismatch and study replicates with a lower call rate were removed. Pegs A and B were then merged using PLINK (Purcell, 2007) (http://pngu.mgh.harvard.edu/~purcell/plink/). PLINK was then used to estimate IBD and find related individuals. We used STRUCTURE (http://pritchardlab.stanford.edu/structure.html) to identify and remove individuals with European ancestry < 80 %. In total, 1 individual was removed for improper genotype to reported sex and 16 samples were removed from both peg A and peg B due to study replicates with a low call rate. A total of 9552 SNPs from peg A and 17761 SNPs from peg B were removed due to poor quality as per Affymetrix guidelines, 1143 peg A SNPs

and 1906 peg B SNPs were then removed due to replicate concordance. Following merging of peg A and peg B, 2072 duplicate SNPs were removed.

GTOOL (V0.75) software was used to transform the data to a format recognizable by the IMPUTE2 program (PED to GEN conversion) (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html) (Freeman & Marchini, 2007). The computer program IMPUTE2 (Pritchard, Stephens, & Donnelly, 2000) (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) was then utilized to impute the whole genome data set. Briefly, the program uses a fine-scale recombination map and reference panel to fill in missing genotypes allowing for more SNPs to be tested for association and increasing study power.

A genetically similar out-bred population from Utah was used as a comparison to the NL population. The CEPH population, collected from 90 individuals in 1980 for the HapMap International Project, was taken from US Utah residents with ancestry from Northern and Western Europe (International HapMap Consortium, 2003). This population will be referred to as the "CEU" population. The CEU samples from Hapmap3 were genotyped on the Affymetrix Human SNP array 6.0 and the Illumina Human 1M-single Beadchip (International HapMap Consortium, 2003). Genotype data were downloaded from the International HapMap Project website (http://hapmap.ncbi.nlm.nih.gov).

2.2.2 Disease Associated Single Nucleotide Polymorphisms

The NIH National Human Genome Research Institute's catalog of published Genome-Wide association studies (https://www.genome.gov/26525384) was used to compile a list of SNPs for the study (Hindorff et al., 2015). As of February 2015 the catalogue included 2111 publications and 15,396 SNPs that associate with various traits. The search for relevant SNPs from the GWAS catalogue was restricted to those reaching genome-wide significance, at $p < 10^{-8}$. This narrowed the list to 6,984 SNPs. The results

were downloaded and further filtered. SNPs that were not found in Caucasian populations (of European descendent) were excluded, further narrowing the list to 4,435 SNPs. Many non-disease associated SNPs were present in the list, including traits such as aging, alcohol consumption, birth weight, educational attainment, etc. These traits, as well as those of personal characteristics such eye colour or biomarkers of disease, were removed. Next, SNPs with multiple haplotypes or reaching genome-wide significance through association with more than one disease or condition were removed. Only those SNPs associated with one specific disease reaching genome-wide significance in a Caucasian population were included. A final list of approximately 1,800 SNPs was used for comparison between populations (Figure 2.1). However, certain allele frequencies were not available from either the NL or CEU data for some of the SNPs. Therefore, they were not included in the analysis. Of the 1,800 SNPs used in the final analysis, allele frequencies from both the CEU population and the NL population were available for only 1,391. The disease-associated variants were determined based on information from the original publication describing the association or from the GWAS catalogue database. The online mendelian inheritance in man (OMIM) (McKusick-Nathans Institute of Genetic Medicine, John Hopkins University (Baltimore, MD), 2015) and Ensembl (v. 80) (Cunningham et al., 2015) databases were also used when necessary.

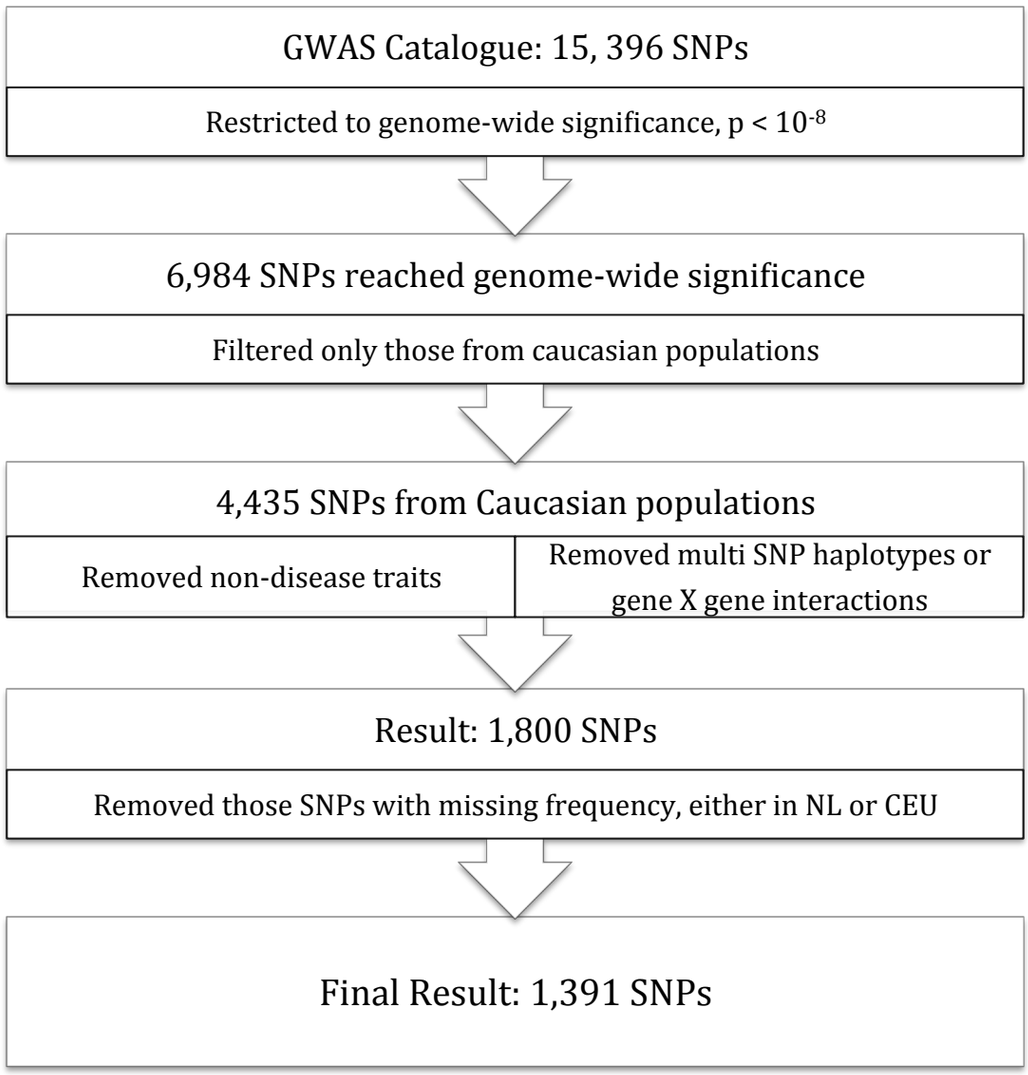


Figure 2.1 Summary of SNP filtering process

2.2.3 Comparison Analysis

Information for an initial 4,435 SNPs downloaded from the GWAS catalogue, as previously mentioned (Figure 2.1), were extracted from the imputed NL population file. The data included SNP rs ID, chromosome location, associated disease, allele frequency and total allele counts. All SNPs reported in the GWAS catalogue for a specific disease were grouped into one category. Then, the risk alleles for all SNPs associated with that disease (as recorded in the GWAS catalogue) were determined and the mean frequencies and standard errors were calculated using STATA/SE 11.2 (Stata Corp, College Station, Texas, USA). For example, all SNPs associated with coronary artery disease were placed

in one category and the mean frequency and standard error were calculated for the risk alleles for each SNP. Each disease has been grouped further by anatomical region or disease type for simplicity. For example, all neurological diseases were grouped and graphed together (the SNP calculations were not combined, they were kept as distinct means for each disease). The mean frequencies of disease-associated alleles from the NL population were then compared to mean frequencies of the CEU population for each disease. The Wilcoxon rank-sum test was used to evaluate the difference between mean risk-allele frequencies from the NL and CEU populations to test whether there was a significant difference between populations. The statistical significance was defined at $p < 0.05$ by STATA software. A summary of SNPs in LD that were used in the same mean frequency calculations are included. LD was verified using pairwise SNP Annotation and Proxy Search (Version 2.2, Pairwise LD function) (A. D. Johnson et al., 2015) (http://www.broadinstitute.org/mpg/snap/ldsearchpw.php). Finally, a comparison is drawn between the mean frequencies of disease alleles in NL and the provincial statistical prevalence of the disease. The provincial statistical average was then compared to the national average. The Canadian national average is used as a comparison population due to the environmental similarity and geographic proximity.

2.3 CHAPTER 3. Results

A total of 1,391 SNPs and 146 disease conditions were included in the final analysis. All SNPs, frequencies and disease-associations are reported in table format. Due to low statistical power, only those diseases with frequencies available from four or more SNPs were used for statistical calculations. Mean frequencies and standard errors are displayed in graphs of mean frequency comparisons.

2.3.1 Autoimmune Conditions

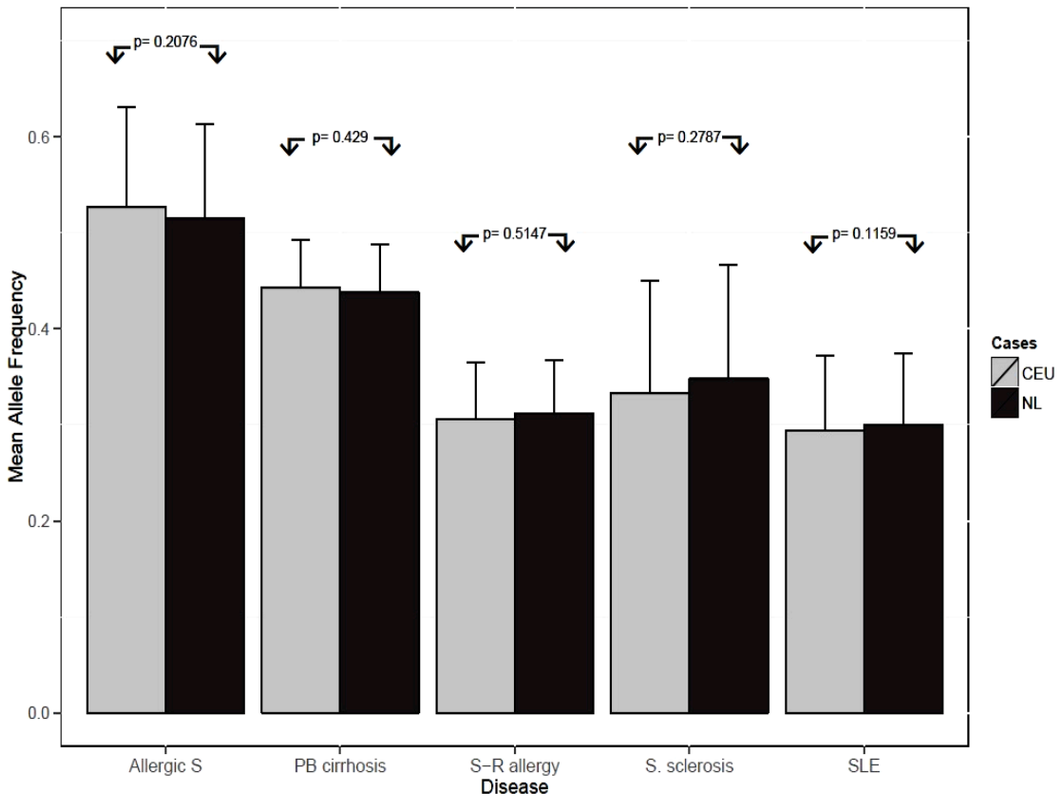


Figure 2.2 Mean risk allele frequencies of autoimmune conditions

"Allergic S" is allergic sensitization, "PB cirrhosis" is primary biliary cirrhosis, "S-R allergy" is self-reported allergy, "S. sclerosis" is systemic sclerosis and "SLE" is systemic lupus erythematosus.

Table 2.1 SNPs association with autoimmune disorders

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU Population
Allergic Sensitization	rs2155219	0.498	0.487
	rs1059513	0.879	0.920
	rs10056340	0.197	0.183
	rs3771175	0.860	0.880

	rs17616434	0.742	0.779
	rs9865818	0.422	0.433
	rs4410871	0.281	0.290
	rs17454584	0.241	0.250
Total Mean Frequencies	N= 8	0.515	0.528
Self-reported allergy	rs10189629	0.143	0.107
	rs10497813	0.472	0.451
	rs9860547	0.432	0.460
	rs2101521	0.238	0.196
	rs7720838	0.385	0.344
	rs1438673	0.454	0.491
	rs7032572	0.150	0.140
	rs9303280	0.475	0.500
	rs6021270	0.058	0.059
Total Mean Frequencies	N= 9	0.312	0.305
Systemic lupus erythematosus	rs7574865	0.252	0.234
	rs12531711	0.880	0.876
	rs13277113	0.238	0.235
	rs10488631	0.120	0.124
	rs9888739	0.104	0.095
	rs5029939	0.025	0.027
	rs3821236	0.206	0.217
	rs4963128	0.696	0.671
	rs6445975	0.321	0.248
	rs12537284	0.154	0.133
	rs3131379	0.107	0.080
	rs11574637	0.128	0.142
	rs10516487	0.668	0.748
Total Mean Frequencies	N= 13	0.300	0.295
Systemic sclerosis	rs11642873	0.793	0.779
	rs10488631	0.120	0.124
	rs2233287	0.093	0.058
	rs7574865	0.252	0.234
	rs2056626	0.625	0.59
	rs3821236	0.206	0.217
Total Mean Frequencies	N= 6	0.348	0.334

A pair of SNPs was found in LD for SLE, rs12531711/rs10488631 (r^2 = 1.0, D' = 1.0).

2.3.2 Brain and Neurological Conditions

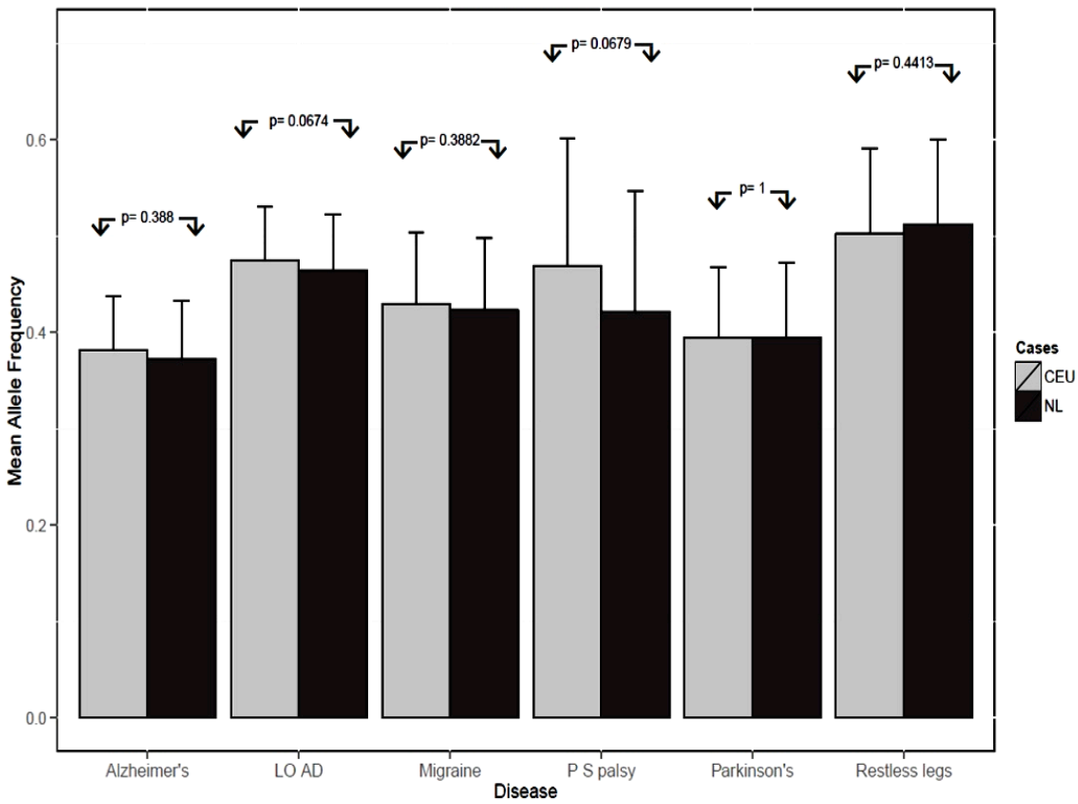


Figure 2.3 Mean risk allele frequencies of brain and neurological conditions.

"Alzheimer's" is Alzheimer's disease, "LO AD" is late-onset Alzheimer's disease, "Migraine" is migraine headaches, "P S palsy" is progressive supranuclear palsy, "Parkinson's" is Parkinson's disease, "Restless legs" is restless leg syndrome

Table 2.2 SNPs associated with brain and neurological conditions

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Alzheimer's disease	rs7561528	0.356	0.366
	rs1562990	0.580	0.509
	rs157580	0.360	0.371
	rs536841	0.683	0.656
	rs744373	0.293	0.317
	rs3764650	0.097	0.107
	rs3818361	0.167	0.255
	rs2075650	0.139	0.161
	rs11136000	0.607	0.647
	rs3851179	0.613	0.585
	rs2075650	0.139	0.161
	rs6859	0.432	0.446
Total Mean Frequencies	N= 12	0.372	0.382
Alzheimer's disease (late onset)	rs6656401	0.160	0.241

	rs10948363	0.263	0.280
	rs11771145	0.638	0.652
	rs9331896	0.605	0.638
	rs10792832	0.612	0.583
	rs2718058	0.604	0.643
	rs1476679	0.746	0.683
	rs17125944	0.090	0.080
	rs11218343	0.963	0.969
	rs10498633	0.798	0.813
	rs9349407	0.263	0.275
	rs11767557	0.784	0.797
	rs4938933	0.585	0.505
	rs3865444	0.654	0.679
	rs6701713	0.167	0.255
	rs7561528	0.355	0.366
	rs561655	0.351	0.375
	rs744373	0.293	0.317
	rs12989701	0.162	0.152
	rs4420638	0.182	0.183
Total Mean Frequencies	N=20	0.464	0.474

Disease	**SNP ID**	**Risk Allele Frequency NL population**	**Risk Allele Frequency CEU population**
Migraine	rs2651899	0.408	0.416
	rs12134493	0.129	0.124
	rs6741751	0.889	0.858
	rs11759769	0.198	0.237
	rs4379368	0.082	0.088
	rs11172113	0.604	0.633
	rs6478241	0.392	0.389
	rs9349379	0.633	0.628
	rs3790455	0.346	0.341
	rs7640543	0.349	0.376
	rs10166942	0.800	0.829
	rs1835740	0.246	0.230
Total Mean Frequencies	N= 12	0.423	0.429
Progressive supranuclear palsy	rs1411478	0.386	0.473
	rs7571971	0.260	0.267
	rs1768208	0.251	0.295
	rs12203592	0.788	0.841
Total Mean Frequencies	N= 4	0.421	0.469
Parkinson's disease	rs11248060	0.094	0.124
	rs356220	0.453	0.403
	rs12456492	0.314	0.319
	rs10513789	0.816	0.757

	rs6812193	0.646	0.602
	rs34372695	0.032	0.031
	rs11724635	0.553	0.500
	rs356219	0.350	0.425
	rs1491942	0.185	0.192
	rs11248051	0.083	0.111
	rs2736990	0.439	0.469
	rs199533	0.774	0.805
Total Mean Frequencies	N= 12	0.395	0.395
Restless legs syndrome	rs2300478	0.235	0.237
	rs9357271	0.806	0.765
	rs1975197	0.185	0.190
	rs12593813	0.642	0.705
	rs6747972	0.442	0.398
	rs3104767	0.605	0.602
	rs4626664	0.164	0.127
	rs3923809	0.723	0.726
	rs9296249	0.808	0.770
Total Mean Frequencies	N= 9	0.512	0.502
Musician's dystonia	rs11655081	0.050	0.053
Creutzfeldt-Jakob disease (variant)	rs6107516	0.738	0.757
Creutzfeldt-Jakob disease	rs1799990	0.644	0.655

Two SNPs are in LD for Parkinson's, rs11248060/rs11248051 (r^2 = 0.884, *D'* =1.0) and for restless legs syndrome, rs9357271/rs9296249 (r^2 = 0.976, *D'* = 1.0).

2.3.3 Arthritis

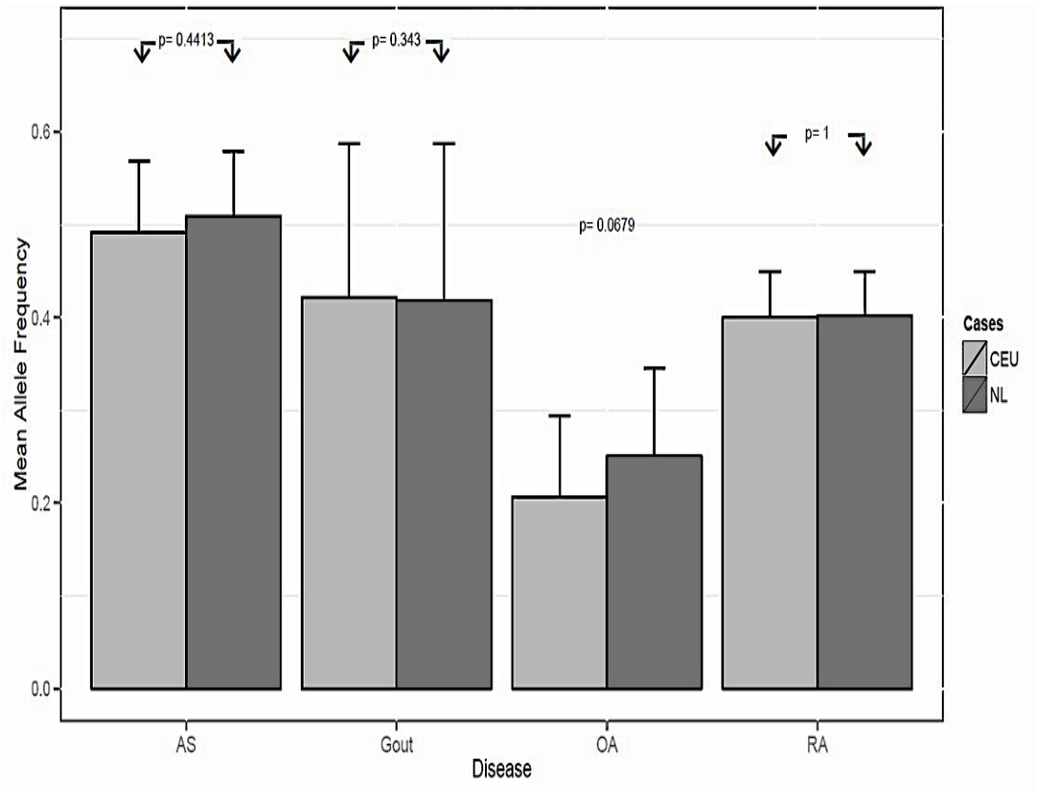


Figure 2.4 Mean risk allele frequencies of forms of arthritis.

"AS" is ankylosing spondylitis, Gout, "OA" is osteoarthritis and "RA" is rheumatoid arthritis

Table 2.3 SNPs associated with types of arthritis

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Ankylosing spondylitis	rs11209026	0.939	0.959
	rs2297909	0.695	0.658
	rs10865331	0.414	0.288
	rs30187	0.360	0.310
	rs378108	0.427	0.496
	rs11249215	0.494	0.482
	rs11616188	0.419	0.433
	rs27434	0.227	0.191
	rs2242944	0.607	0.608
Total Mean Frequencies	N= 9	0.509	0.492
Gout	rs4475146	0.221	0.246
	rs1481012	0.892	0.898
	rs2231142	0.108	0.111
	rs734553	0.760	0.739
	rs2231142	0.108	0.111

Total Mean Frequencies	N= 5	0.418	0.421

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Osteoarthritis (hip)	rs6094710	0.038	0.033
	rs3757837	0.141	0.079
Osteoarthritis	rs11177	0.389	0.339
	rs12982744	0.432	0.375
Total Mean Frequencies	N=4	0.250	0.207
Psoriatic arthritis	rs12188300	0.076	0.112
Rheumatoid Arthritis	rs2476601	0.907	0.882
	rs653178	0.480	0.438
	rs1893217	0.183	0.116
	rs1953126	0.354	0.380
	rs7574865	0.252	0.236
	rs2298428	0.189	0.165
	rs934734	0.476	0.442
	rs6859219	0.795	0.830
	rs874040	0.283	0.295
	rs3093023	0.399	0.410
	rs10488631	0.120	0.121
	rs11676922	0.425	0.457
	rs951005	0.870	0.821
	rs706778	0.424	0.406
	rs6920220	0.248	0.158
	rs4810485	0.250	0.246
	rs13017599	0.358	0.375
	rs231735	0.483	0.487
	rs2736340	0.236	0.241
	rs6679677	0.092	0.116
	rs10499194	0.739	0.825
	rs3761847	0.423	0.473
	rs660895	0.245	0.277
Total Mean Frequencies	N= 23	0.273	0.306

A pair of SNPs in LD were included in the mean calculation for gout; rs1481012/rs2231142 (r^2 = 0.92, D' = 1.0) and for RA: rs2476601/rs6679677 (r^2 = 1.0, D' = 1.0).

2.3.4 Bone Conditions

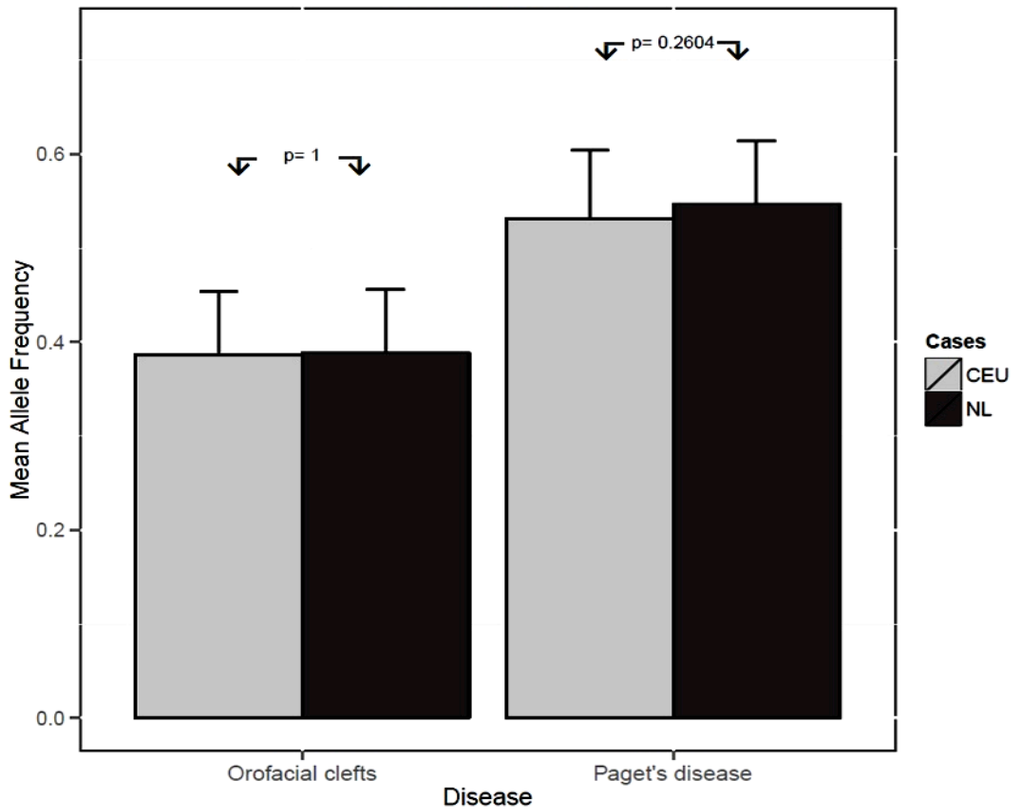


Figure 2.5 Mean risk allele frequencies of bone conditions.

Table 2.4 SNPs associated with disorders of the bone

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Osteoporosis-related phenotypes	rs2062375	0.441	0.467
Orofacial clefts	rs560426	0.573	0.504
	rs861020	0.214	0.221
	rs987525	0.232	0.221
	rs7078160	0.171	0.168
	rs13041247	0.598	0.597
	rs742071	0.417	0.446
	rs8001641	0.512	0.549
Total Mean Frequencies	N= 7	0.388	0.387
Paget's disease	rs10494112	0.202	0.173
	rs4294134	0.846	0.848
	rs2458413	0.578	0.535
	rs1561570	0.570	0.535
	rs10498635	0.821	0.863
	rs5742915	0.573	0.504
	rs3018362	0.367	0.371

	rs2957128	0.417	0.403
	rs484959	0.544	0.553
Total Mean Frequencies	N= 9	0.547	0.532
Disc degeneration (lumbar)	rs17034687	0.922	0.908
Sagittal craniosynostosis	rs1884302	0.355	0.341
	rs10262453	0.693	0.711
Osteosarcoma	rs1906953	0.147	0.095

No SNPs in LD were included in the same mean risk frequency calculations.

2.3.5 Common Types of Cancer

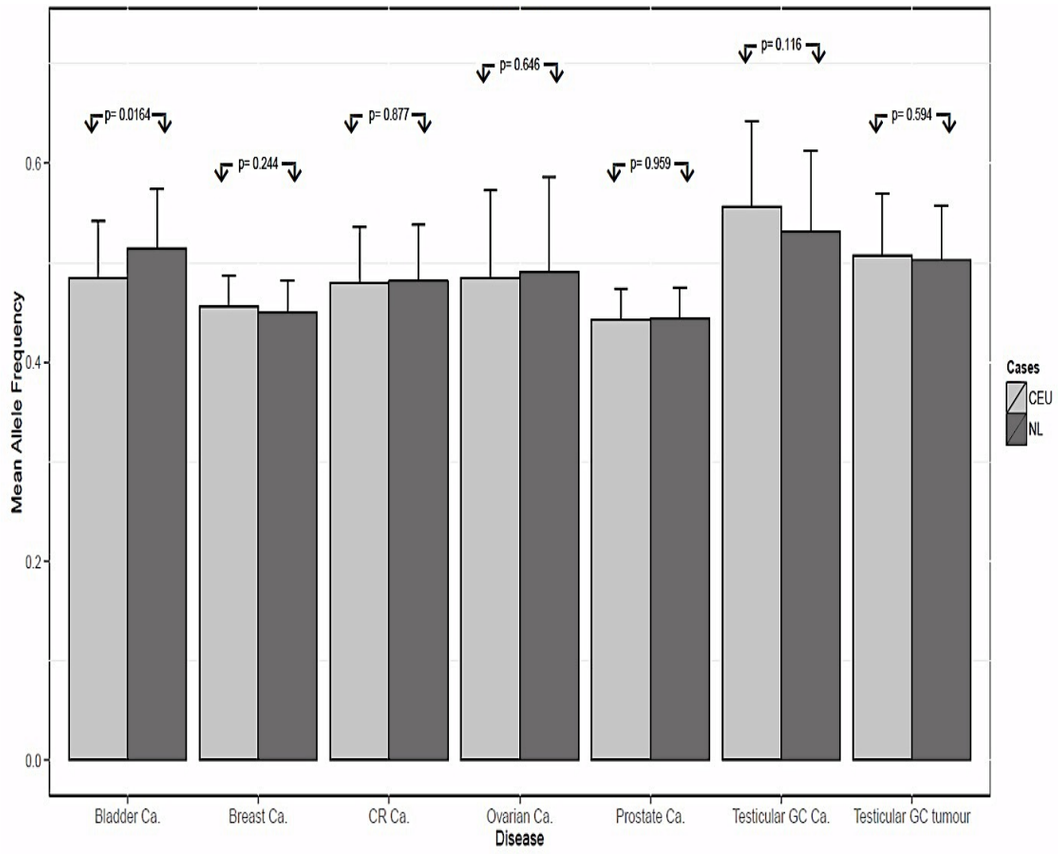


Figure 2.6 Mean risk allele frequencies of common forms of cancer.

Bladder cancer, Breast cancer, "CR ca." is colorectal cancer, Ovarian cancer, prostate cancer, "Testicular GC Ca." is testicular germ cell cancer, "Testicular GC tumour" is testicular germ cell tumour.

Table 2.5 SNPs associated with common types of cancer

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Renal cell carcinoma	rs718314	0.240	0.217
	rs7579899	0.381	0.354
	rs7105934	0.900	0.942
Bladder cancer	rs10936599	0.752	0.730
	rs710521	0.732	0.686
	rs798766	0.199	0.199
	rs401681	0.537	0.566
	rs1495741	0.772	0.712
	rs9642880	0.457	0.425
	rs8102137	0.345	0.292
	rs1014971	0.697	0.655
	rs7238033	0.418	0.350
	rs17674580	0.335	0.292
	rs2294008	0.416	0.420

Total Mean Frequencies	N= 11	0.515	0.484
Breast cancer	rs4245739	0.274	0.305
	rs10069690	0.255	0.301
	rs2046210	0.352	0.288
	rs10771399	0.881	0.833
	rs3803662	0.245	0.248
	rs8170	0.209	0.216
	rs616488	0.656	0.668
	rs4849887	0.910	0.889
	rs16857609	0.242	0.293
	rs6762644	0.401	0.425
	rs6828523	0.875	0.912
	rs1432679	0.478	0.411
	rs11242675	0.626	0.650
	rs204247	0.472	0.482
	rs720475	0.725	0.737
	rs9693444	0.334	0.319
	rs6472903	0.830	0.796
	rs11780156	0.164	0.155
	rs7072776	0.270	0.270
	rs3903072	0.528	0.573
	rs11820646	0.570	0.544
	rs17356907	0.690	0.708
	rs2236007	0.809	0.786
	rs2588809	0.158	0.183
	rs941764	0.359	0.339
	rs17817449	0.592	0.540
	rs13329835	0.211	0.181
	rs527616	0.658	0.633
	rs4808801	0.635	0.655
	rs3760982	0.480	0.460
	rs132390	0.030	0.017
	rs6001930	0.074	0.102
	rs11249433	0.414	0.425
	rs13387042	0.500	0.562
	rs4973768	0.473	0.438
	rs889312	0.283	0.308
	rs17530068	0.232	0.219
	rs3757318	0.073	0.067
	rs13281615	0.386	0.458
	rs865686	0.596	0.606
	rs10995190	0.866	0.872
	rs704010	0.387	0.433
	rs2981579	0.442	0.465
	rs3817198	0.280	0.327
	rs614367	0.167	0.185
	rs1292011	0.575	0.606
	rs999737	0.735	0.739
	rs6504950	0.705	0.693

	rs2823093	0.741	0.728
	rs2943559	0.087	0.071
	rs11814448	0.009	0.033
	rs4415084	0.402	0.376
	rs1562430	0.574	0.647
	rs1219648	0.429	0.465
	rs3112612	0.396	0.407
Total Mean Frequencies	N= 55	0.450	0.455
Breast cancer (male)	rs1314913	0.139	0.155
	rs3803662	0.245	0.248
Colorectal cancer	rs6983267	0.455	0.487
	rs4939827	0.457	0.473
	rs1321311	0.233	0.310
	rs3824999	0.500	0.487
	rs5934683	0.340	0.343
	rs6691170	0.363	0.389
	rs11169552	0.752	0.810
	rs4925386	0.691	0.644
	rs6687758	0.178	0.208
	rs10411210	0.902	0.916
	rs961253	0.407	0.403
	rs4444235	0.476	0.442
	rs7014346	0.403	0.310
	rs3802842	0.284	0.235
	rs10795668	0.337	0.323
	rs16892766	0.931	0.894
Total Mean Frequencies		0.482	0.480
Ovarian cancer	rs3814113	0.661	0.628
	rs10088218	0.867	0.875
	rs2072590	0.287	0.362
	rs7651446	0.043	0.049
	rs8170	0.209	0.216
	rs9303542	0.293	0.269
	rs11782652	0.939	0.892
	rs757210	0.659	0.567
	rs2072590	0.287	0.362
Total Mean Frequencies	N= 9	0.472	0.469
Endometrial cancer	rs4430796	0.483	0.491

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Prostate cancer	rs4245739	0.726	0.695
	rs3771570	0.156	0.142
	rs7611694	0.574	0.584
	rs1894292	0.452	0.460
	rs2273669	0.153	0.150
	rs1933488	0.598	0.597
	rs12155172	0.226	0.226
	rs11135910	0.166	0.173
	rs3850699	0.733	0.730
	rs11568818	0.521	0.531
	rs8008270	0.818	0.792
	rs7141529	0.484	0.478
	rs684232	0.340	0.367
	rs11650494	0.081	0.090
	rs7241993	0.685	0.700
	rs6062509	0.731	0.673
	rs2405942	0.818	0.769
	rs11672691	0.776	0.721
	rs10187424	0.537	0.553
	rs7584330	0.225	0.261
	rs10936632	0.516	0.517
	rs2242652	0.799	0.850
	rs10875943	0.282	0.283
	rs902774	0.140	0.142
	rs651164	0.674	0.655
	rs1016343	0.222	0.261
	rs12621278	0.935	0.960
	rs17021918	0.641	0.646
	rs7679673	0.640	0.624
	rs1512268	0.411	0.420
	rs7127900	0.166	0.235
	rs5759167	0.516	0.549
	rs12500426	0.432	0.460
	rs3123078	0.547	0.580
	rs7130881	0.169	0.190
	rs7501939	0.618	0.566
	rs1327301	0.358	0.373
	rs10934853	0.274	0.239
	rs445114	0.652	0.619
	rs8102476	0.569	0.504
	rs11228565	0.198	0.248
	rs10993994	0.409	0.341
	rs2735839	0.856	0.863
	rs7931342	0.509	0.531
	rs9364554	0.315	0.274
	rs6465657	0.511	0.509
	rs5945619	0.352	0.385
	rs6983267	0.455	0.487
	rs4242384	0.106	0.071

	rs5945572	0.360	0.379
	rs721048	0.170	0.137
	rs10896449	0.519	0.532
	rs4242382	0.106	0.071
	rs4430796	0.483	0.491
	rs1859962	0.483	0.473
	rs16901979	0.035	0.031
	rs1447295	0.106	0.071
Total Mean Frequencies	N= 57	0.444	0.443
Testicular germ cell tumour	rs9905704	0.668	0.699
	rs12699477	0.408	0.292
	rs4888262	0.512	0.553
	rs4624820	0.521	0.563
	rs995030	0.202	0.186
	rs2720460	0.615	0.628
	rs8046148	0.766	0.757
	rs2839186	0.438	0.398
	rs4635969	0.192	0.191
	rs755383	0.613	0.621
	rs2900333	0.593	0.688
Total Mean Frequencies	N= 11	0.503	0.507
Testicular germ cell cancer	rs2736100	0.478	0.473
	rs3782181	0.790	0.801
	rs4624820	0.521	0.563
	rs4635969	0.192	0.191
	rs755383	0.613	0.621
	rs2900333	0.593	0.688
Total Mean Frequencies	N= 6	0.531	0.556
Testicular Cancer	rs4474514	0.791	0.801
	rs7040024	0.754	0.783
	rs755383	0.613	0.621
Upper aero-digestive tract cancers	rs1229984	0.021	0.000
	rs971074	0.882	0.868

Upper aero-digestive tract cancers include cancers of the lips, tongue, major salivary glands, gums and adjacent oral cavity tissues, floor of the mouth, tonsils, oropharynx, nasopharynx, hypopharynx and other oral regions, nasal cavity, accessory sinuses, middle ear, and larynx.

For breast cancer, two SNPs were in LD rs2981579/rs1219648 (r^2 = 0.916, D' = 0.965). Several SNPs were in LD for prostate cancer calculations, rs7931342/rs10896449 (r^2 = 0.966, D' = 1.0), rs4242384/rs4242382 (r^2 = 1.0, D' = 1.0), rs4242384/rs1447295 (r^2 = 1.0, D' = 1.0), and rs4242382/ rs1447295 (r^2 = 1.0, D' = 1.0).

2.3.6 Renal and Digestive Tract Disorders

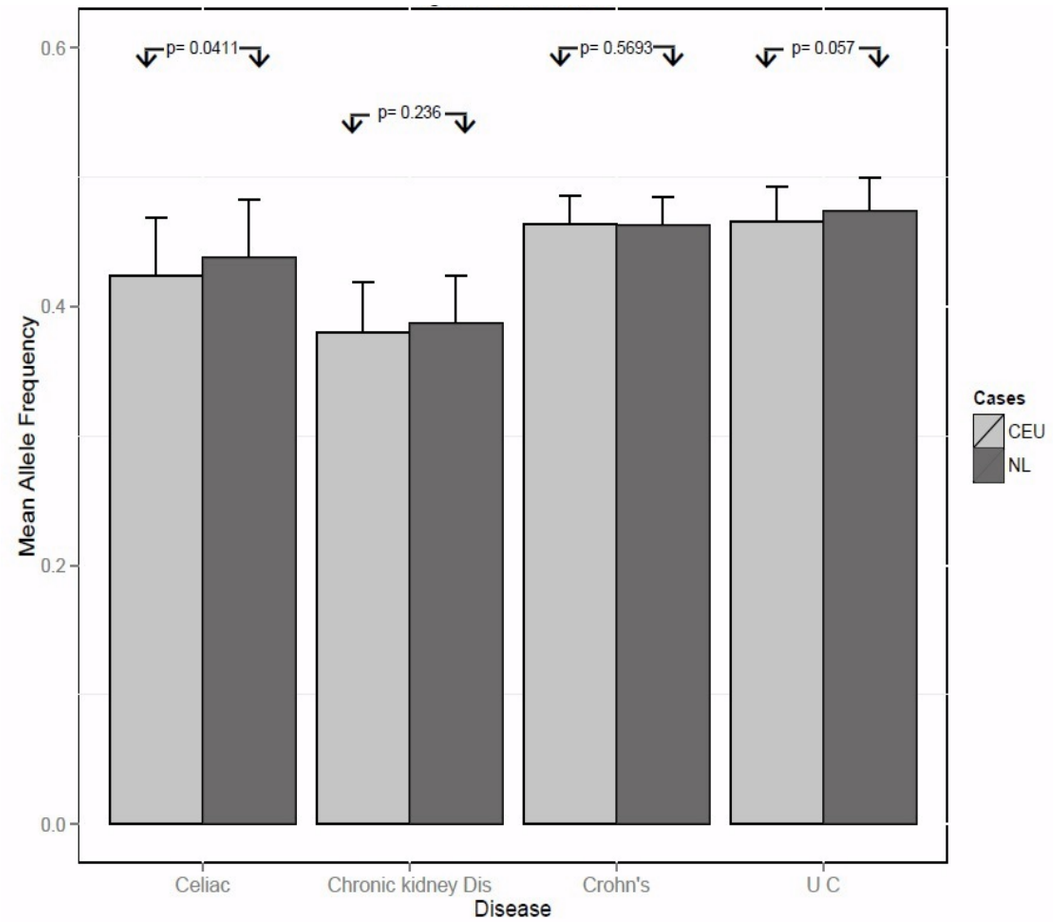


Figure 2.7 Mean risk allele frequencies of renal and digestive tract disorders.

"Celiac" is celiac disease, "Chronic kidney Dis" is chronic kidney disease", "Crohn's" is crohn's disease and "U C" is ulcerative colitis.

Table 2.6 SNPs associated with renal and digestive tract disorders

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Celiac disease	rs2816316	0.192	0.199
	rs13003464	0.412	0.385
	rs917997	0.237	0.204
	rs13010713	0.421	0.434
	rs4675374	0.197	0.186
	rs13098911	0.076	0.076
	rs17810546	0.130	0.100
	rs1464510	0.470	0.456
	rs13151961	0.880	0.866
	rs2327832	0.248	0.155
	rs1738074	0.416	0.465
	rs653178	0.480	0.438
	rs1893217	0.183	0.115

	rs3748816	0.698	0.664
	rs10903122	0.514	0.513
	rs296547	0.653	0.628
	rs17035378	0.748	0.695
	rs13314993	0.451	0.469
	rs11712165	0.388	0.412
	rs10806425	0.434	0.362
	rs802734	0.310	0.292
	rs9792269	0.762	0.752
	rs1250552	0.514	0.544
	rs11221332	0.190	0.212
	rs4819388	0.716	0.752
	rs13015714	0.236	0.204
	rs6822844	0.878	0.854
Total Mean Frequencies	N= 27	0.438	0.423
Kidney stones	rs219780	0.768	0.855
Chronic kidney disease	rs3925584	0.551	0.522
	rs267734	0.230	0.208
	rs1260326	0.422	0.420
	rs13538	0.250	0.208
	rs347685	0.289	0.265
	rs6420094	0.325	0.367
	rs881858	0.325	0.292
	rs7805747	0.305	0.296
	rs4744712	0.387	0.390
	rs653178	0.520	0.562
	rs626277	0.421	0.381
	rs1394125	0.343	0.350
	rs12460876	0.403	0.417
	rs2279463	0.149	0.124
	rs6465825	0.414	0.442
	rs10774021	0.300	0.305
	rs491567	0.237	0.200
	rs9895661	0.169	0.186
	rs4293393	0.828	0.823
	rs12917707	0.827	0.850
	rs17319721	0.439	0.420
	rs2467853	0.378	0.331
Total Mean Frequencies	N= 22	0.336	C0.330
Gallstones	rs11887534	0.059	0.085

Disease	**SNP ID**	**Risk Allele Frequency NL population**	**Risk Allele Frequency CEU population**
Crohn's disease	rs17391694	0.854	0.881
	rs6679677	0.908	0.885
	rs3897478	0.911	0.903

rs9286879	0.253	0.204
rs1728918	0.298	0.281
rs10865331	0.414	0.288
rs6716753	0.184	0.212
rs12994997	0.551	0.575
rs13126505	0.084	0.076
rs10065637	0.796	0.814
rs7702331	0.624	0.593
rs17695092	0.681	0.712
rs12663356	0.543	0.597
rs9491697	0.443	0.473
rs13204742	0.129	0.146
rs212388	0.425	0.438
rs864745	0.481	0.487
rs6651252	0.869	0.876
rs3764147	0.218	0.243
rs16967103	0.185	0.241
rs2945412	0.562	0.584
rs2024092	0.203	0.212
rs4802307	0.671	0.717
rs516246	0.508	0.531
rs2284553	0.605	0.558
rs2076756	0.263	0.323
rs11209026	0.939	0.959
rs3792109	0.559	0.580
rs10761659	0.537	0.553
rs1250550	0.640	0.633
rs4902642	0.583	0.603
rs181359	0.206	0.168
rs1819658	0.818	0.796
rs12720356	0.108	0.124
rs7517810	0.253	0.204
rs3810936	0.712	0.642
rs2062305	0.447	0.487
rs10495903	0.141	0.146
rs7423615	0.177	0.205
rs17293632	0.216	0.230
rs12521868	0.430	0.397
rs4809330	0.723	0.664
rs2476601	0.907	0.883
rs1893217	0.183	0.115
rs11742570	0.628	0.646
rs694739	0.652	0.584
rs6738825	0.481	0.456
rs4409764	0.497	0.455
rs11167764	0.780	0.783
rs11564258	0.023	0.027
rs713875	0.456	0.385
rs3197999	0.258	0.261
rs2413583	0.851	0.854
rs10758669	0.385	0.366
rs7714584	0.070	0.044

rs12722489	0.846	0.827
rs151181	0.403	0.372
rs2058660	0.237	0.192
rs6556412	0.336	0.319
rs2797685	0.168	0.186
rs3024505	0.153	0.181
rs2838519	0.398	0.420
rs2872507	0.449	0.473
rs740495	0.269	0.259
rs780093	0.404	0.394
rs8005161	0.085	0.140
rs281379	0.525	0.559
rs102275	0.318	0.351
rs2549794	0.430	0.429
rs13428812	0.317	0.314
rs1998598	0.321	0.305
rs12242110	0.290	0.319
rs359457	0.559	0.606
rs415890	0.497	0.491
rs3091315	0.720	0.690
rs4077515	0.453	0.482
rs10181042	0.434	0.412
rs7927997	0.392	0.388
rs1847472	0.625	0.692
rs13073817	0.373	0.372
rs17309827	0.632	0.646
rs736289	0.619	0.664
rs1736020	0.572	0.589
rs4871611	0.627	0.619
rs10045431	0.696	0.757
rs2301436	0.435	0.456
rs744166	0.612	0.549
rs7746082	0.255	0.311
rs2274910	0.699	0.634
rs11584383	0.701	0.668
rs6908425	0.804	0.823
rs1456893	0.685	0.650
rs1551398	0.626	0.619
rs17582416	0.325	0.383
rs7927894	0.391	0.394
rs11175593	0.022	0.027
rs1736135	0.568	0.580
rs762421	0.396	0.412
rs11465804	0.946	0.955
rs3828309	0.550	0.571
rs4613763	0.128	0.168
rs2188962	0.433	0.402
rs11747270	0.070	0.049
rs4263839	0.713	0.637
rs11190140	0.500	0.460
rs2542151	0.183	0.115
rs10995271	0.386	0.450

	rs17221417	0.309	0.350
	rs11805303	0.317	0.292
	rs10210302	0.551	0.580
	rs17234657	0.128	0.168
	rs1373692	0.629	0.646
Total Mean Frequencies	N= 112	0.463	0.463

Disease	**SNP ID**	**Risk Allele Frequency NL population**	**Risk Allele Frequency CEU population**
Ulcerative colitis	rs10797432	0.551	0.535
	rs6426833	0.561	0.509
	rs2816958	0.902	0.914
	rs17229285	0.471	0.438
	rs3774959	0.397	0.367
	rs254560	0.418	0.363
	rs6927022	0.499	0.482
	rs798502	0.673	0.712
	rs4380874	0.463	0.438
	rs4728142	0.435	0.398
	rs561722	0.644	0.699
	rs11150589	0.496	0.456
	rs7210086	0.803	0.792
	rs1126510	0.384	0.354
	rs6017342	0.502	0.562
	rs11209026	0.939	0.959
	rs1801274	0.470	0.491
	rs3024505	0.153	0.181
	rs7608910	0.414	0.379
	rs4676406	0.535	0.544
	rs9822268	0.289	0.259
	rs4510766	0.612	0.562
	rs6584283	0.523	0.562
	rs7134599	0.420	0.394
	rs2872507	0.449	0.473
	rs2836878	0.712	0.712
	rs734999	0.550	0.535
	rs35675666	0.803	0.833
	rs7524102	0.845	0.814
	rs2310173	0.478	0.420
	rs11676348	0.466	0.425
	rs267939	0.350	0.393
	rs6451493	0.628	0.646
	rs6871626	0.330	0.367
	rs943072	0.106	0.080
	rs6920220	0.248	0.165
	rs10758669	0.385	0.366
	rs4246905	0.738	0.676
	rs10781499	0.455	0.487
	rs12261843	0.249	0.283
	rs907611	0.294	0.319

	rs2155219	0.499	0.487
	rs17085007	0.188	0.150
	rs941823	0.746	0.779
	rs16940202	0.157	0.164
	rs2297441	0.791	0.752
	rs1297265	0.547	0.562
	rs2838519	0.398	0.42
	rs7554511	0.708	0.673
	rs17207986	0.069	0.058
	rs7809799	0.045	0.076
	rs4654925	0.461	0.467
	rs3024493	0.152	0.169
	rs1317209	0.172	0.137
	rs2201841	0.685	0.704
	rs10800309	0.309	0.296
	rs13003464	0.412	0.385
	rs3197999	0.258	0.261
	rs4957048	0.782	0.757
	rs1558744	0.434	0.403
	rs9858542	0.289	0.257
	rs3806308	0.633	0.602
Total Mean Frequencies	N= 62	0.473	0.465

Disease	**SNP ID**	**Risk Allele Frequency NL population**	**Risk Allele Frequency CEU population**
Barrett's esophagus	rs9936833	0.402	0.375
Infantile hypertrophic pyloric stenosis	rs11712066	0.763	0.774
	rs573872	0.182	0.168
	rs29784	0.487	0.469
Eosinophilic esophagitis (paediatric)	rs3806932	0.600	0.549
Hypospadias	rs1934179	0.335	0.337

A number of SNPs in LD were included in the mean risk allele calculation analysis (see table 2.7).

Table 2.7 Renal and Digestive Tract Disorders – SNPs in LD

Disease	SNP	Proxy	r^2	*D'*
Crohn's Disease	rs6679677	rs2476601	1	1
	rs9286879	rs7517810	1	1
	rs6716753	rs7423615	0.975	1
	rs12994997	rs3792109	0.983	1
	rs12994997	rs10210302	0.983	1
	rs12994997	rs3828309	0.949	0.982
	rs2076756	rs17221417	0.892	1
	rs11209026	rs11465804	0.905	1
	rs3792109	rs3828309	0.932	0.982
	rs12521868	rs2188962	1	1
	rs1893217	rs2542151	1	1
	rs11742570	rs1373692	0.964	0.982
	rs4409764	rs11190140	0.966	0.983
	rs11564258	rs11175593	1	1
	rs7714584	rs11747270	0.905	1
	rs2838519	rs762421	0.965	1
	rs7927997	rs7927894	1	1
	rs1736020	rs1736135	0.966	1
	rs4871611	rs1551398	1	1
	rs11209026	rs11465804	0.905	1
	rs10210302	rs3792109	0.965	0.983
	rs3828309	rs10210302	0.932	0.982
	rs4613763	rs17234657	1	1
Ulcerative colitis	rs13003464	rs7608910	1	1
	rs3197999	rs9858542	0.956	0.978
	rs13003464	rs7608910	1	1
	rs10797432	rs734999	0.966	0.983
	rs9822268	rs9858542	1	1
	rs9822268	rs3197999	0.956	0.978
	rs7134599	rs1558744	0.931	0.982
Celiac Disease	rs917997	rs13015714	1	1
	rs13151961	rs6822844	0.868	0.964

2.3.7 Metabolic syndrome and associated traits

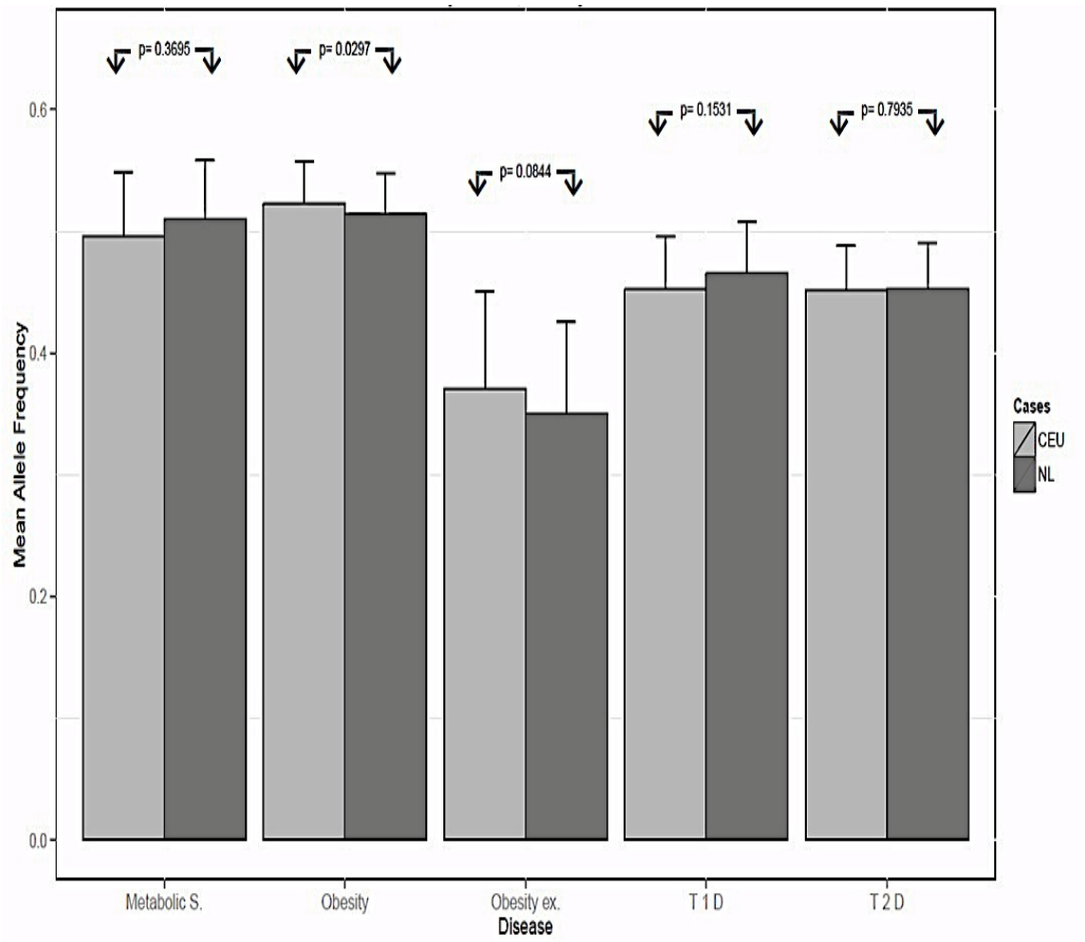


Figure 2.8 Mean risk allele frequencies of metabolic syndrome and associated traits.

"Metabolic S." is metabolic syndrome, obesity, "Obesity ex." is all obesity (extreme) combined, "T 1 D" is type 1 diabetes and "T 2 D" is type 2 diabetes

Table 2.8 SNPs associated with metabolic syndrome and associated traits

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Metabolic syndrome	rs560887	0.722	0.674
	rs6947830	0.519	0.518
	rs3757840	0.491	0.491
	rs1127065	0.556	0.567
	rs10830962	0.374	0.389
	rs673548	0.810	0.795
	rs268	0.290	0.004
	rs1883025	0.272	0.199
	rs10838681	0.675	0.730
	rs1532085	0.635	0.588
	rs247617	0.658	0.659
	rs8060686	0.832	0.867
	rs780094	0.403	0.394
	rs13226650	0.786	0.810

	rs7841189	0.901	0.863
	rs964184	0.143	0.121
	rs9940128	0.440	0.487
	rs295	0.775	0.767
	rs2266788	0.073	0.084
	rs2075290	0.078	0.043
	rs10790162	0.075	0.042
	rs173539	0.650	0.633
Total Mean Frequencies	N= 29	0.507	0.488
Metabolic syndrome (bivariate traits)	rs1387153	0.712	0.728
	rs3764261	0.658	0.655
	rs9939224	0.800	0.822
	rs301	0.224	0.252
	rs10830956	0.713	0.717
	rs2197089	0.461	0.442
	rs11820589	0.065	0.051
Total Mean Frequencies	N= 7	0.519	0.524
Obesity	rs17024258	0.022	0.042
	rs4735692	0.597	0.558
	rs8043757	0.408	0.450
	rs6711012	0.825	0.854
	rs10938397	0.441	0.446
	rs633715	0.225	0.243
	rs2030323	0.781	0.783
	rs2206277	0.197	0.096
	rs7138803	0.396	0.345
	rs10182181	0.491	0.469
	rs7141420	0.540	0.553
	rs7531118	0.522	0.567
	rs1800437	0.809	0.825
	rs9816226	0.832	0.851
	rs7498665	0.406	0.381
	rs8028313	0.765	0.780
	rs2307111	0.633	0.647
	rs11042023	0.638	0.681
	rs887912	0.288	0.319
	rs12446632	0.849	0.858
	rs1514177	0.413	0.460
	rs7185735	0.408	0.442
	rs10189761	0.829	0.854
	rs11152213	0.232	0.283
	rs2207139	0.189	0.102
	rs13130484	0.442	0.429
	rs3101336	0.614	0.637

	rs10423928	0.809	0.824
	rs9568867	0.125	0.10
	rs1516725	0.874	0.908
	rs2112347	0.656	0.619
	rs11639988	0.849	0.858
	rs7184597	0.332	0.271
	rs1421085	0.418	0.460
	rs1558902	0.418	0.450
	rs10871777	0.232	0.270
	rs2568958	0.614	0.637
	rs12446554	0.849	0.867
	rs13078807	0.206	0.226
	rs10875976	0.456	0.451
	rs6731302	0.413	0.438
	rs2370983	0.640	0.704
	rs9568856	0.125	0.126
	rs9299	0.657	0.633
	rs2116830	0.802	0.770
	rs988712	0.762	0.754
	rs17817449	0.408	0.460
	rs17782313	0.231	0.265
	rs1424233	0.520	0.558
Total Mean Frequencies	N= 49	0.514	0.523
Obesity (early onset extreme)	rs476828	0.233	0.270
	rs12463617	0.828	0.881
	rs1993709	0.807	0.851
	rs1957894	0.078	0.083
	rs11208659	0.086	0.062
	rs564343	0.427	0.372
	rs11109072	0.043	0.028
	rs17700144	0.205	0.248
	rs1421085	0.418	0.460
	rs1558902	0.418	0.450
	rs10871777	0.232	0.270
Total Mean Frequencies	N= 11	0.343	0.361
Adiposity	rs987237	0.195	0.164
Obesity (extreme)	rs9941349	0.425	0.473
Type 1 diabetes	rs539514	0.493	0.483
	rs478222	0.602	0.627
	rs924043	0.858	0.894
	rs2304256	0.697	0.679
	rs941576	0.571	0.535
	rs11755527	0.486	0.429

	rs947474	0.169	0.155
	rs3825932	0.329	0.305
	rs6679677	0.092	0.115
	rs3087243	0.433	0.460
	rs17696736	0.441	0.358
	rs2292239	0.350	0.33
	rs12708716	0.354	0.319
	rs1701704	0.342	0.319
	rs2903692	0.659	0.686
	rs1004446	0.629	0.611
	rs11171739	0.430	0.407
	rs2542151	0.183	0.115
	rs2476601	0.093	0.117
	rs1990760	0.574	0.625
Total Mean Frequencies	N= 20	0.439	0.429
Type 1 diabetes autoantibodies	rs7528684	0.525	0.513
	rs3024505	0.847	0.819
	rs3184504	0.481	0.445
	rs763361	0.528	0.504
Total Mean Frequencies	N= 4	0.595	0.570
Type 1 diabetes nephropathy	rs12437854	0.070	0.017
Nephropathy (idiopathic membranous)	rs4664308	0.577	0.535
Nonalcoholic fatty liver disease	rs738409	0.201	0.233

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Type 2 diabetes	rs7903146	0.323	0.279
	rs243021	0.415	0.478
	rs4457053	0.317	0.259
	rs972283	0.503	0.55
	rs896854	0.485	0.436
	rs231362	0.501	0.518
	rs1552224	0.845	0.867
	rs1531343	0.094	0.115
	rs11634397	0.695	0.642
	rs8042680	0.288	0.261
	rs5945326	0.788	0.793
	rs7578326	0.636	0.650
	rs1387153	0.288	0.272
	rs1470579	0.335	0.296
	rs10440833	0.257	0.250
	rs849134	0.522	0.526
	rs10965250	0.823	0.805
	rs5015480	0.571	0.580
	rs8050136	0.407	0.460
	rs864745	0.481	0.487
	rs12779790	0.192	0.229
	rs7961581	0.260	0.252
	rs7578597	0.886	0.876
	rs6931514	0.252	0.279
	rs4506565	0.348	0.296
	rs7754840	0.315	0.336
	rs1111875	0.577	0.584
	rs4402960	0.336	0.296
	rs10811661	0.822	0.801
	rs7756992	0.260	0.279
	rs5215	0.333	0.398
	rs7901695	0.346	0.281
Total Mean Frequencies	N =32	0.453	0.451

Multiple SNPs in LD were included within the same mean calculations (Table 2.9).

Table 2.9 Metabolic Diseases – SNPs in LD

Disease	SNP	Proxy	r^2	*D'*
Obesity	rs6711012	rs10189761	1	1
	rs12446632	rs11639988	1	1
	rs3101336	rs2568958	1	1
	rs1421085	rs17817449	0.932	0.966
	rs10871777	rs17782313	0.979	1
	rs17817449	rs9941349	0.884	0.965
	rs476828	rs17782313	0.979	1
	rs17700144	rs476828	0.856	0.977
	rs17700144	rs17782313	0.834	0.954

	rs1421085	rs17817449	0.932	0.966
	rs10871777	rs476828	1	1
	rs10871777	rs17782313	0.979	1
	rs10871777	rs17700144	0.856	0.977
	rs9941349	rs1421085	0.884	0.965
T1 Diabetes	rs1701704	rs2292239	0.851	0.94
	rs2903692	rs12708716	0.941	0.98
	rs2476601	rs6679677	1	1
T2 Diabetes	rs4506565	rs7903146	0.921	1
	rs1111875	rs5015480	1	1
	rs4402960	rs1470579	1	1
	rs7756992	rs6931514	0.958	0.979
	rs7901695	rs4506565	0.96	1
	rs7901695	rs7903146	0.919	0.979

2.3.8 Eye Disease

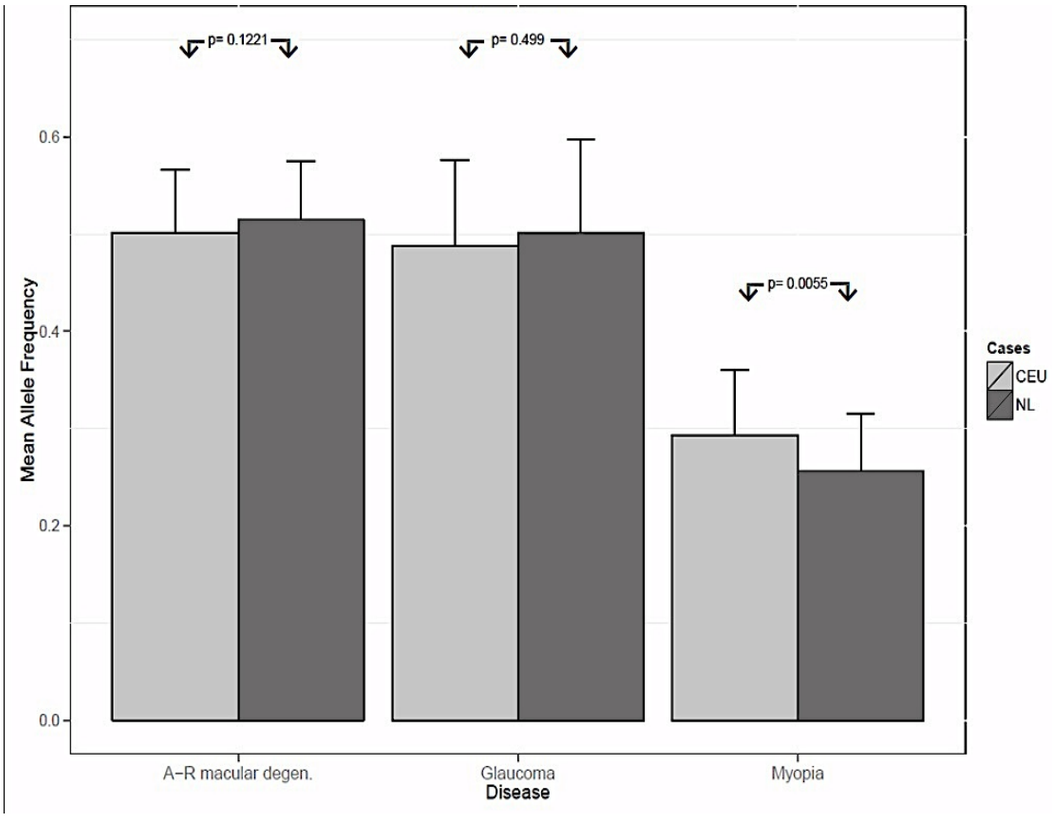


Figure 2.9 Mean risk allele frequencies of eye diseases.

"A-R macular degen" is age-related macular degeneration, glaucoma is all glaucoma SNPs combined, and myopia (pathological).

Table 2.10 SNPs associated with eye disease

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Age-related macular degeneration	rs1831282	0.440	0.416
	rs522162	0.902	0.938
	rs10490924	0.215	0.199
	rs3750848	0.217	0.204
	rs1329424	0.395	0.392
	rs3793917	0.219	0.216
	rs1061147	0.392	0.372
	rs541862	0.902	0.938
	rs10801555	0.392	0.372
	rs1061170	0.392	0.282
	rs641153	0.904	0.942
	rs2230199	0.195	0.175
	rs9332739	0.934	0.933
	rs10468017	0.720	0.690
	rs10033900	0.472	0.451

	rs3764261	0.342	0.345
	rs1329428	0.612	0.576
	rs10737680	0.613	0.58
Total Mean Frequencies	N =18	0.514	0.501
Glaucoma (primary open-angle)	rs2157719	0.605	0.558
	rs10483727	0.415	0.379
	rs284489	0.622	0.642
	rs4236601	0.278	0.281
Total Mean Frequencies	N= 4	0.480	0.465
Glaucoma	rs4656461	0.092	0.146
	rs4977756	0.652	0.584
Glaucoma (exfoliation)	rs3825942	0.844	0.828
Keratoconus	rs4954218	0.653	0.717
Fuchs's corneal dystrophy	rs613872	0.159	0.19
Myopia (pathological)	rs10747502	0.911	0.932
	rs7428796	0.621	0.798
	rs4839680	0.026	0.062
	rs1656966	0.171	0.153
	rs16872571	0.713	0.681
	rs6841898	0.028	0.014
	rs6857559	0.614	0.897
	rs13172324	0.066	0.115
	rs10053502	0.959	0.956
	rs999556	0.293	0.321
	rs12525668	0.072	0.093
	rs1302019	0.027	0.027
	rs10274279	0.114	0.106
	rs4737395	0.106	0.124
	rs12216812	0.084	0.111
	rs872863	0.088	0.085
	rs7077335	0.016	0.009
	rs1472750	0.091	0.130
	rs17788937	0.790	1.000
	rs7323755	0.035	0.053
	rs11838472	0.027	0.040
	rs11618212	0.220	0.248
	rs17822114	0.035	0.058
	rs8050940	0.026	0.018
	rs923375	0.138	0.177

	rs17826255	0.065	0.044
	rs3786800	0.778	0.902
	rs4142248	0.048	0.035
Total Mean Frequencies	N= 28	0.256	0.293

Multiple SNPs from the age-related macular degeneration category were included in the same calculation, rs1831282/ rs10801555 (r^2 = 0.821, D' = 1.0), rs1831282/ rs1061147 (r^2 = 0.821, D' = 1.0), rs522162/ rs541862 (r^2 = 1.0, D' = 1.0), rs10490924/ rs3750848 (r^2 = 0.974, D' = 1.0), rs1061147/ rs10801555 (r^2 = 1.0, D' = 1.0), rs1329428/ rs10737680 (r^2 = 1.0, D' = 1.0).

2.3.9 Lung Disease

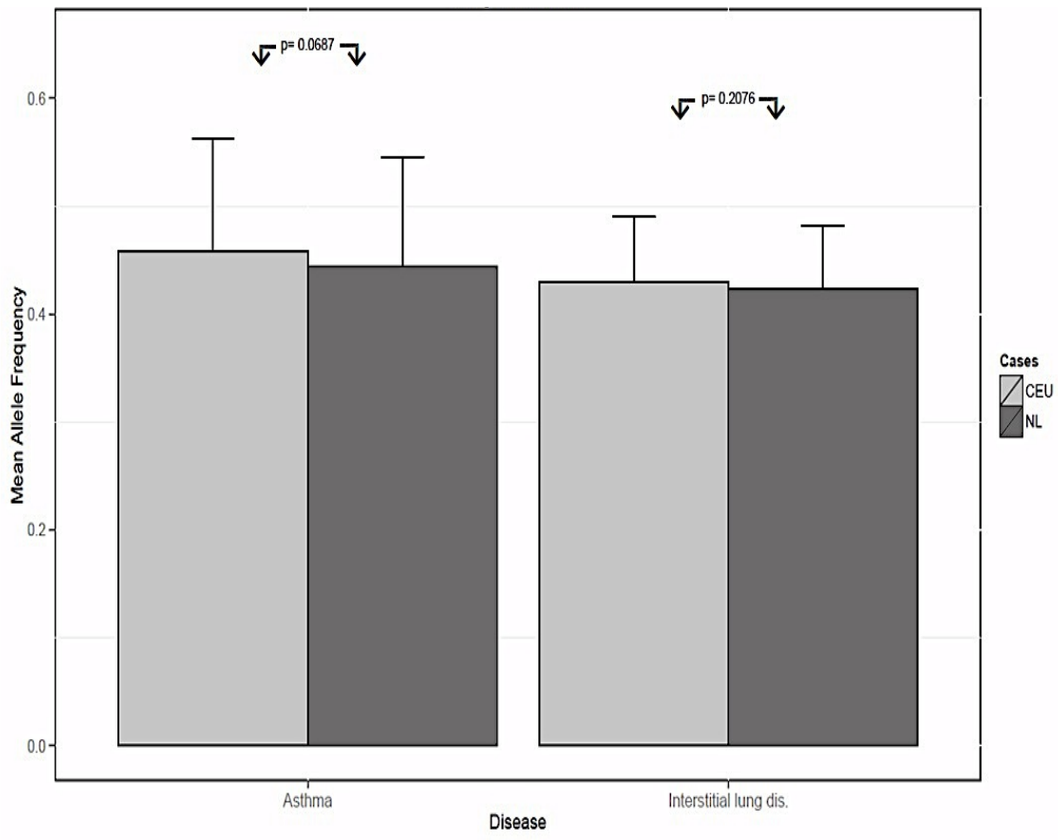


Figure 2.10 Mean risk allele frequencies of lung diseases.

Table 2.11 SNPs associated with lung disease

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Asthma (childhood, severe)	rs6967330	0.172	0.183
	rs2305480	0.556	0.536
	rs928413	0.259	0.286
	rs3894194	0.471	0.487
Total Mean Frequencies	N= 4	0.365	0.373
Asthma (childhood onset)	rs7927044	0.009	0.009
Asthma	rs13408661	0.860	0.907
	rs2786098	0.741	0.768
	rs7216389	0.487	0.491
Total Mean Frequencies	N= 3	0.696	0.722
Idiopathic pulmonary fibrosis	rs5743894	0.157	0.155

	rs5743890	0.858	0.845
Interstitial lung disease	rs6793295	0.267	0.277
	rs2609255	0.247	0.212
	rs2736100	0.478	0.473
	rs2076295	0.435	0.451
	rs7934606	0.413	0.463
	rs1278769	0.763	0.77
	rs2034650	0.477	0.478
	rs12610495	0.308	0.314
Total Mean Frequencies	N= 8	0.423	0.430
Chronic obstructive pulmonary disease	rs7937	0.551	0.553
	rs7671167	0.526	0.518
Lung adenocarcinoma	rs31489	0.559	0.607
	rs1051730	0.295	0.385
	rs2736100	0.478	0.473
Lung cancer	rs8034191	0.289	0.418
	rs4975616	0.558	0.58
	rs1051730	0.295	0.385
	rs401681	0.537	0.566
	rs8042374	0.251	0.201
	rs1051730	0.295	0.385
Total Mean Frequencies	N= 6	0.371	0.423

Two pairs of SNPs from lung cancer were in LD, rs4975616/rs401681 (r^2 = 0.882, D' = 0.964) and rs1051730/rs8034191 (r^2 = 0.899, D' = 1.0).

2.3.10 Lymphomas, myelomas and leukemias

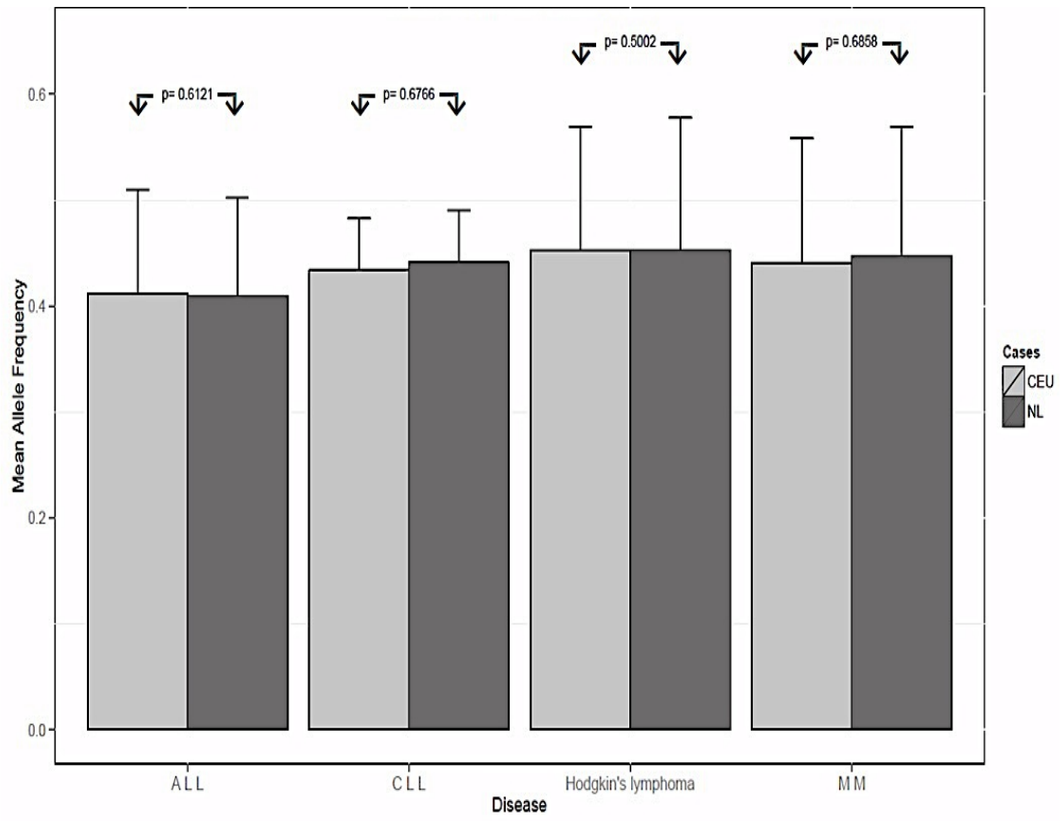


Figure 2.11 Mean risk allele frequencies of lymphomas, myelomas and leukemias.

"ALL" is acute lymphoblastic leukemia, "CLL" is chronic lymphocytic leukemia, Hodgkin's lymphoma, and " MM" is multiple myelomas.

Table 2.12 SNPs associated with lymphomas, myelomas and leukemias

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Acute lymphoblastic leukemia (childhood)	rs17505102	0.847	0.910
	rs4132601	0.263403	0.304
	rs7089424	0.339161	0.302
	rs10821936	0.337995	0.310
	rs11978267	0.265252	0.304
Total Mean Frequencies	N= 5	0.411	0.426
Acute lymphoblastic leukemia (B-cell precursor)	rs10828317	0.652	0.612
	rs3824662	0.163	0.143

Myeloproliferative neoplasms	rs10974944	0.269	0.305

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Chronic lymphocytic leukemia	rs10936599	0.752	0.732
	rs2236256	0.432	0.446
	rs1439287	0.480	0.473
	rs13397985	0.180	0.207
	rs872071	0.515	0.513
	rs735665	0.178	0.174
	rs7176508	0.415	0.371
	rs1044873	0.632	0.612
	rs4406737	0.554	0.571
	rs4987855	0.897	0.883
	rs7944004	0.492	0.491
	rs898518	0.586	0.612
	rs3769825	0.530	0.513
	rs13401811	0.856	0.775
	rs17483466	0.205	0.190
	rs674313	0.245	0.138
	rs9273012	0.254	0.161
	rs11636802	0.094	0.150
	rs7169431	0.079	0.103
	rs305061	0.667	0.683
	rs391023	0.626	0.681
	rs4987852	0.076	0.080
	rs210142	0.733	0.721
	rs391525	0.329	0.308
	rs11083846	0.300	0.257
Total Mean Frequencies	N= 25	0.441	0.434
Hodgkin's lymphoma	rs3806624	0.441	0.442
	rs7745098	0.536	0.544
	rs501764	0.173	0.194
	rs2019960	0.240	0.243
Total Mean Frequencies	N= 4	0.348	0.358
Nodular sclerosis Hodgkin lymphoma	rs2858870	0.875	0.841
Follicular lymphoma	rs735665	0.178	0.173
Meningioma	rs11012732	0.672	0.673

Ewing sarcoma	rs9430161	0.780	0.792
	rs224278	0.591	0.588
	rs4924410	0.258	0.314
Multiple myeloma	rs10936599	0.752	0.730
	rs2285803	0.267	0.274
	rs4273077	0.094	0.113
	rs877529	0.452	0.416
	rs4487645	0.670	0.673
Total Mean Frequencies	N= 5	0.447	0.441
Multiple myeloma (hyperdiploidy)	rs6599175	0.166	0.243
	rs139371	0.527	0.544
	rs2272007	0.169	0.243
	rs4487645	0.670	0.673
Total Mean Frequencies	N= 4	0.383	0.426
Amyotrophic lateral sclerosis	rs12608932	0.339	0.345
	rs2814707	0.250	0.239
Myasthenia gravis	rs2476601	0.093	0.117
	rs4958881	0.119	0.075

Two SNPs from acute lymphoblastic leukemia (childhood) were in LD, rs4132601/rs11978267 (r^2 = 1.0, D' = 1.0). Also, two SNPs from multiple myeloma (hyperdiploidy) were in perfect LD (r^2 = 1.0, D' = 1.0).

2.3.11 Thyroid Disease

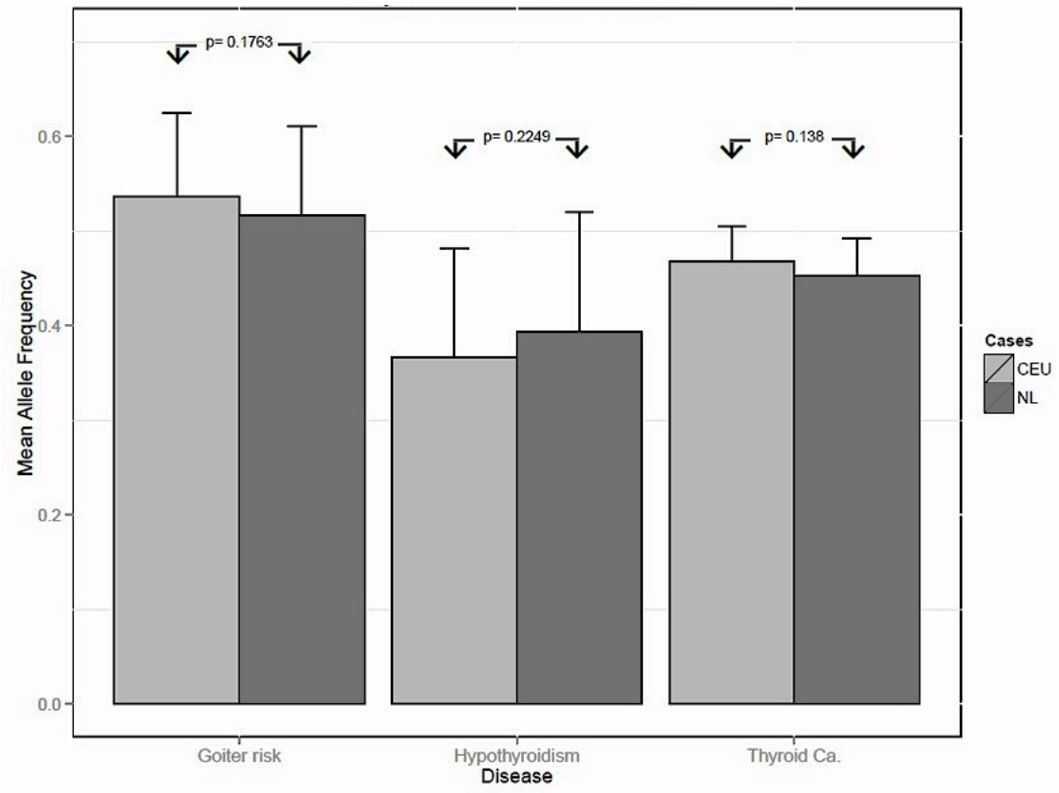


Figure 2.12 Mean risk allele frequencies of thyroid disease.

Table 2.13 SNPs associated with thyroid disease

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Hypo-thyroidism	rs925489	0.652	0.567
	rs4915077	0.087	0.072
	rs3184504	0.481	0.445
	rs6679677	0.092	0.115
	rs7850258	0.652	0.633
Total Mean Frequencies	N= 5	0.393	0.366
Goiter risk	rs12138950	0.836	0.814
	rs4338740	0.745	0.746
	rs17767419	0.323	0.367
	rs10917468	0.227	0.270
	rs4338740	0.255	0.254
	rs3813579	0.547	0.603
	rs12045440	0.683	0.699
Total Mean Frequencies	N= 7	0.517	0.536
Thyroid cancer	rs6759952	0.412	0.447
	rs965513	0.349	0.367

	rs966423	0.429	0.456
	rs2439302	0.484	0.465
	rs944289	0.587	0.602
Total Mean Frequencies	N= 5	0.452	0.467

No SNPs in LD were included in the same mean frequency calculations.

2.3.12 Skin Disease

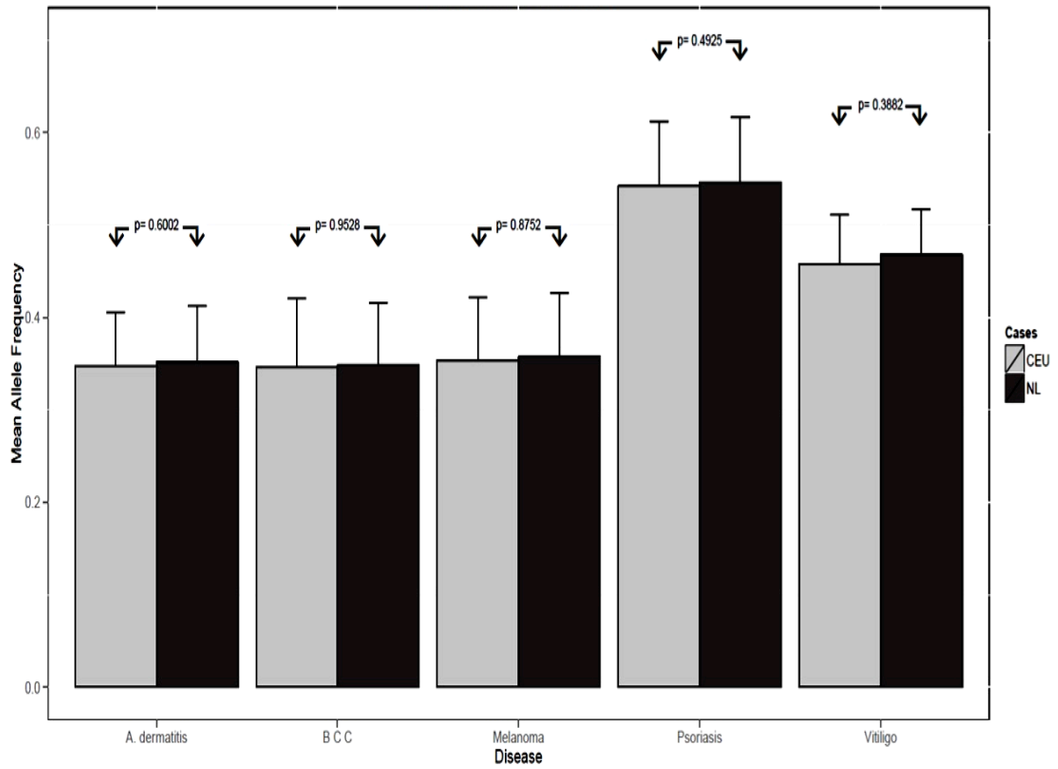


Figure 2.13 Mean risk allele frequencies of skin conditions and cancers.

"A. dermatitis" is atopic dermatitis, "BCC" is basal cell carcinoma, melanoma, psoriasis and vitiligo

Table 2.14 SNPs associated with skin disease

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Atopic dermatitis	rs1295686	0.211	0.217
	rs7130588	0.359	0.381
	rs6661961	0.423	0.412
	rs479844	0.567	0.535
	rs2164983	0.161	0.147
	rs7927894	0.391	0.394
Total Mean Frequencies	N= 6	0.352	0.348
Alopecia areata	rs1024161	0.412	0.403
	rs9479482	0.564	0.504
	rs3118470	0.331	0.310
Basal cell carcinoma	rs214782	0.214	0.144
	rs7006527	0.163	0.128
	rs7538876	0.378	0.427
	rs801114	0.357	0.342

	rs401681	0.537	0.571
	rs2151280	0.516	0.500
	rs157935	0.689	0.727
	rs1805007	0.088	0.122
	rs12210050	0.190	0.152
Total Mean Frequencies	N= 9	0.348	0.346
Non-melanoma skin cancer	rs12203592	0.212	0.156
	rs1805007	0.088	0.122
Melanoma	rs16953002	0.149	0.143
	rs13016963	0.409	0.420
	rs1801516	0.835	0.811
	rs45430	0.647	0.670
	rs7023329	0.522	0.509
	rs1393350	0.276	0.219
	rs258322	0.111	0.138
	rs7412746	0.587	0.527
	rs17119461	0.011	0.018
	rs1393350	0.276	0.219
	rs258322	0.111	0.138
	rs4785763	0.308	0.380
	rs2284063	0.660	0.679
	rs910873	0.103	0.080
Total Mean Frequencies	N= 14	0.357	0.354
Psoriasis	rs2546890	0.521	0.562
	rs702873	0.564	0.571
	rs17716942	0.838	0.805
	rs27524	0.385	0.332
	rs240993	0.233	0.250
	rs8016947	0.558	0.535
	rs12720356	0.892	0.876
	rs458017	0.065	0.045
	rs465969	0.922	0.942
	rs280519	0.485	0.487
	rs4112788	0.679	0.597
	rs3213094	0.198	0.190
	rs2082412	0.802	0.805
	rs17728338	0.074	0.080
	rs20541	0.795	0.783
	rs610604	0.328	0.425
	rs2066808	0.934	0.929
Total Mean Frequencies	N= 17	0.545	0.542
Vitiligo	rs2111485	0.569	0.615
	rs10768122	0.410	0.385
	rs4409785	0.177	0.142

	rs2456973	0.343	0.297
	rs4766578	0.468	0.422
	rs4822024	0.786	0.758
	rs4908760	0.372	0.376
	rs706779	0.545	0.541
	rs1393350	0.724	0.774
	rs11203203	0.360	0.323
	rs229527	0.391	0.403
	rs1464510	0.470	0.456
Total Mean Frequencies	N= 12	0.468	0.458

Two SNPs from atopic dermatitis were in LD, rs7130588/rs7927894 (r^2 = 0.948, D' = 1.0).

2.3.13 Aneurysms, embolisms and stroke

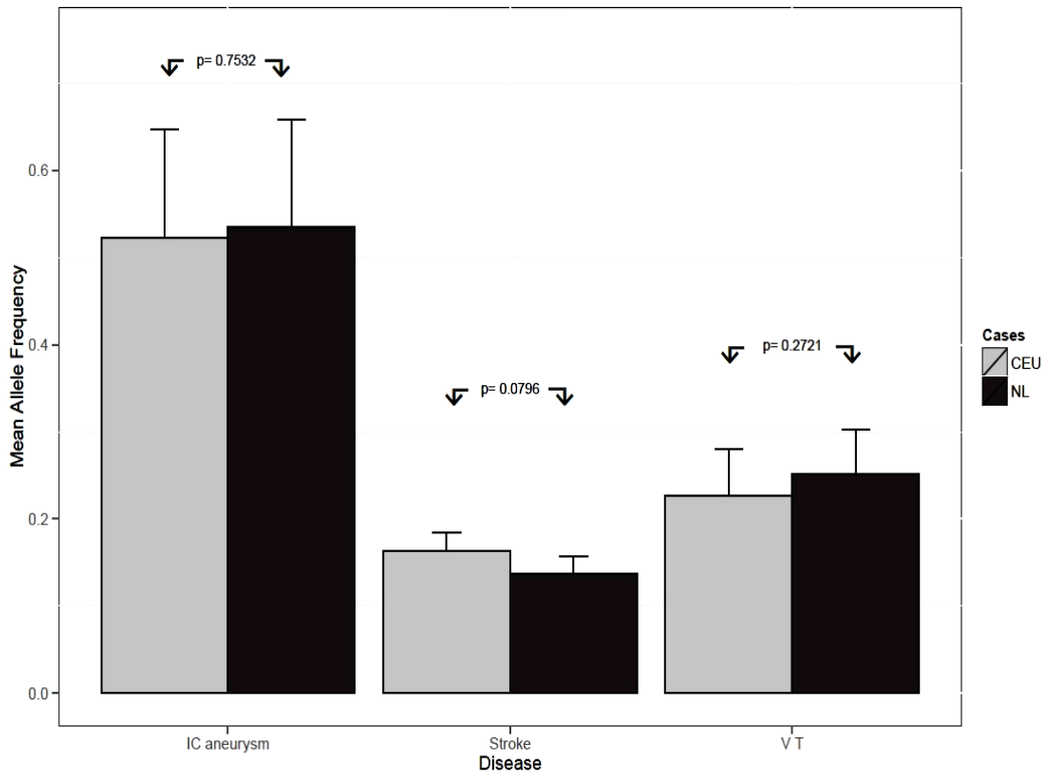


Figure 2.14 Mean risk allele frequencies of aneurysms, embolisms and stroke.

"IC aneurysms" is intracranial aneurysms, all stroke-associate SNPs, and "V T" is venous thrombosis.

Table 2.15 SNPs associated with aneurysms, embolisms and stroke

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Intracranial aneurysm	rs9298506	0.783	0.795
	rs1333040	0.645	0.558
	rs12413409	0.916	0.925
	rs9315204	0.219	0.212
	rs11661542	0.474	0.487
	rs10958409	0.173	0.159
Total Mean Frequencies	N= 6	0.535	0.523
Abdominal aortic aneurysm	rs6511720	0.887	0.894
	rs1466535	0.669	0.686
	rs7025486	0.242	0.248
Thoracic aortic aneurysms	rs1036476	0.119	0.075

Venous thrombosis	rs3813948	0.057	0.107
Venous thrombo-embolism	rs6427196	0.064	0.062
	rs687621	0.350	0.356
	rs4253399	0.418	0.394
	rs6536024	0.556	0.562
	rs6025	0.025	0.022
	rs495828	0.232	0.186
	rs16861990	0.064	0.080
	rs2519093	0.205	0.000
	rs1018827	0.064	0.076
	rs7659024	0.258	0.221
	rs505922	0.350	0.363
	rs3756008	0.441	0.398
Total Mean Frequencies	N= 12	0.252	0.227
Stroke (ischemic)	rs2107595	0.138	0.200
	rs6843082	0.175	0.212
	rs2200733	0.100	0.115
Stroke	rs11984041	0.088	0.112
	rs12425791	0.185	0.177
Total Mean Frequencies	N= 5	0.137	0.163

Two SNPs in LD were included in the calculation for venous thromboembolism, rs4253399/rs3756008 (r^2 = 0.947, D' = 0.982).

2.3.14 Heart and Circulatory Conditions

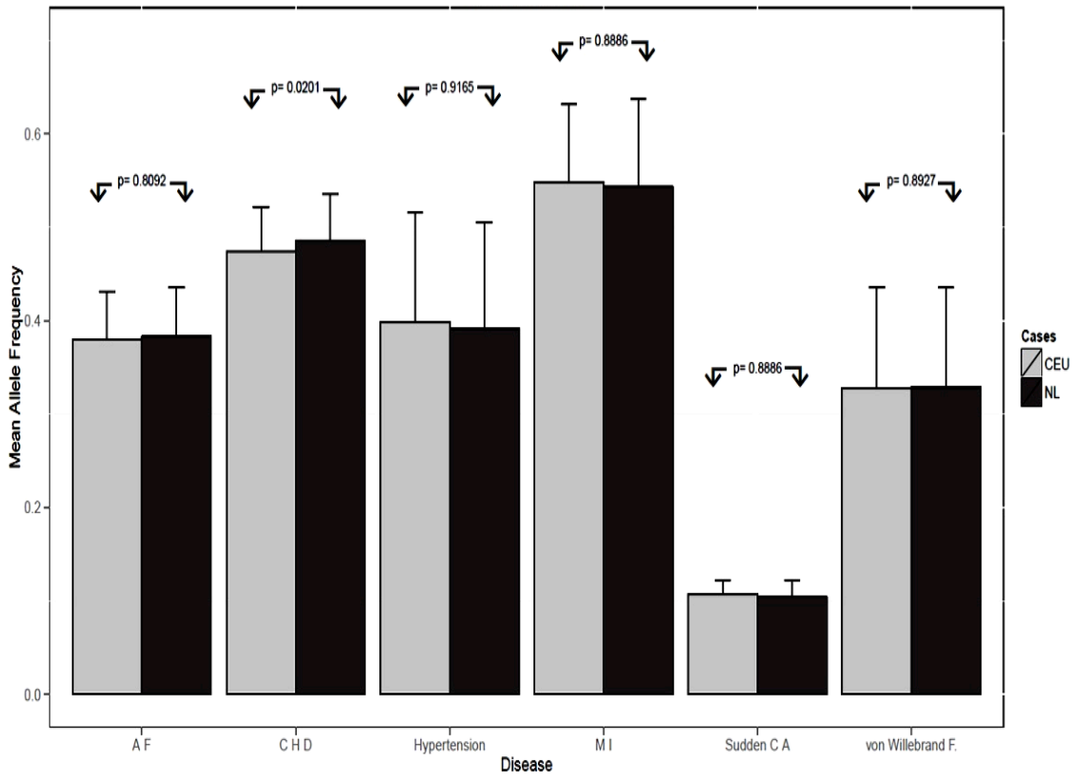


Figure 2.15 Mean risk allele frequencies of heart and circulatory conditions.

"AF" is atrial fibrillation, "CHD" is coronary heart disease, hypertension, "MI" is myocardial infarction, "sudden CA" is sudden cardiac arrest and von Willebrand factor.

Table 2.16 SNPs associated with heart and circulatory conditions

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Brugada syndrome	rs11708996	0.147	0.200
	rs10428132	0.456	0.420
	rs9388451	0.489	0.465
Congenital heart disease	rs870142	0.229	0.265
Tetralogy of Fallot	rs7982677	0.262	0.259
	rs11065987	0.436	0.336
von Willebrand factor	rs687289	0.345	0.358
	rs8176704	0.075	0.088
	rs1063856	0.655	0.655
	rs867186	0.127	0.097
	rs9390459	0.444	0.442
Total Mean Frequencies	N= 5	0.329	0.328

Factor VII	rs561241	0.894	0.883
Aortic-valve calcification	rs10455872	0.084	0.083
Pulmonary arterial hypertension	rs2217560	0.084	0.111
Coronary artery calcification	rs1537370	0.460	0.456
	rs1333049	0.470	0.456
	rs9349379	0.633	0.628
Hypertri-glyceridemia	rs964184	0.143	0.121
	rs1260326	0.422	0.420
Kawasaki disease	rs1801274	0.470	0.491
Coronary heart disease	rs1333049	0.470	0.451
	rs7865618	0.606	0.549
	rs1746048	0.893	0.853
	rs17114036	0.893	0.895
	rs12190287	0.636	0.603
	rs11556924	0.617	0.612
	rs579459	0.232	0.205
	rs12413409	0.916	0.924
	rs964184	0.143	0.122
	rs4773144	0.409	0.424
	rs2895811	0.406	0.419
	rs3825807	0.564	0.550
	rs216172	0.365	0.328
	rs12936587	0.479	0.467
	rs599839	0.802	0.719
	rs6725887	0.125	0.156
	rs12526453	0.653	0.634
	rs3798220	0.029	0.000
	rs4977574	0.463	0.455
	rs1122608	0.745	0.746
	rs9982601	0.146	0.206
	rs1231206	0.363	0.325
	rs1994016	0.587	0.572
	rs514659	0.351	0.364
	rs3739998	0.572	0.563
	rs9818870	0.174	0.170
Total Mean Frequencies	N= 26	0.486	0.474

Myocardial infarction (early onset)	rs646776	0.194	0.257
	rs17465637	0.719	0.734
	rs1746048	0.893	0.853
	rs12526453	0.653	0.634
	rs4977574	0.463	0.455
	rs1122608	0.745	0.746
	rs9982601	0.146	0.206
Total Mean Frequencies	N= 7	0.545	0.555
Myocardial infarction	rs10757278	0.533	0.500
Dilated cardio-myopathy	rs2234962	0.760	0.792
	rs10927875	0.708	0.683
Sudden cardiac arrest	rs4665058	0.016	0.025
	rs16866933	0.077	0.076
	rs12429889	0.119	0.134
	rs7307780	0.183	0.165
	rs10765792	0.090	0.121
	rs12189362	0.134	0.124
	rs2982694	0.136	0.121
	rs16942421	0.080	0.091
Total Mean Frequencies	N= 8	0.104	0.107
Atrial fibrillation	rs2200733	0.100	0.116
	rs10033464	0.075	0.094
	rs3903239	0.449	0.448
	rs3807989	0.589	0.580
	rs10821415	0.450	0.397
	rs10824026	0.862	0.835
	rs1152591	0.485	0.487
	rs7164883	0.159	0.170
	rs6666258	0.328	0.298
	rs6817105	0.100	0.119
	rs2106261	0.186	0.179
	rs6843082	0.175	0.210
	rs13376333	0.328	0.339
	rs7193343	0.174	0.161
	rs1805126	0.626	0.679
	rs6906287	0.455	0.500
	rs1321313	0.770	0.695
	rs2207790	0.519	0.500
	rs6795970	0.448	0.420

Total Mean Frequencies	N =19	0.318	0.317
Hypertension	rs6015450	0.128	0.071
	rs633185	0.261	0.317
	rs1799945	0.136	0.182
	rs1173771	0.595	0.513
	rs932764	0.400	0.425
	rs2681472	0.831	0.884
Total Mean Frequencies	N= 6	0.392	0.399

For the mean risk allele frequency calculation for coronary heart disease two SNPs were in LD, rs4977574/ rs1333049 (r^2 = 0.9, D' = 0.9).

2.3.15 Inflammatory Conditions

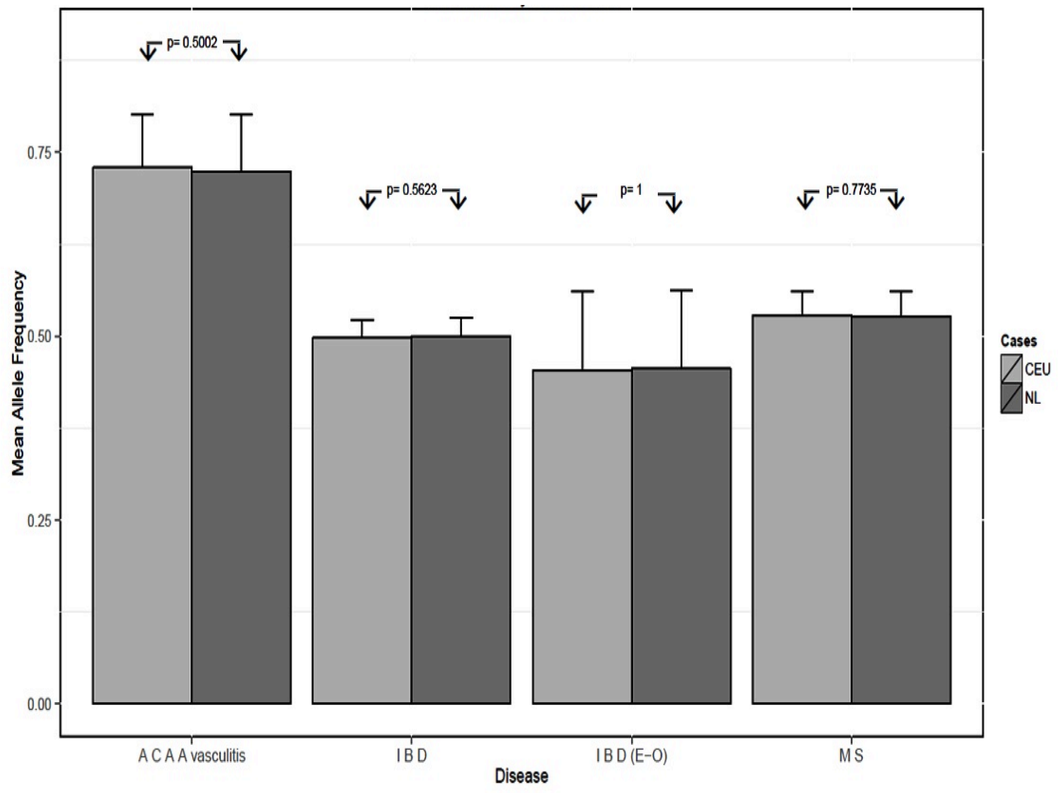


Figure 2.16 Mean risk allele frequencies of inflammatory conditions.

"ACAA" is antineutrophil cytoplasmic antibody-associated vasculitis, "IBD" is irritable bowel syndrome, "IBD (E-O)" is irritable bowel syndrome (early-onset) and "MS" is multiple sclerosis.

Table 2.17 SNPs associated with inflammatory conditions

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Antineutrophil cytoplasmic antibody-associated vasculitis	rs6638512	0.501	0.536
	rs1972809	0.601	0.62
	rs2755459	0.768	0.713
	rs7151526	0.937	0.938
	rs5954596	0.816	0.839
Total Mean Frequencies	N= 5	0.724	0.729
Inflammatory bowel disease	rs35675666	0.803	0.833
	rs12568930	0.845	0.814
	rs11209026	0.940	0.959

rs4845604	0.831	0.867
rs670523	0.358	0.363
rs4656958	0.702	0.646
rs1801274	0.470	0.491
rs2488389	0.213	0.207
rs7554511	0.708	0.673
rs3024505	0.153	0.181
rs6545800	0.437	0.416
rs925255	0.566	0.527
rs10495903	0.141	0.146
rs7608910	0.414	0.379
rs917997	0.237	0.204
rs1517352	0.591	0.624
rs2382817	0.389	0.345
rs3749171	0.163	0.210
rs4256159	0.156	0.164
rs3197999	0.258	0.261
rs7657746	0.770	0.790
rs11742570	0.628	0.646
rs1363907	0.429	0.442
rs4836519	0.770	0.730
rs2188962	0.433	0.402
rs6863411	0.629	0.633
rs11741861	0.069	0.044
rs6871626	0.330	0.367
rs17119	0.825	0.770
rs9358372	0.388	0.350
rs1847472	0.625	0.692
rs6568421	0.255	0.310
rs3851228	0.058	0.040
rs6920220	0.248	0.165
rs1819333	0.494	0.491
rs1456896	0.678	0.646
rs9297145	0.241	0.305
rs1734907	0.117	0.162
rs921720	0.627	0.619
rs1991866	0.430	0.447
rs10758669	0.385	0.366
rs4743820	0.706	0.699
rs4246905	0.738	0.676
rs10781499	0.455	0.487
rs12722515	0.844	0.850
rs1042058	0.604	0.628
rs11010067	0.325	0.354
rs2790216	0.800	0.779
rs10761659	0.537	0.553
rs2227564	0.759	0.761
rs1250546	0.561	0.558
rs6586030	0.826	0.823
rs4409764	0.497	0.455
rs907611	0.294	0.319
rs10896794	0.798	0.735

	rs11230563	0.644	0.637
	rs4246215	0.323	0.372
	rs559928	0.816	0.796
	rs2231884	0.147	0.173
	rs2155219	0.499	0.487
	rs630923	0.838	0.863
	rs11564258	0.023	0.027
	rs11168249	0.526	0.504
	rs7134599	0.420	0.394
	rs17085007	0.188	0.150
	rs941823	0.746	0.779
	rs9557195	0.791	0.774
	rs194749	0.227	0.235
	rs8005161	0.085	0.140
	rs17293632	0.216	0.230
	rs7495132	0.897	0.876
	rs529866	0.781	0.805
	rs7404095	0.570	0.637
	rs26528	0.464	0.451
	rs10521318	0.910	0.907
	rs3091316	0.722	0.686
	rs12946510	0.440	0.500
	rs12942547	0.633	0.545
	rs1292053	0.439	0.451
	rs1893217	0.183	0.115
	rs7240004	0.641	0.626
	rs727088	0.528	0.500
	rs11879191	0.801	0.801
	rs17694108	0.267	0.305
	rs11672983	0.381	0.407
	rs6142618	0.515	0.535
	rs4911259	0.413	0.345
	rs1569723	0.251	0.239
	rs259964	0.441	0.464
	rs6062504	0.725	0.655
	rs2823286	0.684	0.681
	rs2836878	0.712	0.712
	rs7282490	0.399	0.429
	rs2266959	0.204	0.168
	rs2412970	0.456	0.385
	rs2413583	0.851	0.854
	rs2315008	0.723	0.659
	rs5743289	0.196	0.226
	rs2076756	0.263	0.323
Total Mean Frequencies	N= 99	0.500	0.498
Inflammatory bowel disease (early onset)	rs10500264	0.161	0.179
	rs2412973	0.439	0.385
	rs8049439	0.583	0.615
	rs1250550	0.640	0.633

Total Mean Frequencies	N= 4	0.456	0.453

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Multiple sclerosis	rs9292777	0.630	0.649
	rs2300747	0.896	0.867
	rs12368653	0.500	0.447
	rs11154801	0.371	0.397
	rs7522462	0.708	0.673
	rs2293370	0.788	0.850
	rs650258	0.628	0.638
	rs1335532	0.892	0.863
	rs9282641	0.936	0.902
	rs7200786	0.580	0.540
	rs2248359	0.639	0.588
	rs2303759	0.247	0.296
	rs11129295	0.356	0.363
	rs11810217	0.256	0.257
	rs2119704	0.913	0.932
	rs7923837	0.593	0.633
	rs2546890	0.521	0.562
	rs17066096	0.252	0.181
	rs7090512	0.312	0.314
	rs7238078	0.780	0.796
	rs2283792	0.535	0.527
	rs4648356	0.727	0.668
	rs4410871	0.719	0.712
	rs12466022	0.746	0.748
	rs669607	0.491	0.487
	rs2425752	0.267	0.252
	rs771767	0.283	0.350
	rs7595037	0.532	0.549
	rs4613763	0.128	0.168
	rs802734	0.690	0.708
	rs2019960	0.240	0.243
	rs11581062	0.286	0.292
	rs10201872	0.156	0.196
	rs9891119	0.329	0.389
	rs1738074	0.584	0.535
	rs1800693	0.387	0.482
	rs1077667	0.774	0.786
	rs4902647	0.531	0.562
	rs1250550	0.360	0.367
	rs354033	0.776	0.765
	rs3118470	0.331	0.310
	rs703842	0.686	0.664
	rs17445836	0.788	0.733
	rs17824933	0.231	0.195
	rs10492972	0.314	0.336
Total Mean Frequencies	N= 45	0.526	0.528

Sarcoidosis	rs479777	0.693	0.633
Primary sclerosing cholangitis	rs3197999	0.258	0.261
Chronic hepatitis C infection	rs8099917	0.205	0.150
Pancreatitis	rs10273639	0.570	0.589
	rs12688220	0.268	0.333
Periodontitis	rs1537415	0.576	0.571

Two SNPs for inflammatory bowel disease are in LD, rs6062504/rs2315008 (r^2 = 0.981, D' = 1.0). The SNPs in LD for multiple sclerosis calculations were rs2300747/rs1335532 (r^2 = 0.964, D' = 1.0).

2.3.16 Brain Tumours and Others

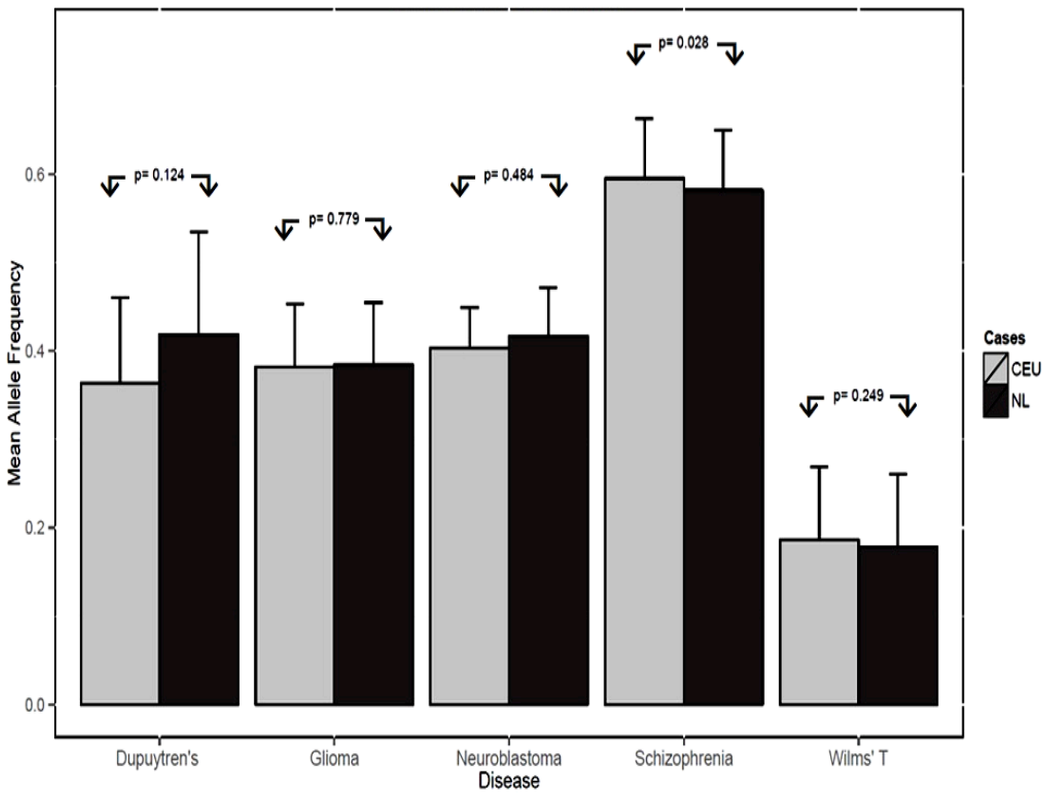


Figure 2.17 Mean risk allele frequencies of brain tumours and other conditions.

"Dupuytren's" is Dupuytren's contracture, "Glioma" is all SNPs associated with forms of gliomas, Schizophrenia and "Wilms' T" is Wilms' tumour.

Table 2.18 SNPs associated with brain tumours and other conditions

Disease	SNP ID	Risk Allele Frequency NL population	Risk Allele Frequency CEU population
Dupuytren's disease	rs7524102	0.155	0.186
	rs16879765	0.115	0.045
	rs611744	0.885	0.554
	rs2912522	0.758	0.739
	rs10809650	0.740	0.705
	rs11672517	0.228	0.221
	rs8124695	0.080	0.097
	rs6519955	0.390	0.363
Total Mean Frequencies	N= 8	0.419	0.364
Glioma	rs2736100	0.522	0.527
	rs6010620	0.797	0.757
	rs4295627	0.156	0.128
	rs2157719	0.395	0.442
	rs498872	0.332	0.310
	rs2853676	0.295	0.286
	rs891835	0.229	0.192

	rs4977756	0.349	0.416
Total Mean Frequencies	N= 8	0.384	0.382
Glioma (high-grade)	rs1412829	0.383	0.427
	rs4809324	0.107	0.088
	rs6010620	0.797	0.757
Neuro-blastoma	rs4336470	0.635	0.625
	rs7587476	0.251	0.265
	rs9295536	0.472	0.398
	rs110419	0.515	0.504
	rs4712653	0.492	0.425
	rs3768716	0.228	0.243
	rs6939340	0.500	0.460
Total Mean Frequencies	N= 7	0.442	0.417
Neuro-blastoma (high-risk)	rs6435862	0.240	0.304
Schizo-phrenia	rs7085104	0.655	0.688
	rs6461049	0.530	0.580
	rs1198588	0.797	0.808
	rs4129585	0.441	0.393
	rs10789369	0.382	0.406
	rs7940866	0.506	0.554
	rs17504622	0.029	0.033
	rs2905424	0.349	0.362
	rs2373000	0.418	0.428
	rs1261117	0.966	0.955
	rs7527939	0.837	0.795
	rs11191580	0.912	0.924
	rs1625579	0.821	0.833
	rs6932590	0.742	0.799
	rs12807809	0.835	0.826
	rs9960767	0.052	0.067
	rs1006737	0.338	0.347
	rs17693963	0.868	0.920
Total Mean Frequencies	N= 18	0.582	0.595
Bipolar disorder	rs1064395	0.168	0.152
	rs420259	0.723	0.804
	rs10994336	0.069	0.067
Major depressive	rs1545843	0.475	0.478

disorder			
Autism	rs4307059	0.615	0.628
Wilms tumour	rs3755132	0.130	0.150
	rs1027643	0.054	0.075
	rs790356	0.516	0.527
	rs2283873	0.018	0.031
	rs5955543	0.022	0.006
	rs807624	0.330	0.326
Total Mean Frequencies	N= 6	0.178	0.186
Epilepsy (generalized)	rs13026414	0.578	0.650
	rs10496964	0.799	0.801
Hippo-campal atrophy	rs6703865	0.073	0.034
Cystic fibrosis severity	rs12793173	0.213	0.257
Essential tremor	rs9652490	0.228	0.199
Narcolepsy	rs1154155	0.158	0.146

For the neuroblastoma calculations two pairs of SNPs were in LD, rs7587476/rs3768716 (r^2 = 0.896, *D'* = 1.0) and rs9295536/ rs4712653 (r^2 = 0.835, *D'* = 0.963).

2.3.17 Summary of Statistically Significant Results

Several diseases showed statistically significant results including bladder cancer, coronary heart disease and celiac disease. All three had $p < 0.05$, and ulcerative colitis had a near significant result at $p = 0.06$. The mean risk allele frequencies for these diseases were all higher in Newfoundland than in the CEU population (see figure 2.18).

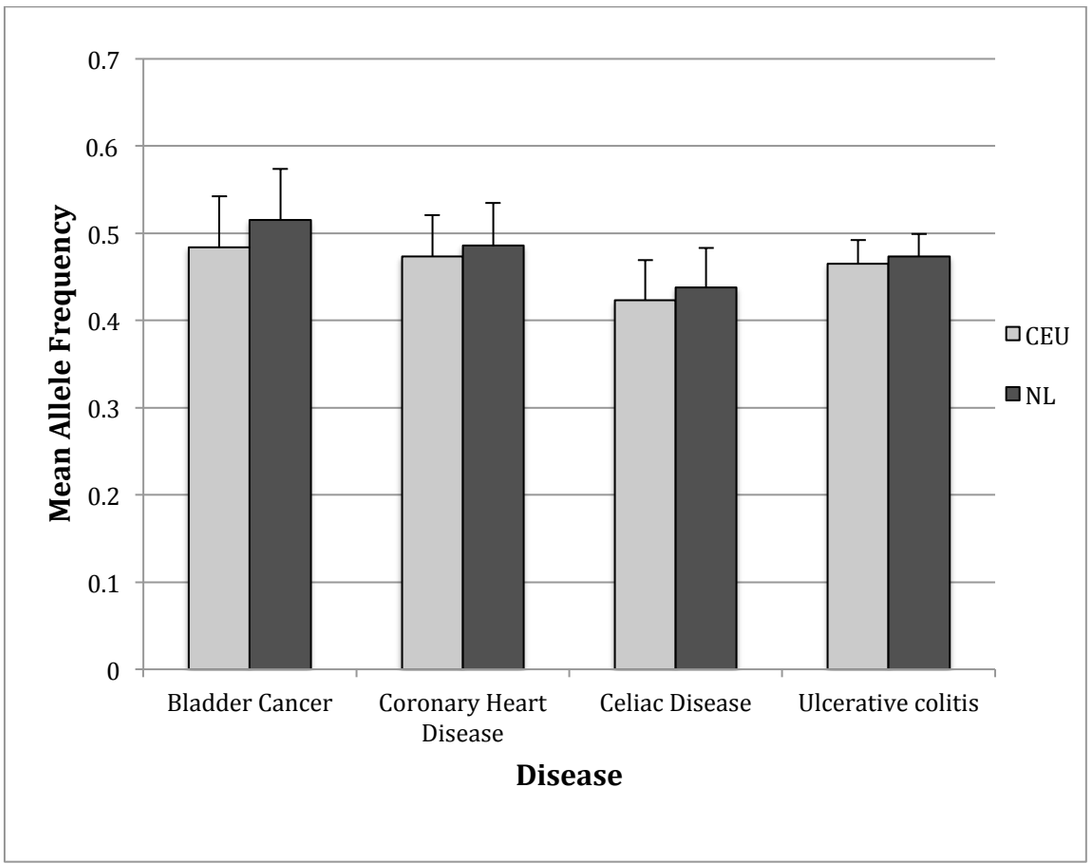


Figure 2.18 Significant Results - mean risk allele frequencies more prevalent in Newfoundland

Several diseases also showed significantly lower mean risk allele frequencies ($p < 0.05$) in Newfoundland as compared to the outbred population, including myopia, obesity and schizophrenia (see figure 2.19).

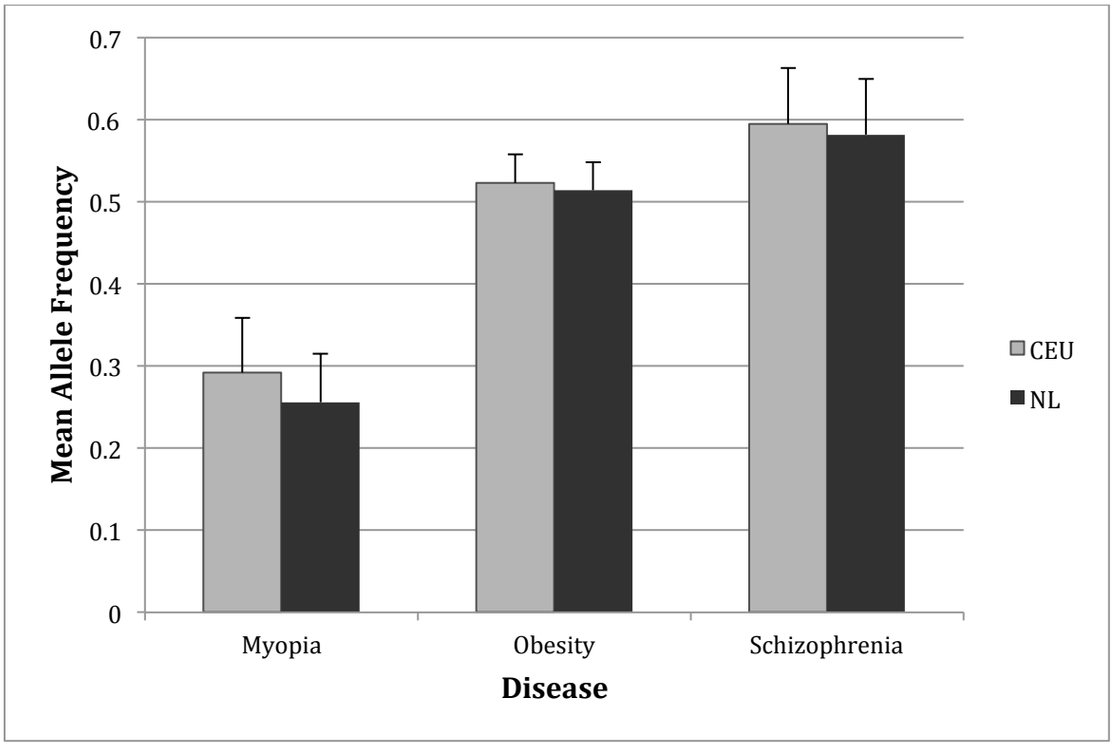


Figure 2.19 Significant Results - mean risk allele frequencies less prevalent in Newfoundland

2.4 CHAPTER 4. Discussion

Based on a current hypothesis of the evolution of complex disease, that complex diseases have developed due to an accumulation of multiple, low-penetrant polymorphisms, this study assesses the mean frequencies of the known risk alleles associated with disease in the Newfoundland and Labrador population. Many variants associated with complex diseases have already been discovered through GWA studies, with over 15,000 currently in the NIH catalogue, making it a concise resource for hundreds of SNPs associated with many different diseases. However, even when factoring in all risk alleles for a particular complex disease, the combined contribution of identified loci to disease variation is often less than 10% (Bjorkegren et al., 2015). By combining SNPs for a given disease the statistical power is increased and a broader picture of the burden of disease in a given population may be achieved. For example, a study of the complex trait of height utilized a "polygenic" model to combine all height-associated SNPs and explained up to 40% of genetic variance (Yang et al., 2010). This approach can be expanded to estimate the genetic contribution of risk alleles for common complex diseases. The unique genetic structure of founder populations could result in higher rates of certain diseases due to higher frequencies of risk alleles, making them particularly useful populations to study the genetic basis of complex disease. This has driven studies of monogenic disorders, rare in out-bred populations, but present at a higher rate in genetically isolated populations due to founder effects. These populations may also exhibit extended linkage disequilibrium than that found in out-bred populations. If proven, this would mean that fewer SNPs are required for studying genetic association. In simulations of recent bottleneck populations, natural selection has not had sufficient time to decrease the proportion of risk alleles to the equilibrium found in the out-bred populations (Lohmueller et al., 2008). Newfoundland has been identified as a founder population where a number of diseases are reported at higher rates and therefore, it can be assumed that there is a higher frequency of certain risk variants for complex traits. These risk variants likely rose in frequency during the population bottleneck Newfoundland

experienced during its settlement and may have spread through the population due to the low genetic diversity and higher rate of inbreeding subsequent to the bottleneck. By calculating the mean frequencies of risk alleles associated with a given disease in the Newfoundland population we identified several diseases with a significantly higher mean frequency of risk alleles in the NL population compared to the CEU sample. This is a novel method to study potential disease burden, as there are few reported studies using this method. One article by Lim and colleagues studied low-frequency loss-of-function variants. They were able to successfully identify an enrichment of risk variants and several health-related consequences in the Finnish population (Lim et al., 2014). Our study uses a similar method, involving common variants to assess disease allele burden in the Newfoundland and Labrador founder population as compared to the out-bred CEU population of similar ancestry. The Canadian prevalence of disease is also discussed as a reference to the NL disease prevalence. A comparison is drawn between the statistical prevalence and mean disease-associated SNP prevalence. The Canadian population was chosen due to its close geographic proximity to Newfoundland, similar environments and ethnic distribution in an attempt to compare disease burden in a founder population (NL) to an outbred population (Canada). Due to the high number of diseases included in the study, only those with significant or unexpected results will be discussed in detail.

2.4.1 Arthritis

Arthritis, particularly psoriatic arthritis and OA are reported at a higher frequency in NL (as explained in part 1) as compared to other areas of Canada. When combining risk alleles for ankylosing spondylitis, gout, OA and RA, none of the allele associations reached significance. However, of particular interest was OA for which there were only 4 SNPs included in the analysis, but still reached a p-value of 0.068. However, all four SNPs were found at a higher frequency in the NL population. And the difference between the populations was one of the largest recorded in this study, with the mean risk

frequency in NL at 0.25 compared to CEU at 0.21. It is possible that this difference could become significant if more SNPs were included in the analysis.

2.4.2 Common Cancers

The cancers included in this study were bladder, breast, colorectal, ovarian, prostate and testicular cancers. Of these, only bladder cancer risk alleles were found at a significantly different rate. The mean risk frequency for the CEU population is 0.484, while the NL population is 0.515, with a p-value of 0.016. The incidence of bladder cancer is NL was reported at 17.8 new cases/ 100,000 (age-standardized) in 2010, while Canada-wide it was 16.1/ 100,000 in 2010 (Statistics Canada, 2010). Therefore, according to Statistics Canada, the province of Newfoundland and Labrador has a higher incidence of bladder cancer than the national average. The higher mean frequency of risk variants in NL as compared to CEU found in the study support a higher rate of bladder cancer incidence in NL.

Interestingly, colorectal cancer, which is reported at a significantly higher rate in NL, did not show a significantly different frequency of risk alleles in this study. These results could be due to the NL study population being taken from a control group of a colorectal cancer research study. Any participants diagnosed with colorectal cancer were excluded from the study as controls. Therefore, the study population from NL is not necessarily a true representation of a random sample of the population in terms of colorectal cancer development. This could alter the results for the colorectal cancer frequencies, and also other related types of cancer may be skewed. For example, Lynch syndrome, observed at a high rate in NL as compared to the rest of Canada has an increased risk primarily of colorectal cancer, but also an increased risk of endometrial cancer. Therefore, the data may not accurately reflect the frequency of colorectal cancer and related cancer types.

2.4.3 Eye Diseases

Several genetic eye conditions are known to be higher in NL, such as achromatopsia (Doucette et al., 2013). Interestingly, the mean frequency of risk alleles for pathological myopia in NL is significantly lower than in the CEU population (Doucette et al., 2013). Canada-wide, the risk of myopia is reported at 30% (Doctors of optometry Canada,). Unfortunately, statistical data on the rates by province of myopia are not available. Therefore, the results cannot be compared to a provincial prevalence.

2.4.4 Heart and Circulatory Diseases

Atrial fibrillation, coronary heart disease, hypertension, myocardial infarction, sudden cardiac arrest and von Willebrand Factor were included in heart and circulatory conditions. The mean frequency for risk alleles for coronary heart disease was significantly higher in NL at 0.486 versus CEU at 0.473, with a p-value of 0.02. NL is well known for cardiomyopathies and heart disease. Age-standardized mortality rates of ischemic heart disease in 2011 in Canada are 2.1/ 100,000, while in NL the rate is 3.3/ 100,000 (Statistics Canada, 2011a). In addition the rates of self-reported heart disease in Canada are 4.9%, in NL the Central region is 7.9%, the Eastern region is 5.8%, Labrador is 4.5% and the Western region 7.4% (Statistics Canada, 2011b). The mean risk allele frequencies in the study lend support to the high rate of cardiac disease in NL and suggest that coronary heart disease in NL may be significantly influenced by genetics.

2.4.5 Renal and Digestive Tract Disorders

Celiac disease affects as many as 1 in every 100-200 people in North America and as many as 300,000 Canadians (Statistics Canada, 2012b). The clinical presentation of the disease is varied, which can make a diagnosis difficult (Green, Lebwohl, & Greywoode, 2015). Celiac disease is generally under diagnosed; in the U.S. less than 20% of people with celiac disease receive a diagnosis (Rubio-Tapia, Ludvigsson, Brantner, Murray, &

Everhart, 2012). The prevalence of celiac disease has not previously been reported in individual provinces. Therefore, the prevalence in NL is not available for analysis or for comparison to other out-bred population. Although, our data indicate that the mean risk allele frequency is higher in NL at 0.438 than in the CEU population 0.423, at $p = 0.04$. This could suggest that the rate of Celiac disease may be higher in NL as compared to other populations. Further research into the provincial rates of celiac disease is necessary.

Of the other conditions – chronic kidney disease, Crohn's disease and ulcerative colitis – none of the mean risk allele frequencies were significantly different between the two populations. Interestingly, the prevalence of ulcerative colitis in NL was also higher than in the CEU population with a p-value near significant of 0.057. The national rate of self-reported bowel disorders in 2011/2012 is 4.3% (Statistics Canada, 2012b). While the rate in Central NL is 9.8%, Eastern 7.8%, Labrador 2.1% and Western NL is 6.8% (Statistics Canada, 2012b), indicating that NL has a higher rate of bowel disorders in general and celiac disease and ulcerative colitis may represent a significant proportion of this.

2.4.6 Metabolic Syndrome and associated conditions

Mean frequencies of metabolic syndrome, type 1 diabetes and type 2 diabetes risk alleles were not significantly different. This is surprising considering the well-documented high prevalence of diabetes in Newfoundland. Although the mean risk frequency was higher, it was not significant. Perhaps even more surprising was mean risk frequency for obesity, found to be significantly lower in the NL population at $p = 0.03$. NL has the highest rate of obesity in Canada and it is considered a significant health burden. These results suggest the high rate of obesity and diabetes in Newfoundland are more likely due to the environment than to risk-associated genetic variants. Extreme obesity, while not significant at $p = 0.08$, was also higher in the CEU population.

Alternatively, there may be specific genetic factors responsible for obesity and diabetes in the NL that have not been identified yet.

2.4.7 Schizophrenia

Of the mood disorders evaluated, only the mean risk frequency for schizophrenia was different, at a significantly lower rate in NL of 0.58 than in CEU at 0.60, $p = 0.03$. Statistics Canada reports the prevalence of schizophrenia or psychosis at 0.8% (age >15, 2012) as compared to the national average at 1.3% (Statistics Canada, 2012a). This analysis supports a lower rate of schizophrenia in NL.

2.4.8 Other Conditions

The remaining conditions reported in the results included bone conditions, leukemias, lymphomas, myelomas, lung conditions, skin conditions and cancers, inflammatory conditions, thyroid disorders, aneurysms, embolisms, stroke, autoimmune disorders, and brain or neurological conditions. Of the mean risk allele frequencies from complex diseases in these categories, none were found to be significantly different between the NL population and the CEU population.

2.5 CHAPTER 5. Conclusions

2.5.1 Study Strengths

Assessing the impact of individual SNPs with low penetrance may not be particularly useful to describe a complex disease in a population as a whole, as the disease may be related to multiple, low-penetrance variants. Even when factoring in all risk alleles for a particular complex disease, the combined contribution of identified loci to disease variation is often less than 10% (Bjorkegren et al., 2015). This could explain the small variation observed in the mean risk frequencies between the NL and CEU population for most complex diseases studied here. However, by combining SNPs for a given disease the statistical power is increased and a broader picture of the burden of disease in a given population may be achieved. For example, a study of the complex trait of height utilized a "polygenic" model to combine all height-associated SNPs and was able to explain up to 40% of genetic variance (Yang et al., 2010). This approach can be expanded to estimate the genetic contribution of risk alleles for common complex disease. The mean frequencies can be compared between populations and an estimate of the overall prevalence can be calculated. The current study is unique, as it seems this method has not be used to study population genetics of common complex diseases previously. A similar study was recently published by Lim *et al.*, comparing loss-of-function variants in the rare variant 0.5-5% range (Lim et al., 2014). However, they found no significant difference in the rates of common variants between the Finnish and other out-bred European populations, possibly due to their comparisons of individual SNPs as opposed to grouped (Lim et al., 2014). Whereas the current study takes into account all SNPs reported to be associated with given disease. The data obtained with this study could help direct public policy by identifying complex conditions that have a higher overall risk allele frequency in the NL population. In genetically isolated communities, such as many areas in NL, the large majority of couples are members of the same population. This can have implications for disease risk to children. Ancestry based targeted screening is already used for specific mutations for some disorders, and is considered an effective way

to inform and identify couples at risk for rare disorders (Mathijssen et al 2015). While the current usage is for specific, rare or severe disease, it could be expanded to complex diseases or to a population-scale risk assessment.

2.5.2 Study limitations

This study has several limitations. Any limitations from SNPs reported in the genome-wide association studies catalogue can impact this study. For example, careful selection of phenotypic qualification of a disease is important. Diseases where a lot of ambiguity or heterogeneity exists can make classification difficult. Also, a study on available catalogues of disease alleles found many are imperfect or incomplete and as much as 27% of database entries were potentially unreliable by containing entries that have been erroneously included as disease variants (Bell et al., 2011). Sequencing studies are limited by the large numbers of variants with modest or low effects, allelic heterogeneity and neutral variants, and the information presented in the GWAS catalogue is not always reliably reviewed or monitored. Due to the study's reliance on previously reported SNPs associated with disease, any SNPs that have been incorrectly reported could potentially skew our data. Also, the effect sizes of the risk alleles were not considered in the analysis, risk alleles were evaluated as if they had the same effect. It would have been beneficial to include effect sizes in the analysis.

Another limitation is our study population. Our genotype data is from a random sample of Newfoundlanders. However, because the samples were originally controls from a colorectal cancer research study, participants were excluded if they had a personal history of colorectal cancer. This could affect the results, particularly for the colorectal cancer data, and related types of cancers.

The data on the CEPH/CEU population from the HapMap International Project may be limited. In the "Guidelines for Referring to the HapMap Populations in

Publications and Presentations" there is a cautionary note that the CEU genotypes may not accurately reflect the patterns of genetic variation in people with northern and western European ancestry due to the limited information collected regarding ancestral geography of the study participants (International HapMap Consortium, 2003). No phenotypic information or medical information was collected with the samples, and at the time, collection of ancestry information was not considered important to the research. A better study population, with more accurate ancestry information would be useful for future analysis. Furthermore, the HapMap population was used as the reference genome for the imputation analysis. This may introduce some bias as the same population was used for SNP imputation and risk allele population comparison.

The Newfoundland population may have lower LD than other recent founder populations. A significant inbreeding rate may be more responsible for high frequencies of risk variants (Service et al., 2006). A study by Lohmueller *et al*., of a simulated bottleneck population, suggests that the interaction of demographic processes and purifying selection can have an important impact on the distribution of deleterious variation, even in populations that did not undergo a severe founder effect (Lohmueller et al., 2008). It is suggested that association studies from these populations could be more useful than out-bred populations (Service et al., 2006) even with no evidence of linkage disequilibrium (Pope et al., 2011). Therefore, Newfoundland is considered a useful resource for studying risk variants even if the NL population does not exhibit extended linkage disequilibrium.

2.5.3 Future Directions

These results have implications for the burden of disease due to common risk variants, in Newfoundland in particular. A greater predictive capacity of disease burden could be achieved through assessing prevalence of disease in Newfoundland from groups of SNPs compared to individual SNPs alone. This could help inform public health policy.

Obesity, for example, is prevalent in Newfoundland and Labrador, at a rate of 27.7% in 2011 compared to the national average at 18.3%. NL has the highest obesity rate of any province and by 2019, it is estimated that 71% of the adult population in NL will be either overweight or obese. Based on our data, the mean overall frequency of risk variants for obesity is actually significantly lower than the outbred CEU population. This suggests that obesity in NL may be due to environmental factors rather than genetics. This information can help guide public health policy in addressing the obesity epidemic in NL by focusing on environmental risk factors that could be changed. On the other hand, it could suggest that there may be specific genetic factors responsible for the NL population not covered in this research. Further genetic studies using the NL population to help identify potential genes related to obesity could help clarify this question. An important factor to consider in future studies is the effect size of the risk alleles, which would need to be calculated in future comparisons using this method.

There are many possible future directions for this study. The method is novel, and therefore, studying common risk variants in other founder populations to see if similar results can be obtained is important. Alternatively, comparing two founder populations from the same out-bred population, such as Finland and Newfoundland could provide interesting results. Additionally, studying low-frequency variants in the Newfoundland population could yield further insights into certain disease prevalences in the population. For example, if the high rate of Type 1 Diabetes in Newfoundland were a product of a number of rare variants at a frequency of less than 5 %, the current study would not yield significant results. Regardless, further application of this method could yield interesting results in the future.

References

Ahmad, R., Qureshi, H. Y., El Mabrouk, M., Sylvester, J., Ahmad, M., & Zafarullah, M. (2007). Inhibition of interleukin 1-induced matrix metalloproteinase 13 expression in human chondrocytes by interferon gamma. *Annals of the Rheumatic Diseases, 66*(6), 782-789. doi:10.1136/ard.2006.060269

Aicher, W. K., & Rolauffs, B. (2014). The spatial organisation of joint surface chondrocytes: Review of its potential roles in tissue functioning, disease and early, preclinical diagnosis of osteoarthritis. *Annals of the Rheumatic Diseases, 73*(4), 645-653. doi:10.1136/annrheumdis-2013-204308; 10.1136/annrheumdis-2013-204308

Aigner, T., Hemmel, M., Neureiter, D., Gebhard, P. M., Zeiler, G., Kirchner, T., & McKenna, L. (2001). Apoptotic cell death is not a widespread phenomenon in normal aging and osteoarthritis human articular knee cartilage: A study of proliferation, programmed cell death (apoptosis), and viability of chondrocytes in normal and osteoarthritic human knee cartilage. *Arthritis and Rheumatism, 44*(6), 1304-1312. doi:2-T

Akhtar, N., Rasheed, Z., Ramamurthy, S., Anbazhagan, A. N., Voss, F. R., & Haqqi, T. M. (2010). MicroRNA-27b regulates the expression of matrix metalloproteinase 13 in human osteoarthritis chondrocytes. *Arthritis and Rheumatism, 62*(5), 1361-1371. doi:10.1002/art.27329; 10.1002/art.27329

Altman, R., Alarcon, G., Appelrouth, D., Bloch, D., Borenstein, D., Brandt, K., . . . Feldman, D. (1991). The american college of rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis and Rheumatism, 34*(5), 505-514.

Altman, R. D., & Gold, G. E. (2007). Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 15 Suppl A*, A1-56. doi:10.1016/j.joca.2006.11.009

Altman, R. D., Hochberg, M., Murphy, W. A.,Jr, Wolfe, F., & Lequesne, M. (1995). Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 3 Suppl A*, 3-70.

Aminoff, M., Carter, J. E., Chadwick, R. B., Johnson, C., Grasbeck, R., Abdelaal, M. A., . . . Krahe, R. (1999). Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. *Nature Genetics, 21*(3), 309-313. doi:10.1038/6831

arcOGEN Consortium, arcOGEN Collaborators, Zeggini, E., Panoutsopoulou, K., Southam, L., Rayner, N. W., . . . Loughlin, J. (2012). Identification of new susceptibility loci for osteoarthritis (arcOGEN): A genome-wide association study. *Lancet, 380*(9844), 815-823. doi:10.1016/S0140-6736(12)60681-3; 10.1016/S0140-6736(12)60681-3

Arnsdorf, E. J., Tummala, P., Castillo, A. B., Zhang, F., & Jacobs, C. R. (2010). The epigenetic mechanism of mechanically induced osteogenic differentiation. *Journal of Biomechanics, 43*(15), 2881-2886. doi:10.1016/j.jbiomech.2010.07.033; 10.1016/j.jbiomech.2010.07.033

Arthritis Alliance of Canada. (2011). The impact of arthritis in canada: Today and over the next 30 years. Retrieved from http://www.arthritisalliance.ca/en/

Arthritis Community Research and Evaluation Unit (ACREU). (2013a). Arthritis in canada (prepared for the arthritis society). Retrieved from http://www.arthritis.ca/facts

Arthritis Community Research and Evaluation Unit (ACREU). (2013b). Arthritis in newfoundland and labrador. Retrieved from https://www.arthritis.ca/document.doc?id=906

Arthritis Research UK. (2011). Obesity leads to a four-fold increase in knee osteoarthritis risk. Retrieved from http://www.arthritisresearchuk.org/news/general-news/2011/april/obesity-leads-to-a-four-fold-increase-in-knee-osteoarthritis-risk.aspx

Barter, M. J., Bui, C., & Young, D. A. (2012). Epigenetic mechanisms in cartilage and osteoarthritis: DNA methylation, histone modifications and microRNAs. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 20*(5), 339-349. doi:10.1016/j.joca.2011.12.012; 10.1016/j.joca.2011.12.012

Bateman, J. F., Rowley, L., Belluoccio, D., Chan, B., Bell, K., Fosang, A. J., & Little, C. B. (2013). Transcriptomics of wild-type mice and mice lacking ADAMTS-5 activity identifies

genes involved in osteoarthritis initiation and cartilage destruction. *Arthritis and Rheumatism, 65*(6), 1547-1560. doi:10.1002/art.37900; 10.1002/art.37900

Bear, J. C., Nemec, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M., & et al. (1987). Persistent genetic isolation in outport newfoundland. *Am. J. Med. Genet., 27*(4), 807-830. doi:10.1002/ajmg.1320270410.PMID:3425597

Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., . . . Kingsmore, S. F. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Science Translational Medicine, 3*(65), 65ra4. doi:10.1126/scitranslmed.3001756; 10.1126/scitranslmed.3001756

Billinghurst, R. C., Dahlberg, L., Ionescu, M., Reiner, A., Bourne, R., Rorabeck, C., . . . Poole, A. R. (1997). Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. *The Journal of Clinical Investigation, 99*(7), 1534-1545. doi:10.1172/JCI119316

Bjorkegren, J. L., Kovacic, J. C., Dudley, J. T., & Schadt, E. E. (2015). Genome-wide significant loci: How important are they?: Systems genetics to understand heritability of coronary artery disease and other common complex disorders. *Journal of the American College of Cardiology, 65*(8), 830-845. doi:10.1016/j.jacc.2014.12.033; 10.1016/j.jacc.2014.12.033

Blanco, F. J., Rego, I., & Ruiz-Romero, C. (2011). The role of mitochondria in osteoarthritis. *Nature Reviews.Rheumatology, 7*(3), 161-169. doi:10.1038/nrrheum.2010.213; 10.1038/nrrheum.2010.213

Blanco, F. J., & Rego-Perez, I. (2014). Editorial: Is it time for epigenetics in osteoarthritis? *Arthritis & Rheumatology (Hoboken, N.J.), 66*(9), 2324-2327. doi:10.1002/art.38710; 10.1002/art.38710

Bottini, N., & Firestein, G. S. (2013). Epigenetics in rheumatoid arthritis: A primer for rheumatologists. *Current Rheumatology Reports, 15*(11), 372-013-0372-9. doi:10.1007/s11926-013-0372-9; 10.1007/s11926-013-0372-9

Brunet, A., & Berger, S. L. (2014). Epigenetics of aging and aging-related disease. *The Journals of Gerontology.Series A, Biological Sciences and Medical Sciences, 69 Suppl 1*, S17-20. doi:10.1093/gerona/glu042; 10.1093/gerona/glu042

Bui, C., Barter, M. J., Scott, J. L., Xu, Y., Galler, M., Reynard, L. N., . . . Young, D. A. (2012). cAMP response element-binding (CREB) recruitment following a specific CpG demethylation leads to the elevated expression of the matrix metalloproteinase 13 in human articular chondrocytes and osteoarthritis. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology, 26*(7), 3000-3011. doi:10.1096/fj.12-206367; 10.1096/fj.12-206367

Burdett, T., Hall, P. N., Hasting, E., Hindorff, L. A., Junkins, H. A., Klemm, A. K., . . . Welter, D. (2015). The NHGRI-EBI catalog of published genome-wide association studies. Retrieved from www.ebi.ac.uk/gwas

Bygren, L. O., Tinghog, P., Carstensen, J., Edvinsson, S., Kaati, G., Pembrey, M. E., & Sjostrom, M. (2014). Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genetics, 15*, 12-2156-15-12. doi:10.1186/1471-2156-15-12; 10.1186/1471-2156-15-12

Carlo, M. D.,Jr, & Loeser, R. F. (2003). Increased oxidative stress with aging reduces chondrocyte survival: Correlation with intracellular glutathione levels. *Arthritis and Rheumatism, 48*(12), 3419-3430. doi:10.1002/art.11338

Carmeli, D. B. (2004). Prevalence of jews as subjects in genetic research: Figures, explanation, and potential implications. *American Journal of Medical Genetics.Part A, 130A*(1), 76-83. doi:10.1002/ajmg.a.20291

Casals, F., Hodgkinson, A., Hussin, J., Idaghdour, Y., Bruat, V., de Maillard, T., . . . Awadalla, P. (2013). Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. *PLoS Genetics, 9*(9), e1003815. doi:10.1371/journal.pgen.1003815 [doi]

Chen, C. G., Thuillier, D., Chin, E. N., & Alliston, T. (2012). Chondrocyte-intrinsic Smad3 represses Runx2-inducible matrix metalloproteinase 13 expression to maintain articular cartilage and prevent osteoarthritis. *Arthritis and Rheumatism, 64*(10), 3278-3289. doi:10.1002/art.34566; 10.1002/art.34566

Chen, W., Zhong, X., Wei, Y., Liu, Y., Yi, Q., Zhang, G., . . . Luo, J. (2015). TGF-beta regulates survivin to affect cell cycle and the expression of EGFR and MMP9 in glioblastoma. *Molecular Neurobiology,* doi:10.1007/s12035-015-9121-6

Cheung, K. S., Hashimoto, K., Yamada, N., & Roach, H. I. (2009). Expression of ADAMTS-4 by chondrocytes in the surface zone of human osteoarthritic cartilage is regulated by epigenetic DNA de-methylation. *Rheumatology International, 29*(5), 525-534. doi:10.1007/s00296-008-0744-z; 10.1007/s00296-008-0744-z

Chubinskaya, S., Kumar, B., Merrihew, C., Heretis, K., Rueger, D. C., & Kuettner, K. E. (2002). Age-related changes in cartilage endogenous osteogenic protein-1 (OP-1). *Biochimica Et Biophysica Acta, 1588*(2), 126-134.

Collins, F. S., Guyer, M. S., & Charkravarti, A. (1997). Variations on a theme: Cataloging human DNA sequence variation. *Science (New York, N.Y.), 278*(5343), 1580-1581.

Cornelis, F. M., Luyten, F. P., & Lories, R. J. (2011). Functional effects of susceptibility genes in osteoarthritis. *Discovery Medicine, 12*(63), 129-139.

Cox, L. G., van Donkelaar, C. C., van Rietbergen, B., Emans, P. J., & Ito, K. (2012). Decreased bone tissue mineralization can partly explain subchondral sclerosis observed in osteoarthritis. *Bone, 50*(5), 1152-1161. doi:10.1016/j.bone.2012.01.024; 10.1016/j.bone.2012.01.024

Crawford, M. H., Koertvelyessy, T., Huntsman, R. G., Collins, M., Duggirala, R., Martin, L., & et al. (1995). Effects of religion, economics, and geography on genetic structure of fogo island, newfoundland. *Am. J. Humn. Biol, 7*(4), 437-451. doi:10.1002/ajhb.1310070405

Culvenor, A. G., Engen, C. N., Oiestad, B. E., Engebretsen, L., & Risberg, M. A. (2014). Defining the presence of radiographic knee osteoarthritis: A comparison between the

kellgren and lawrence system and OARSI atlas criteria. *Knee Surgery, Sports Traumatology, Arthroscopy : Official Journal of the ESSKA,* doi:10.1007/s00167-014-3205-0

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research, 43*(Database issue), D662-9. doi:10.1093/nar/gku1010 [doi]

da Silva, M. A., Yamada, N., Clarke, N. M., & Roach, H. I. (2009). Cellular and epigenetic features of a young healthy and a young osteoarthritic cartilage compared with aged control and OA cartilage. *Journal of Orthopaedic Research : Official Publication of the Orthopaedic Research Society, 27*(5), 593-601. doi:10.1002/jor.20799; 10.1002/jor.20799

Davidson, E. N. B., Scharstuhl, A., Vitters, E. L., van der Kraan, P. M., & van den Berg, W. B. (2005). Reduced transforming growth factor-beta signaling in cartilage of old mice: Role in impaired repair capacity. *Arthritis Res.Ther., 7*(6), R1338-R1347.

Day-Williams, A. G., Southam, L., Panoutsopoulou, K., Rayner, N. W., Esko, T., Estrada, K., . . . Zeggini, E. (2011). A variant in MCF2L is associated with osteoarthritis. *American Journal of Human Genetics, 89*(3), 446-450. doi:10.1016/j.ajhg.2011.08.001; 10.1016/j.ajhg.2011.08.001

de la Chapelle, A., & Wright, F. A. (1998). Linkage disequilibrium mapping in isolated populations: The example of finland revisited. *Proceedings of the National Academy of Sciences of the United States of America, 95*(21), 12416-12423.

Dejica, V. M., Mort, J. S., Laverty, S., Antoniou, J., Zukor, D. J., Tanzer, M., & Poole, A. R. (2012). Increased type II collagen cleavage by cathepsin K and collagenase activities with aging and osteoarthritis in human articular cartilage. *Arthritis Research & Therapy, 14*(3), R113. doi:10.1186/ar3839; 10.1186/ar3839

Delcuve, G. P., Rastegar, M., & Davie, J. R. (2009). Epigenetic control. *Journal of Cellular Physiology, 219*(2), 243-250. doi:10.1002/jcp.21678; 10.1002/jcp.21678

Delgado-Calle, J., Sanudo, C., Fernandez, A. F., Garcia-Renedo, R., Fraga, M. F., & Riancho, J. A. (2012). Role of DNA methylation in the regulation of the RANKL-OPG system in human bone. *Epigenetics : Official Journal of the DNA Methylation Society, 7*(1), 83-91. doi:10.4161/epi.7.1.18753; 10.4161/epi.7.1.18753

den Hollander, W., Ramos, Y. F., Bos, S. D., Bomer, N., van der Breggen, R., Lakenberg, N., . . . Meulenbelt, I. (2014). Knee and hip articular cartilage have distinct epigenomic landscapes: Implications for future cartilage regeneration approaches. *Annals of the Rheumatic Diseases, 73*(12), 2208-2212. doi:10.1136/annrheumdis-2014-205980; 10.1136/annrheumdis-2014-205980

Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H., & Goldstein, D. B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biology, 8*(1), e1000294. doi:10.1371/journal.pbio.1000294 [doi]

Ding, C., Cicuttini, F., Scott, F., Cooley, H., & Jones, G. (2005). Association between age and knee structural change: A cross sectional MRI based study. *Annals of the Rheumatic Diseases, 64*(4), 549-555. doi:10.1136/ard.2004.023069

Doctors of optometry Canada.Myopia (nearsightedness). Retrieved from http://doctorsofoptometry.ca/myopia-nearsightedness/

Doucette, L., Green, J., Black, C., Schwartzentruber, J., Johnson, G. J., Galutira, D., & Young, T. L. (2013). Molecular genetics of achromatopsia in newfoundland reveal genetic heterogeneity, founder effects and the first cases of jalili syndrome in north america. *Ophthalmic Genetics, 34*(3), 119-129. doi:10.3109/13816810.2013.763993; 10.3109/13816810.2013.763993

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews.Genetics, 11*(6), 446-450. doi:10.1038/nrg2809; 10.1038/nrg2809

Eyre, D. (2002). Collagen of articular cartilage. *Arthritis Research, 4*(1), 30-35.

Feng, S., Jacobsen, S. E., & Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science (New York, N.Y.), 330*(6004), 622-627. doi:10.1126/science.1190614; 10.1126/science.1190614

Fernandez, B. A., Fox, G., Bhatia, R., Sala, E., Noble, B., Denic, N., . . . Woods, M. O. (2012). A newfoundland cohort of familial and sporadic idiopathic pulmonary fibrosis patients: Clinical and genetic features. *Respiratory Research, 13*, 64-9921-13-64. doi:10.1186/1465-9921-13-64; 10.1186/1465-9921-13-64

Fortier, L. A., Barker, J. U., Strauss, E. J., McCarrel, T. M., & Cole, B. J. (2011). The role of growth factors in cartilage repair. *Clinical Orthopaedics and Related Research, 469*(10), 2706-2715. doi:10.1007/s11999-011-1857-3; 10.1007/s11999-011-1857-3

Freeman, C., & Marchini, J. (2007). *GTOOL v0.7.5*

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity, 105*(1), 4-13. doi:10.1038/hdy.2010.54; 10.1038/hdy.2010.54

Glasson, S. S., Blanchet, T. J., & Morris, E. A. (2007). The surgical destabilization of the medial meniscus (DMM) model of osteoarthritis in the 129/SvEv mouse. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 15*(9), 1061-1069. doi:10.1016/j.joca.2007.03.006

Goldstein, D. B. (2009). Common genetic variation and human traits. *The New England Journal of Medicine, 360*(17), 1696-1698. doi:10.1056/NEJMp0806284 [doi]

Gonzalez, A. (2013). Osteoarthritis year 2013 in review: Genetics and genomics. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 21*(10), 1443-1451. doi:10.1016/j.joca.2013.07.001; 10.1016/j.joca.2013.07.001

Goodwin, W., McCabe, D., Sauter, E., Reese, E., Walter, M., Buckwalter, J. A., & Martin, J. A. (2010). Rotenone prevents impact-induced chondrocyte death. *Journal of Orthopaedic Research : Official Publication of the Orthopaedic Research Society, 28*(8), 1057-1063. doi:10.1002/jor.21091; 10.1002/jor.21091

Grazio, S., & Balen, D. (2009). Obesity: Risk factor and predictor of osteoarthritis. [Debljina: cimbenik rizika i prediktor razvoja osteoartritisa] *Lijecnicki Vjesnik, 131*(1-2), 22-26.

Green, P. H., Lebwohl, B., & Greywoode, R. (2015). Celiac disease. *The Journal of Allergy and Clinical Immunology, 135*(5), 1099-1106. doi:S0091-6749(15)00230-4 [pii]

Guccione, A. A., Felson, D. T., Anderson, J. J., Anthony, J. M., Zhang, Y., Wilson, P. W., . . . Kannel, W. B. (1994). The effects of specific medical conditions on the functional limitations of elders in the framingham study. *American Journal of Public Health, 84*(3), 351-358.

Hanna, F. S., Teichtahl, A. J., Wluka, A. E., Wang, Y., Urquhart, D. M., English, D. R., . . . Cicuttini, F. M. (2009). Women have increased rates of cartilage loss and progression of cartilage defects at the knee than men: A gender study of adults without clinical knee osteoarthritis. *Menopause (New York, N.Y.), 16*(4), 666-670.

Harada, H., Miyamoto, K., Yamashita, Y., Taniyama, K., Mihara, K., Nishimura, M., & Okada, M. (2015). Prognostic signature of protocadherin 10 methylation in curatively resected pathological stage I non-small-cell lung cancer. *Cancer Medicine,* doi:10.1002/cam4.507 [doi]

Hellman, A., & Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science (New York, N.Y.), 315*(5815), 1141-1143. doi:10.1126/science.1136352

Henrotin, Y., Kurz, B., & Aigner, T. (2005). Oxygen and reactive oxygen species in cartilage degradation: Friends or foes? *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 13*(8), 643-654. doi:10.1016/j.joca.2005.04.002

Hindorff, L. A., MacArthur, J., Morales, J., Junkins, H. A., Hall, P. N., Klemm, A. K. & Manolio, T. A. (2015). A catalog of published genome-wide association studies. Retrieved from www.genome.gov/gwastudies

Huang, C. Y., Lai, K. Y., Hung, L. F., Wu, W. L., Liu, F. C., & Ho, L. J. (2011). Advanced glycation end products cause collagen II reduction by activating janus kinase/signal

transducer and activator of transcription 3 pathway in porcine chondrocytes. *Rheumatology (Oxford, England), 50*(8), 1379-1389. doi:10.1093/rheumatology/ker134; 10.1093/rheumatology/ker134

Hudelmaier, M., Glaser, C., Hohe, J., Englmeier, K. H., Reiser, M., Putz, R., & Eckstein, F. (2001). Age-related changes in the morphology and deformational behavior of knee joint cartilage. *Arthritis and Rheumatism, 44*(11), 2556-2561.

Hugle, T., Geurts, J., Nuesch, C., Muller-Gerbl, M., & Valderrabano, V. (2012). Aging and osteoarthritis: An inevitable encounter? *Journal of Aging Research, 2012*, 950192. doi:10.1155/2012/950192; 10.1155/2012/950192

International HapMap Consortium. (2003). The international HapMap project. *Nature, 426*(6968), 789-796. doi:10.1038/nature02168

Itsara, A., Cooper, G. M., Baker, C., Girirajan, S., Li, J., Absher, D., . . . Eichler, E. E. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *American Journal of Human Genetics, 84*(2), 148-161. doi:10.1016/j.ajhg.2008.12.014; 10.1016/j.ajhg.2008.12.014

Jablonka, E., & Raz, G. (2009). Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology, 84*(2), 131-176.

Jeffries, M. A., Donica, M., Baker, L. W., Stevenson, M. E., Annan, A. C., Humphrey, M. B., . . . Sawalha, A. H. (2014). Genome-wide DNA methylation study identifies significant epigenomic changes in osteoarthritic cartilage. *Arthritis & Rheumatology (Hoboken, N.J.), 66*(10), 2804-2815. doi:10.1002/art.38762; 10.1002/art.38762

Jeffries, M. A., & Sawalha, A. H. (2011). Epigenetics in systemic lupus erythematosus: Leading the way for specific therapeutic agents. *International Journal of Clinical Rheumatology, 6*(4), 423-439. doi:10.2217/ijr.11.32

Johansson, A., Enroth, S., & Gyllensten, U. (2013). Continuous aging of the human DNA methylome throughout the human lifespan. *PloS One, 8*(6), e67378. doi:10.1371/journal.pone.0067378

Johnson, A. D., Handsaker, R. E., Pulit, S., Nizzari, M. M., O'Donnell, C. J. & de Bakker, P. I. W. (2015). SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Retrieved from http://www.broadinstitute.org/mpg/snap/index

Johnson, A. A., Akman, K., Calimport, S. R., Wuttke, D., Stolzing, A., & de Magalhaes, J. P. (2012). The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Research, 15*(5), 483-494. doi:10.1089/rej.2012.1324; 10.1089/rej.2012.1324

Jordan, J. M., Helmick, C. G., Renner, J. B., Luta, G., Dragomir, A. D., Woodard, J., . . . Hochberg, M. C. (2009). Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in african americans and caucasians: The johnston county osteoarthritis project. *The Journal of Rheumatology, 36*(4), 809-815. doi:10.3899/jrheum.080677; 10.3899/jrheum.080677

Kaurah, P., MacMillan, A., Boyd, N., Senz, J., De Luca, A., Chun, N., . . . Huntsman, D. (2007). Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *Jama, 297*(21), 2360-2372. doi:10.1001/jama.297.21.2360

Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases, 16*(4), 494-502.

Kiefer, J. C. (2007). Epigenetics in development. *Developmental Dynamics : An Official Publication of the American Association of Anatomists, 236*(4), 1144-1156. doi:10.1002/dvdy.21094

Kimura, M., & King, J. L. (1979). Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift. *Proceedings of the National Academy of Sciences of the United States of America, 76*(6), 2858-2861.

Knauper, V., Lopez-Otin, C., Smith, B., Knight, G., & Murphy, G. (1996). Biochemical characterization of human collagenase-3. *The Journal of Biological Chemistry, 271*(3), 1544-1550.

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., . . . Stefansson, K. (2009). Parental origin of sequence variants associated with complex diseases. *Nature, 462*(7275), 868-874. doi:10.1038/nature08625; 10.1038/nature08625

Kosseim, P., Pullman, D., Perrot-Daley, A., Hodgkinson, K., Street, C., & Rahman, P. (2013). Privacy protection and public goods: Building a genetic database for health research in newfoundland and labrador. *Journal of the American Medical Informatics Association : JAMIA, 20*(1), 38-43. doi:10.1136/amiajnl-2012-001009; 10.1136/amiajnl-2012-001009

Kraus, V. B., Jordan, J. M., Doherty, M., Wilson, A. G., Moskowitz, R., Hochberg, M., . . . Atif, U. (2007). The genetics of generalized osteoarthritis (GOGO) study: Study design and evaluation of osteoarthritis phenotypes. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 15*(2), 120-127. doi:10.1016/j.joca.2006.10.002

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *American Journal of Human Genetics, 80*(4), 727-739. doi:10.1086/513473

Kwon, J. M., & Goate, A. M. (2000). The candidate gene approach. *Alcohol Research & Health : The Journal of the National Institute on Alcohol Abuse and Alcoholism, 24*(3), 164-168.

Leahy, A. A., Esfahani, S. A., Foote, A. T., Hui, C. K., Rainbow, R. S., Nakamura, D. S., . . . Zeng, L. (2015). Analysis of the trajectory of osteoarthritis development in a mouse model by serial near-infrared fluorescence imaging of matrix metalloproteinase activities. *Arthritis & Rheumatology (Hoboken, N.J.), 67*(2), 442-453. doi:10.1002/art.38957; 10.1002/art.38957

Legre-Boyer, V. (2015). Viscosupplementation: Techniques, indications, results. *Orthopaedics & Traumatology, Surgery & Research : OTSR, 101*(1S), S101-S108. doi:10.1016/j.otsr.2014.07.027; 10.1016/j.otsr.2014.07.027

Li, Y., Wei, X., Zhou, J., & Wei, L. (2013). The age-related changes in cartilage and osteoarthritis. *BioMed Research International, 2013*, 916530. doi:10.1155/2013/916530; 10.1155/2013/916530

Lian, K., Lui, L., Zmuda, J. M., Nevitt, M. C., Hochberg, M. C., Lee, J. M., . . . Lane, N. E. (2007). Estrogen receptor alpha genotype is associated with a reduced prevalence of radiographic hip osteoarthritis in elderly caucasian women. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 15*(8), 972-978. doi:10.1016/j.joca.2007.02.020

Liang, Z. J., Zhuang, H., Wang, G. X., Li, Z., Zhang, H. T., Yu, T. Q., & Zhang, B. D. (2012). MiRNA-140 is a negative feedback regulator of MMP-13 in IL-1beta-stimulated human articular chondrocyte C28/I2 cells. *Inflammation Research : Official Journal of the European Histamine Research Society ...[Et Al.], 61*(5), 503-509. doi:10.1007/s00011-012-0438-6; 10.1007/s00011-012-0438-6

Lim, E. T., Wurtz, P., Havulinna, A. S., Palta, P., Tukiainen, T., Rehnstrom, K., . . . Sequencing Initiative Suomi (SISu) Project. (2014). Distribution and medical impact of loss-of-function variants in the finnish founder population. *PLoS Genetics, 10*(7), e1004494. doi:10.1371/journal.pgen.1004494; 10.1371/journal.pgen.1004494

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., . . . Feinberg, A. P. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology, 31*(2), 142-147. doi:10.1038/nbt.2487; 10.1038/nbt.2487

Loeser, R. F., Olex, A. L., McNulty, M. A., Carlson, C. S., Callahan, M., Ferguson, C., & Fetrow, J. S. (2013). Disease progression and phasic changes in gene expression in a mouse model of osteoarthritis. *PloS One, 8*(1), e54633. doi:10.1371/journal.pone.0054633; 10.1371/journal.pone.0054633

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., . . . Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in european than

in african populations. *Nature, 451*(7181), 994-997. doi:10.1038/nature06611; 10.1038/nature06611

Losina, E., Thornhill, T. S., Rome, B. N., Wright, J., & Katz, J. N. (2012). The dramatic increase in total knee replacement utilization rates in the united states cannot be fully explained by growth in population size and the obesity epidemic. *The Journal of Bone and Joint Surgery.American Volume, 94*(3), 201-207. doi:10.2106/JBJS.J.01958; 10.2106/JBJS.J.01958

Luo, L., Wu, J., Xie, J., Xia, L., Qian, X., Cai, Z., & Li, Z. (2015). Downregulated ECRG4 is associated with poor prognosis in renal cell cancer and is regulated by promoter DNA methylation. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine,* doi:10.1007/s13277-015-3913-1 [doi]

Mackie, E. J., Tatarczuch, L., & Mirams, M. (2011). The skeleton: A multi-functional complex organ: The growth plate chondrocyte and endochondral ossification. *The Journal of Endocrinology, 211*(2), 109-121. doi:10.1530/JOE-11-0048; 10.1530/JOE-11-0048

Mannion, J. J. (Ed.). (1977). *The peopling of newfoundland*

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009a). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753. doi:10.1038/nature08494 [doi]

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., . . . Visscher, P. M. (2009b). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753. doi:10.1038/nature08494; 10.1038/nature08494

Marchi, N., Pissard, S., Cliquennois, M., Vasseur, C., Le Metayer, N., Mereau, C., . . . Rose, C. (2014). Confirmation of a founder effect in a northern european population of a new beta-globin variant: HBB:C.23_26dup (codons 8/9 (+AGAA)). *European Journal of Human Genetics : EJHG,* doi:10.1038/ejhg.2014.263; 10.1038/ejhg.2014.263

Martin, L. J., Crawford, M. H., Koertvelyessy, T., Keeping, D., Collins, M., & Huntsman, R. (2000). The population structure of ten newfoundland outports. *R Human Biol, 72*(6), 997-1016.

Martin, J. A., & Buckwalter, J. A. (2003). The role of chondrocyte senescence in the pathogenesis of osteoarthritis and in limiting cartilage repair. *The Journal of Bone and Joint Surgery.American Volume, 85-A Suppl 2*, 106-110.

Mathijssen, I. B., Henneman, L., van Eeten-Nijman, J. M., Lakeman, P., Ottenheim, C. P., Redeker, E. J., . . . van Maarle, M. C. (2015). Targeted carrier screening for four recessive disorders: High detection rate within a founder population. *European Journal of Medical Genetics,* doi:10.1016/j.ejmg.2015.01.004; 10.1016/j.ejmg.2015.01.004

McKusick-Nathans Institute of Genetic Medicine, John Hopkins University (Baltimore, MD). (2015). Online mendelian inheritance in man, OMIM ®. Retrieved from http://omim.org/

Mendizabal, I., Lao, O., Marigorta, U. M., Kayser, M., & Comas, D. (2013). Implications of population history of european romani on genetic susceptibility to disease. *Human Heredity, 76*(3-4), 194-200. doi:10.1159/000360762 [doi]

Merner, N. D., Hodgkinson, K. A., Haywood, A. F., Connors, S., French, V. M., Drenckhahn, J. D., . . . Young, T. L. (2008). Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *American Journal of Human Genetics, 82*(4), 809-821. doi:10.1016/j.ajhg.2008.01.010; 10.1016/j.ajhg.2008.01.010

Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., . . . Ikegawa, S. (2007). A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nature Genetics, 39*(4), 529-533. doi:10.1038/2005

Miyamoto, Y., Shi, D., Nakajima, M., Ozaki, K., Sudo, A., Kotani, A., . . . Ikegawa, S. (2008). Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nature Genetics, 40*(8), 994-998. doi:10.1038/ng.176; 10.1038/ng.176

Moldovan, F., Pelletier, J. P., Hambor, J., Cloutier, J. M., & Martel-Pelletier, J. (1997). Collagenase-3 (matrix metalloprotease 13) is preferentially localized in the deep layer of human arthritic cartilage in situ: In vitro mimicking effect by transforming growth factor beta. *Arthritis and Rheumatism, 40*(9), 1653-1661. doi:2-4

Murphy, G., Knauper, V., Atkinson, S., Butler, G., English, W., Hutton, M., . . . Clark, I. (2002). Matrix metalloproteinases in arthritic disease. *Arthritis Research, 4 Suppl 3*, S39-49.

Muthuri, S. G., Doherty, S., Zhang, W., Maciewicz, R. A., Muir, K. R., & Doherty, M. (2013). Gene-environment interaction between body mass index and transforming growth factor beta 1 (TGFbeta1) gene in knee and hip osteoarthritis. *Arthritis Research & Therapy, 15*(2), R52. doi:10.1186/ar4214; 10.1186/ar4214

Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J. J., Furuichi, T., . . . Ikegawa, S. (2010). New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PloS One, 5*(3), e9723. doi:10.1371/journal.pone.0009723; 10.1371/journal.pone.0009723

National Cancer Institute (NCI). (2014). Ashkenazi jewish genetic panel (AJGP) - what are ashkenazi jewish genetic diseases? Retrieved from http://www.webmd.com/children/tc/ashkenazi-jewish-genetic-panel-ajgp-what-are-ashkenazi-jewish-genetic-diseases

Nelson, K. K., & Melendez, J. A. (2004). Mitochondrial redox control of matrix metalloproteinases. *Free Radical Biology & Medicine, 37*(6), 768-784. doi:10.1016/j.freeradbiomed.2004.06.008

Niu, J., Zhang, Y. Q., Torner, J., Nevitt, M., Lewis, C. E., Aliabadi, P., . . . Felson, D. T. (2009). Is obesity a risk factor for progressive radiographic knee osteoarthritis? *Arthritis and Rheumatism, 61*(3), 329-335. doi:10.1002/art.24337; 10.1002/art.24337

O'Neill, P., O'Neill, P., & O'Neill, P. (2003). *The oldest city*. St. Philip's, Nfld.: Boulder Publications. Retrieved from http://qe2a-proxy.mun.ca/login?url=http://site.ebrary.com/lib/memorial/Doc?id=10128393

Panoutsopoulou, K., Metrustry, S., Doherty, S. A., Laslett, L. L., Maciewicz, R. A., Hart, D. J., . . . arcOGEN Consortium. (2014). The effect of FTO variation on increased osteoarthritis risk is mediated through body mass index: A mendelian randomisation study. *Annals of the Rheumatic Diseases, 73*(12), 2082-2086. doi:10.1136/annrheumdis-2013-203772 [doi]

Panoutsopoulou, K., & Zeggini, E. (2013). Advances in osteoarthritis genetics. *Journal of Medical Genetics, 50*(11), 715-724. doi:10.1136/jmedgenet-2013-101754; 10.1136/jmedgenet-2013-101754

Pearson, R. G., Kurien, T., Shu, K. S., & Scammell, B. E. (2011). Histopathology grading systems for characterisation of human knee osteoarthritis--reproducibility, variability, reliability, correlation, and validity. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 19*(3), 324-331. doi:10.1016/j.joca.2010.12.005; 10.1016/j.joca.2010.12.005

Peltonen, L., Jalanko, A., & Varilo, T. (1999). Molecular genetics of the finnish disease heritage. *Human Molecular Genetics, 8*(10), 1913-1923.

Peter, B., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution; International Journal of Organic Evolution,* doi:10.1111/evo.12609; 10.1111/evo.12609

Pitsillides, A. A., & Beier, F. (2011). Cartilage biology in osteoarthritis--lessons from developmental biology. *Nature Reviews.Rheumatology, 7*(11), 654-663. doi:10.1038/nrrheum.2011.129; 10.1038/nrrheum.2011.129

Polvi, A., Linturi, H., Varilo, T., Anttonen, A. K., Byrne, M., Fokkema, I. F., . . . Muilu, J. (2013). The finnish disease heritage database (FinDis) update-a database for the genes mutated in the finnish disease heritage brought to the next-generation sequencing era. *Human Mutation, 34*(11), 1458-1466. doi:10.1002/humu.22389; 10.1002/humu.22389

Poole, A. R. (2012). Osteoarthritis as a whole joint disease. *HSS Journal : The Musculoskeletal Journal of Hospital for Special Surgery, 8*(1), 4-6. doi:10.1007/s11420-011-9248-6; 10.1007/s11420-011-9248-6

Poonpet, T., & Honsawek, S. (2014). Adipokines: Biomarkers for osteoarthritis? *World Journal of Orthopedics, 5*(3), 319-327. doi:10.5312/wjo.v5.i3.319; 10.5312/wjo.v5.i3.319

Pope, A. M., Carr, S. M., Smith, K. N., & Marshall, H. D. (2011). Mitogenomic and microsatellite variation in descendants of the founder population of newfoundland: High genetic diversity in an historically isolated population. *Genome / National Research Council Canada = Genome / Conseil National De Recherches Canada, 54*(2), 110-119. doi:10.1139/G10-102; 10.1139/G10-102

Pouliot, M. C., Labrie, Y., Diorio, C., & Durocher, F. (2015). The role of methylation in breast cancer susceptibility and treatment. *Anticancer Research, 35*(9), 4569-4574. doi:35/9/4569 [pii]

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). In Falush D., Hubisz M. J.(Eds.), *Inference of population structure using multilocus genotype data. STRUCTURE v2.3.4* (155th ed.). University of Oxford: Genetics.

Pritzker, K. P., Gay, S., Jimenez, S. A., Ostergaard, K., Pelletier, J. P., Revell, P. A., . . . van den Berg, W. B. (2006). Osteoarthritis cartilage histopathology: Grading and staging. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 14*(1), 13-29. doi:10.1016/j.joca.2005.07.014

Public Health Agency of Canada. (2011). Life with arthritis in canada: A personal and public health challenge. Retrieved from http://www.phac-aspc.gc.ca/cd-mc/arthritis-arthrite/lwaic-vaaac-10/3-eng.php#t12

Purcell, S. (2007). In Neale B., . . . Sham P. C. (Ed.), *PLINK: A toolset for whole-genome association and population-based linkage analysis. v1.07* (81st ed.) American Journal of Human Genetics.

Rahman, P., Jones, A., Curtis, J., Bartlett, S., Peddle, L., Fernandez, B. A., & Freimer, N. B. (2003). The newfoundland population: A unique resource for genetic investigation of complex diseases. *Human Molecular Genetics, 12 Spec No 2*, R167-72. doi:10.1093/hmg/ddg257

Rego-Perez, I., Fernandez-Moreno, M., Deberg, M., Pertega, S., Fernandez-Lopez, C., Oreiro, N., . . . Blanco, F. J. (2011). Mitochondrial DNA haplogroups and serum levels of proteolytic enzymes in patients with osteoarthritis. *Annals of the Rheumatic Diseases, 70*(4), 646-652. doi:10.1136/ard.2010.133637; 10.1136/ard.2010.133637

Reynard, L. N., & Loughlin, J. (2013). Insights from human genetic studies into the pathways involved in osteoarthritis. *Nature Reviews.Rheumatology, 9*(10), 573-583. doi:10.1038/nrrheum.2013.121; 10.1038/nrrheum.2013.121

Roach, H. I., Yamada, N., Cheung, K. S., Tilley, S., Clarke, N. M., Oreffo, R. O., . . . Bronner, F. (2005). Association between the abnormal expression of matrix-degrading enzymes by human osteoarthritic chondrocytes and demethylation of specific CpG sites in the promoter regions. *Arthritis and Rheumatism, 52*(10), 3110-3124. doi:10.1002/art.21300

Rowe, F. W. (1980). *A history of newfoundland and labrador*. Toronto: McGraw-Hill Ryerson.

Rubio-Tapia, A., Ludvigsson, J. F., Brantner, T. L., Murray, J. A., & Everhart, J. E. (2012). The prevalence of celiac disease in the united states. *The American Journal of Gastroenterology, 107*(10), 1538-44; quiz 1537, 1545. doi:10.1038/ajg.2012.219 [doi]

Rushton, M. D., Reynard, L. N., Barter, M. J., Refaie, R., Rankin, K. S., Young, D. A., & Loughlin, J. (2014). Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis & Rheumatology (Hoboken, N.J.), 66*(9), 2450-2460. doi:10.1002/art.38713; 10.1002/art.38713

Ryu, J. H., Lee, A., Huh, M. S., Chu, J., Kim, K., Kim, B. S., . . . Youn, I. (2012). Measurement of MMP activity in synovial fluid in cases of osteoarthritis and acute inflammatory

conditions of the knee joints using a fluorogenic peptide probe-immobilized diagnostic kit. *Theranostics, 2*(2), 198-206. doi:10.7150/thno.3477; 10.7150/thno.3477

Sacks, J. J., Luo, Y. H., & Helmick, C. G. (2010). Prevalence of specific types of arthritis and other rheumatic conditions in the ambulatory health care system in the united states, 2001-2005. *Arthritis Care & Research, 62*(4), 460-464. doi:10.1002/acr.20041; 10.1002/acr.20041

Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E. S., & Peltonen, L. (1998). CLN5, a novel gene encoding a putative transmembrane protein mutated in finnish variant late infantile neuronal ceroid lipofuscinosis. *Nature Genetics, 19*(3), 286-288. doi:10.1038/975

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., . . . Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.), 316*(5823), 445-449. doi:10.1126/science.1138659

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., . . . Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.), 305*(5683), 525-528. doi:10.1126/science.1098918

Service, S., DeYoung, J., Karayiorgou, M., Roos, J. L., Pretorious, H., Bedoya, G., . . . Freimer, N. (2006). Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature Genetics, 38*(5), 556-560. doi:10.1038/ng1770

Simons, Y. B., Turchin, M. C., Pritchard, J. K., & Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. *Nature Genetics, 46*(3), 220-224. doi:10.1038/ng.2896; 10.1038/ng.2896

Singh, J. A., Noorbaloochi, S., MacDonald, R., & Maxwell, L. J. (2015). Chondroitin for osteoarthritis. *The Cochrane Database of Systematic Reviews, 1*, CD005614. doi:10.1002/14651858.CD005614.pub2; 10.1002/14651858.CD005614.pub2

Singh, S., Kumar, D., & Sharma, N. R. (2014). Role of hyaluronic acid in early diagnosis of knee osteoarthritis. *Journal of Clinical and Diagnostic Research : JCDR, 8*(12), LC04-7. doi:10.7860/JCDR/2014/11732.5342; 10.7860/JCDR/2014/11732.5342

Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J., & Hart, D. (1996). Genetic influences on osteoarthritis in women: A twin study. *BMJ (Clinical Research Ed.), 312*(7036), 940-943.

Statistics Canada. (2010). Bladder cancer incidence. Retrieved from http://infobase.phac-aspc.gc.ca/index-en.html

Statistics Canada. (2011a). Ischemic heart disease mortality by geography age group and gender. Retrieved from http://infobase.phac-aspc.gc.ca/index-en.html

Statistics Canada. (2011b). Major cardiovascular disease mortality by geography age group and gender. Retrieved from http://infobase.phac-aspc.gc.ca/index-en.html

Statistics Canada. (2012a). CCHS mental health 2012. Retrieved from http://infobase.phac-aspc.gc.ca/index-en.html

Statistics Canada. (2012b). Self-reported bowel disorders. Retrieved from http://infobase.phac-aspc.gc.ca/index-en.html

Stuckless, S., Parfrey, P. S., Woods, M. O., Cox, J., Fitzgerald, G. W., Green, J. S., & Green, R. C. (2007). The phenotypic expression of three MSH2 mutations in large newfoundland families with lynch syndrome. *Familial Cancer, 6*(1), 1-12. doi:10.1007/s10689-006-0014-8

Sun, H. B. (2010). Mechanical loading, cartilage degradation, and arthritis. *Annals of the New York Academy of Sciences, 1211*, 37-50. doi:10.1111/j.1749-6632.2010.05808.x; 10.1111/j.1749-6632.2010.05808.x

Tahvanainen, E., Ranta, S., Hirvasniemi, A., Karila, E., Leisti, J., Sistonen, P., . . . de la Chapelle, A. (1994). The gene for a recessively inherited human childhood progressive epilepsy with mental retardation maps to the distal short arm of chromosome 8. *Proceedings of the National Academy of Sciences of the United States of America, 91*(15), 7267-7270.

Taylor, S. E., Smeriglio, P., Dhulipala, L., Rath, M., & Bhutani, N. (2014). A global increase in 5-hydroxymethylcytosine levels marks osteoarthritic chondrocytes. *Arthritis & Rheumatology (Hoboken, N.J.), 66*(1), 90-100. doi:10.1002/art.38200; 10.1002/art.38200

The Arthritis Society. (February 2015). Arthritis in canada, facts and figures. Retrieved from http://www.arthritis.ca/document.doc?id=925

Thomas, E., Peat, G., & Croft, P. (2014). Defining and mapping the person with osteoarthritis for population studies and public health. *Rheumatol, 53*(2), 338-345. doi:10.1093/rheumatology/ket346

Tsezou, A. (2014). Osteoarthritis year in review 2014: Genetics and genomics. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society, 22*(12), 2017-2024. doi:10.1016/j.joca.2014.07.024; 10.1016/j.joca.2014.07.024

Valdes, A. M., Evangelou, E., Kerkhof, H. J., Tamm, A., Doherty, S. A., Kisand, K., . . . Doherty, M. (2011). The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Annals of the Rheumatic Diseases, 70*(5), 873-875. doi:10.1136/ard.2010.134155; 10.1136/ard.2010.134155

Valdes, A. M., & Spector, T. D. (2011). Genetic epidemiology of hip and knee osteoarthritis. *Nature Reviews.Rheumatology, 7*(1), 23-32. doi:10.1038/nrrheum.2010.191; 10.1038/nrrheum.2010.191

van Beuningen, H. M., van der Kraan, P. M., Arntz, O. J., & van den Berg, W. B. (1994). Transforming growth factor-beta 1 stimulates articular chondrocyte proteoglycan synthesis and induces osteophyte formation in the murine knee joint. *Laboratory Investigation; a Journal of Technical Methods and Pathology, 71*(2), 279-290.

van der Kraan, P. M. (2014). Age-related alterations in TGF beta signaling as a causal factor of cartilage degeneration in osteoarthritis. *Bio-Medical Materials and Engineering, 24*(1 Suppl), 75-80. doi:10.3233/BME-140976; 10.3233/BME-140976

Van Lint, P., & Libert, C. (2007). Chemokine and cytokine processing by matrix metalloproteinases and its effect on leukocyte migration and inflammation. *Journal of Leukocyte Biology, 82*(6), 1375-1381. doi:10.1189/jlb.0607338

Vrtacnik, P., Marc, J., & Ostanek, B. (2014). Epigenetic mechanisms in bone. *Clinical Chemistry and Laboratory Medicine : CCLM / FESCC, 52*(5), 589-608. doi:10.1515/cclm-2013-0770; 10.1515/cclm-2013-0770

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., . . . Parkinson, H. (2014). The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research, 42*(Database issue), D1001-6. doi:10.1093/nar/gkt1229; 10.1093/nar/gkt1229

Woods, M. O., Hyde, A. J., Curtis, F. K., Stuckless, S., Green, J. S., Pollett, A. F., . . . Parfrey, P. S. (2005). High frequency of hereditary colorectal cancer in newfoundland likely involves novel susceptibility genes. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 11*(19 Pt 1), 6853-6861. doi:10.1158/1078-0432.CCR-05-0726

World Health Organization. (2011). Chronic rheumatic conditions. Retrieved from http://www.who.int/chp/topics/rheumatic/en/.

Wu, W., Billinghurst, R. C., Pidoux, I., Antoniou, J., Zukor, D., Tanzer, M., & Poole, A. R. (2002). Sites of collagenase cleavage and denaturation of type II collagen in aging and osteoarthritic articular cartilage and their relationship to the distribution of matrix metalloproteinase 1 and matrix metalloproteinase 13. *Arthritis and Rheumatism, 46*(8), 2087-2094. doi:10.1002/art.10428

Xie, Y. -., Zheng, H., Leggo, J., Scully, M. -., & Lillicrap, D. (2002). Letters to the editor: A founder factor VIII mutation, valine 2016 to alanine, in a population with an extraordinarily high prevalence of mild hemophilia A. *Thromb Haemost, 87*, 178-179.

Xu, L., & Bergan, R. C. (2006). Genistein inhibits matrix metalloproteinase type 2 activation and prostate cancer cell invasion by blocking the transforming growth factor beta-mediated

activation of mitogen-activated protein kinase-activated protein kinase 2-27-kDa heat shock protein pathway. *Molecular Pharmacology, 70*(3), 869-877. doi:10.1124/mol.106.023861

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E. V., Mort, M., . . . 1000 Genomes Project Consortium. (2012). Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing. *American Journal of Human Genetics, 91*(6), 1022-1032. doi:10.1016/j.ajhg.2012.10.015; 10.1016/j.ajhg.2012.10.015

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., . . . Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics, 42*(7), 565-569. doi:10.1038/ng.608 [doi]

Yelin, E., Murphy, L., Cisternas, M. G., Foreman, A. J., Pasta, D. J., & Helmick, C. G. (2007). Medical care expenditures and earnings losses among persons with arthritis and other rheumatic conditions in 2003, and comparisons with 1997. *Arthritis and Rheumatism, 56*(5), 1397-1407. doi:10.1002/art.22565

Zhai, G., Rabie, Z., Aref Eshghi, E., Green, R., Furey, A., Martin, G., . . . Rahman, P. (2013). An epigenome-wide association study of osteoarthritis [Abstract]. *Osteoarthritis and Cartilage, 21*(S63-S312) S170. doi:10.1016/j.joca.2013.02.366

Zhang, W., Likhodii, S., Aref-Eshghi, E., Zhang, Y., Harper, P. E., Randell, E., . . . Zhai, G. (2015). Relationship between blood plasma and synovial fluid metabolite concentrations in patients with osteoarthritis. *The Journal of Rheumatology,* doi:10.3899/jrheum.141252

Zhang, W., Likhodii, S., Zhang, Y., Aref-Eshghi, E., Harper, P. E., Randell, E., . . . Zhai, G. (2014). Classification of osteoarthritis phenotypes by metabolomics analysis. *BMJ Open, 4*(11), e006286-2014-006286. doi:10.1136/bmjopen-2014-006286; 10.1136/bmjopen-2014-006286

APPENDICES

A. Sample Calculations

Example of relative quantity (RQ) calculations for MMP-13

1. $Ct_m = (Ct_1 + Ct_2 + Ct_3) / 3$

Sample Name	Target Name	C_T	CT_m
Sample1	MMP-13	34.26573	34.171412
Sample1	MMP-13	34.00884	
Sample1	MMP-13	34.23967	
Sample2	MMP-13	34.55347	34.420247
Sample2	MMP-13	34.34153	
Sample2	MMP-13	34.36574	
Ctl	MMP-13	27.68807	27.772725
Ctl	MMP-13	27.84452	
Ctl	MMP-13	27.78559	
Sample1	Gapdh	24.52023	24.56497
Sample1	Gapdh	24.57550	
Sample1	Gapdh	24.59918	
Sample2	Gapdh	26.10388	26.187399
Sample2	Gapdh	26.29309	
Sample2	Gapdh	26.16522	
Ctl	Gapdh	25.65107	25.595953
Ctl	Gapdh	25.60254	
Ctl	Gapdh	25.53425	

2. $\Delta Ct = Ct_{m1} - Ct_{mGAPDH}$

Sample Name	$CT_{mMMP-13}$	CT_{mGAPDH}	ΔCt
Sample1	34.171412	24.56497	9.606442
Sample2	34.420247	26.187399	8.232848
Ctl	27.772725	25.595953	2.176772

3. $\Delta\Delta Ct = \Delta Ct_1 - \Delta Ct_{Ctl}$

Sample Name	ΔCt_1	ΔCt_{Ctl}	$\Delta\Delta Ct$
Sample1	9.606442	2.176772	7.42967
Sample2	8.232848	2.176772	6.056076
Ctl	2.176772	2.176772	0

4. $RQ = 2^{(-\Delta\Delta Ct)}$

Sample Name	$\Delta\Delta Ct$	RQ
Sample1	7.42967	0.0058003

Sample2	6.056076	0.0150293
Ctl	0	1

Example of body mass index calculation

BMI = weight (kg)/height (m) ^{2}

Sample ID	**Height**	**Weight**	**kg/m^2**	**kg/m^2 *10000**	**BMI**
P2	177.5	99.7	99.7/(177.5^2)	0.00316445 * 10000	31.65
P3	177	63.2	63.2/(177^2)	0.0020173 * 10000	20.17
C2	187	90	90/(187^2)	0.00257371 *10000	25.7

B. Consent Form for Study Participants

MEMORIAL UNIVERSITY
Faculty of Medicine

Discipline of Genetics
Health Sciences Centre
St. John's, NL Canada A1B 3V6
Tel: 709 777-6807 Fax:709 777-7497
www.med.mun.ca

Consent to Take Part in Research

TITLE: Newfoundland Osteoarthritis Study (NFOAS)

INVESTIGATOR(S): Drs. G. Zhai, A. Furey, G. Martin, P. Rahman, R. Green, and G. Sun.

You have been invited to take part in a research study. It is up to you to decide whether to be in the study or not. Before you decide, you need to understand what the study is for, what risks you might take and what benefits you might receive. This consent form explains the study.

The researchers will:

- discuss the study with you
- answer your questions
- keep confidential any information which could identify you personally
- be available during the study to deal with problems and answer questions

If you decide not to take part or to leave the study, your normal treatment will not be affected.

1. **Introduction/Background:**
   You have a form of arthritis called osteoarthritis. Many people in Newfoundland and other places have arthritis, but we still don't know much about what causes it. We know that what you inherit from your parents plays a part because if you have a close relative with arthritis, you are more likely to get the disease yourself. You inherit genes (made up of DNA) from both your parents. As part of this study we will be examining DNA from osteoarthritis patients and comparing it with DNA from those who do not have arthritis.

2. **Purpose of study:**
   We will recruit about 1000 people who have osteoarthritis and 1000 people who don't have, and be looking to see how the genes you inherit from your parents contribute to the cause of arthritis. We will also look to see whether small changes to these genes within the joints themselves can make a difference.

3. **Description of the study procedures and tests:**
   You are going to have a knee or hip joint replaced because of arthritis in the joint. <u>You will get exactly the same treatment whether you decide to join our research study or not.</u>

If you join the study we will ask you for some extra tubes of blood (up to 4 extra tubes) when you have your routine blood collection. From this blood we will prepare your DNA (the material that makes up your genes) that we will store in a freezer. We will test the DNA from hundreds of arthritis patients to find changes that could be involved in causing arthritis.

We will ask you to complete a questionnaire about your general medical history plus a short form just about your arthritis. We will help to complete the questionnaires, if you would like.

During your joint surgery we will be keeping small samples of the cartilage, subchondral bone, synovial membrane, and of the fluid inside the joint, which are normally thrown away. This material may also be tested for its genetic content.

We also ask your consent to access your medical records, both now and as long as this study continues (at least 5 years). This lets us compare the symptoms you have or may develop, and when you get them, to any genetic changes we may find.

4. **Length of time:**
   Completing the questionnaires should take about 30 minutes.
   There is nothing else for you to do.

5. **Possible risks and discomforts:**
   Bruising/discomfort after blood sample.

6. **Benefits:**
   It is unlikely that this study will benefit you personally. We hope that our findings may help in preventing arthritis in the future or in finding better treatments.

7. **Liability statement:**
   Signing this form gives us your consent to be in this study. It tells us that you understand the information about the research study. When you sign this form, you do not give up your legal rights. Researchers or agencies involved in this research study still have their legal and professional responsibilities.

8. **What about my privacy and confidentiality?**
   Protecting your privacy is an important part of this study. Every effort to protect your privacy will be made. However it cannot be guaranteed. For example we may be required by law to allow access to research records.

   When you sign this consent form you give us permission to
   - Collect information from you
   - Collect information from your health record
   - Share information with the people conducting the study
   - Share information with the people responsible for protecting your safety

Access to records
The members of the research team will see health and study records that identify you by name. Other people may need to look at the study records that identify you by name. This might include the research ethics board. You may ask to see the list of these people. They can look at your records only when one of the research team is present.

Use of records
The research team will collect and use only the information they need for this research study.

This information will include your

- date of birth
- sex
- medical conditions
- medications
- the results of tests and procedures you had before and during the study
- information from study interviews and questionnaires

Your name and contact information will be kept secure by the research team in Newfoundland and Labrador. It will not be shared with others without your permission. Your name will never appear in any report or article published as a result of this study.

Information collected for this study will be kept for at least five years after the end of the study.

If you decide to withdraw from the study, the information collected up to that time will continue to be used by the research team. It may not be removed. This information will only be used for the purposes of this study. At any time, you may ask that your DNA and any other specimens be destroyed.

After your part in this study ends, we may continue to review your health records to check that the information we collected is correct.

Information collected and used by the research team will be stored by Dr. Zhai who is the person responsible for keeping it secure**.**

Your access to records
You may ask the study doctor or researcher to see the information that has been collected about you.

9. Questions:

If you have any questions about taking part in this study, you can meet with the investigator, Dr. Guangju Zhai, 709-864-6683 who is in charge of the study at this institution.

Or you can talk to someone who is not involved with the study, but can advise you on your rights as a participant in a research study. This person can be reached through: Health Research Ethics Authority (HREA) at 709-777-6974 or email**:** **info**@hrea.ca

Future use of DNA or other samples:

In order to preserve a valuable resource, your DNA and other samples may be stored at the end of this research project. It is possible that these samples may be used in a future research project. **Any future research would first have to be approved by a Research Ethics Board (REB).**

Please tick **one** of the following two options:

	I agree that my samples can be used for an approved research project without contacting me again, but **only if my name* cannot be linked, in any way, to the samples.**
	Under no circumstances may my samples be used for future research. **My samples must be destroyed at the end of this present project**.

*Includes name, MCP number or any other identifying information.

The DNA and other samples from this study will be stored in St. John's, NL for an indefinite period of time.

At any time, you may ask that your DNA and any other specimens be destroyed. To do this please call Dr. Zhai at (709) 864-6683, contact anyone else on the research team, or contact the Office of the HREA at 709-777-6974 or email: info@hrea.ca

To protect your samples, we will ask that you confirm your request in writing before we destroy them.

After signing this consent you will be given a copy.

Signature Page

Study title: Genetics of Osteoarthritis in Newfoundland

Name of principal investigator: Dr. G. Zhai

To be filled out and signed by the participant:

	Please check as appropriate:	
I have read the consent form.	Yes { }	No { }
I have had the opportunity to ask questions and to discuss this study.	Yes { }	No { }
I have received satisfactory answers to all of my questions.	Yes { }	No { }
I have received enough information about the study.	Yes { }	No { }
I have spoken to _____ and he/she has answered my questions	Yes { }	No { }
I understand that I am free to withdraw from the study • at any time • without having to give a reason • without affecting my future care	Yes { }	No { }
I understand that it is my choice to be in the study and that I may not benefit.	Yes { }	No { }
I agree that the study doctor or investigator may read the parts of my hospital records which are relevant to the study.	Yes { }	No { }
I agree to take part in this study.	Yes { }	No { }

_____ _____
Signature of participant Date

_____ _____
Signature of witness Date

<u>To be signed by the investigator or person obtaining consent</u>

I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

_____ _____
Signature of investigator/person obtaining consent Date

Telephone number: _____

C. Blood Requisition Form for Study Participants

BLOOD REQUISITION
For DNA research study
Genetics of Osteoarthritis

Requested by:
Dr. Guangju Zhai
777-7286
Genetics
Faculty of Medicine
Memorial University

Study Number: _____

Patient Name _____

MCP _____

-OR-

D.O.B _____
Day-Month-Year

Please do not eat or drink anything for 6 hours before blood collection.
Thank you.

Date blood drawn _____

Where collected _____

INSTRUCTIONS FOR COLLECTION:

1. Draw **20 mL** blood in EDTA (lavender) tubes.
2. Please label all tubes with patient name.
3. Place in bag with this form.

Send blood to:

Lab Office
First Floor
Health Sciences Centre
St. John's, NL

Health Sciences Lab Office

On arrival, please call
Maggie Liu at 777-**6774**

D. General Questionnaire for Study Participants

ID Number: | | | | | | |

MEMORIAL UNIVERSITY
Faculty of Medicine
Discipline of Genetics
Health Sciences Centre
St. John's, NL Canada A1B 3V6
Tel: 709 777-6807 Fax:709 777-7497
www.med.mun.ca

Genetic Study of Osteoarthritis in the Newfoundland Population

General Questionnaire

Date form completed: | | / | | / | | | | |
(dd/mm/yyyy)

ID Number: | | | | | | |

Instruction for completing the questionnaire:

Please answer all questions to the best of your ability (leave blank if unknown).
Please write in block letters using the boxes where provided.
Use a black/blue pen.
Cross out any mistakes & write correct answers just below the relevant boxes.
Indicate your response by filling in the box next to the most appropriate answer or by writing clearly in the boxes or space provided.
Your answers will be completely confidential.

Self administered: ☐

Research assistant administered: ☐

ID Number:

Name and address

Surname

Given name

Title

Maiden Name (if applicable)

Address

Province

Postal code

Date of Birth (dd/mm/yyyy)
/ /

Place of Birth
City/Town

Province/Country

Gender: Male ☐ Female ☐

MCP number:

ID Number: ☐☐☐☐☐☐

Section 1: Demographics

1. Ethnic: White ☐ Black ☐ Other ☐, please specify ☐

2. Height: ☐☐☐ cm

3. Weight: ☐☐☐ kg

4a. Smoker: yes (current) ☐ no ☐ ex-smoker ☐

4b. If Yes, how many cigarettes do/did you smoke a day? ☐☐☐

5a. In the past 4 weeks approximately how many units of alcohol did you drink per week? (1 unit = 1 glass of wine/½ pint of beer /1 shot of sprit)? ☐☐☐

5b. Do you think your drinking habits in the last 4 weeks reflect your typical drinking habit?

Yes ☐

no, less than usual ☐

no, more than usual ☐

6a. How heavy were you when you were born?

☐☐☐ grams or ☐☐ lbs ☐☐ ozs

6b. If weight unknown, were you Light ☐ Average ☐ Heavy ☐

6c. Were you born prematurely (more than 1 week early) Yes ☐ No ☐

7a. How heavy were you at age 20 yrs? ☐☐☐ kg

7b. How heavy were you at age 50 yrs? ☐☐☐ kg

For women only:

8. At what age did your period start? ☐☐

ID Number: | | | | | | |

Section 1: Demographics (continued)

9. At what age did your period stop? ☐☐

10a. Have you had a hysterectomy (removal of the womb)? Yes ☐ No ☐

10b. If Yes, how old were you? ☐☐

10c. Did the hysterectomy include removal of the ovaries?

Yes ☐ No ☐ or Unknown ☐

11. Have you ever taken an oral contraceptive pill? Yes ☐ No ☐

12a. Have you ever taken hormone replacement therapy? Yes ☐ No ☐

12b. If Yes, how long in total did you take it for?

Less than 3 months ☐

3 to 12 months ☐

1 to 5 years ☐

Longer than 5 years ☐

13. How many live births have you had? ☐☐

ID Number: | | | | | | |

Section 2 - Occupation

14a. What was your current/last occupation (job title)?

14b. In what industry did you carry out this occupation (eg farming, shipyard, car factory, shoe shop, hospital, insurance office)?

14c. Number of years in job: ☐☐

15a. What was the main occupation that you held for the longest period of time (job title)?

15b. In what industry did you carry out this occupation (eg farming, shipyard, car factory, shoe shop, hospital, insurance office)?

15c. Number of years in job: ☐☐

For your main occupation in an average working day, did you:

16. Sit for more than two hours in total? Yes ☐ No ☐ Don't know ☐

17. Stand or walk for more than two hours in total? Yes ☐ No ☐ Don't know ☐

18. Kneel for more than one hour in total? Yes ☐ No ☐ Don't know ☐

19. Squat for more than one hour in total? Yes ☐ No ☐ Don't know ☐

20. Drive for more than 4 hours in total? Yes ☐ No ☐ Don't know ☐

21. Walk more than 2 miles in total? Yes ☐ No ☐ Don't know ☐

ID Number: | | | | | | |

Section 2 – Occupation (continued)

22. In the course of your work how often on average did you lift or carry weights of 10 kg or more?

Never ☐

Less than once per week ☐

1 to 10 times per week ☐

More than 10 times per week ☐

23. In the course of your work how often on average did you lift or carry weights of 25kg or more (Equivalent to half a bag of cement)

Never ☐

Less than once per week ☐

1 to 10 times per week ☐

More than 10 times per week ☐

ID Number: | | | | | | |

Section 3 – Medical history (1)

Please list in the box below all medication that the patient is currently taking:

Have you **EVER** been told by a Doctor or other health professional that you have **ANY** of the following conditions (please tick all that apply to you):

Cardiology

24. Congenital Heart Disease ☐
25. Coronary Heart Disease ☐
26. Heart Attack ☐
27. Hypertension (high blood pressure) ☐
28. High Blood Pressure in Pregnancy ☐
29. Angina ☐
30. High Cholesterol ☐
31. Deep Vein Thrombosis ☐
32. Varicose Veins ☐
33. Pulmonary Embolism ☐

Immunology/Chest Medicine

34. Asthma ☐
35. Hayfever ☐
36. Eczema ☐
37. Sinusitis ☐

Gastroenterology/Endocronology

38. Heartburn ☐
39. Irritable Bowel Syndrome ☐
40. Crohn's ☐
41. Diabetes ☐

Neurology/Psychiatry

42. Dyslexia ☐
43. Clinical Depression ☐
44. Anxiety/Stress Disorder ☐
45. Epilepsy ☐
46. Stroke ☐
47. Motion Sickness ☐
48. Migraine ☐

ID Number: | | | | | | |

Section 3 – Medical history (2)

Have you **EVER** been told by a Doctor or other health professional that you have **ANY** of the following conditions (please tick all that apply to you):

Oncology/Cancers

49. Breast Cancer ☐
50. Colon Cancer ☐
51a. Skin Cancer ☐ if yes, was it:
51b. Melanoma ☐
51c. Basal Cell Carcinoma ☐
51d. Squamous Cell Carcinoma ☐

Rheumatology

52. Gout ☐
53. Paget's Disease ☐
54. Bunions ☐
55. Frozen Shoulder ☐
56. Osteoporosis ☐
57. Carpal Tunnel ☐
58. Tennis Elbow ☐
59. Golfer's Elbow ☐

Dermatology/Skin

60. Acne (that caused scarring) ☐
61. Viral Warts ☐
62. Cold Sores ☐

Hearing

63. Hearing Loss ☐
64. Tinnitus (ringing in ears) ☐

Opthalmology/Eyes

65. Glaucoma ☐
66. Cataract ☐
67. Myopia (short sightedness) ☐
68. Age-related Macular Degeneration (AMD) ☐

Urology

69. Incontinence (leak urine) ☐
70. Polycystic ovary syndrome ☐

ID Number: | | | | | | |

Section 3 – Medical history (3)

Please answer the following questions by ticking the appropriate box:

71a. Have you ever lost the use of an arm, leg, vision, or ability to speak?

Yes ☐ No ☐

71b. If Yes, how long for : less than 24 hours ☐ or more than 24 hours ☐

72a. Do you usually bring up phlegm from your chest in winter? Yes ☐ No ☐

72b. Do you usually bring up phlegm on most days for at least 3 months a year?

Yes ☐ No ☐

73a. Have you had heartburn or acid regurgitation in the last year? Yes ☐ No ☐

73b. If Yes, how many times have you had heartburn/acid regurgitation in the last year?

Less than once a month ☐

About once a month ☐

Once a week or more ☐

74a. Have you been bothered by recurrent headaches? Yes ☐ No ☐

74b. If Yes, do you still have recurrent headaches? Yes ☐ No ☐

74c. If Yes, are your most troubling headaches

One sided ☐

Accompanied by sensitivity to light/noise ☐

4 to 72 hours in duration if untreated ☐

ID Number: | | | | | | |

Section 3 – Medical history (4)

Please answer the following questions by ticking the appropriate box:

75. Since turning 16 have you ever fractured or broken a bone? Yes ☐ No ☐

If Yes, please tick which of the following bones you have fractured or broken

Wrist ☐ Arm ☐ Ribs ☐ Hip ☐ Ankle ☐ Vertebra ☐ Other ☐

76. In the past 3 months have you had pain in your back on most days?

Yes ☐ No ☐

If Yes, does this pain typically radiate to either leg? Yes ☐ No ☐

77. In the past 3 months have you had any pain in any part of your body lasting at least 24 hours? Yes ☐ No ☐

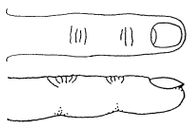
ID Number: | | | | | | |

Section 4 – Nodal status

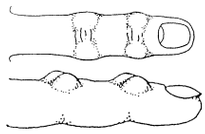
We are interested in knowing whether you have any finger nodes. These sometimes relate to arthritis at the hand and other joints. A finger node is a firm, bobbly swelling on the back of the finger joint.

For example:

A finger **without** nodes:



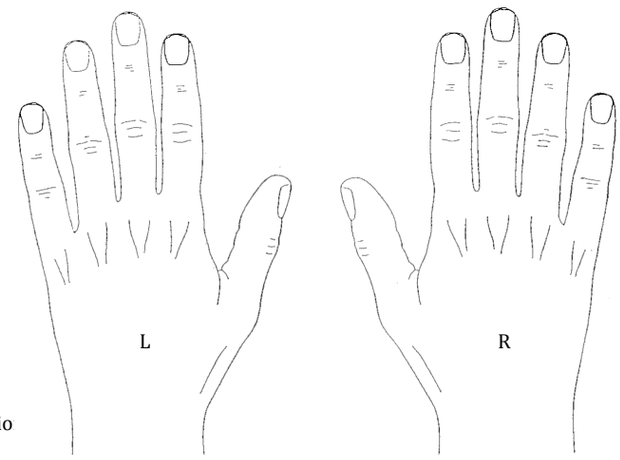
A finger **with** nodes:



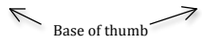
When you meet with the research assistant, please look at your hands and then answer the following questions:

78a. Do you think you have any nodes/swellings on your hands? Yes ☐ No ☐

If Yes, for each hand please circle the finger joint(s) where you have these nodes. (You may circle several joints).



ID Number: | | | | | | |



Section 4 – Nodal status (continued)

78b. If Yes, at what age did the nodes first develop? ☐☐

78c. How many nodes do you have on the: left hand ☐☐

right hand ☐☐

79. Which hand do you write with? Left ☐

Right ☐

80a. Have you suffered from pain in the fingers for most days for at least one month?

Yes ☐ No ☐

80b. If Yes, at what age did you first develop 'significant' pain in your fingers? ☐☐

81. Do you have pain in the base of your thumb (as arrow on drawing)?

Yes ☐ No ☐

ID Number: | | | | | | |

Section 5 – Family History of Osteoarthritis

82. Does/did your mother suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your mother had/have a total joint replacement of the knee/hip?

Yes ☐ No ☐ Don't know ☐

83. Does/did your father suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your father had/have a total joint replacement of the knee/hip?

Yes ☐ No ☐ Don't know ☐

84. Does/did your brothers/sisters suffer from osteoarthritis of the knee/hip?

Yes ☐ No ☐ Don't know ☐

If Yes, has/did your brothers/sisters had/have a total joint replacement of the knee/hip?

Yes ☐ No ☐ Don't know ☐

E. WOMAC Questionnaire

ID Number: ☐☐☐☐☐☐

Genetic Study of Osteoarthritis in the Newfoundland Population

The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

Name and address

Surname

Given name

Title

Maiden Name (if applicable)

Address

Province

Postal code

Date of Birth (dd/mm/yyyy)

☐☐ / ☐☐ / ☐☐☐☐

Place of Birth
City/Town

Province/Country

Gender: Male ☐ Female ☐

MCP number: ☐☐☐☐☐☐☐☐☐☐☐☐

Version 08th August 2011 1

ID Number: | | | | | | |

Section 1 – WOMAC for knee

This section assesses pain, stiffness, and functional deficit on a scale from 0 to 4.
Example:

	None 0	1	2	3	Severe 4
Example of no pain	☒	☐	☐	☐	☐
Example of severe pain	☐	☐	☐	☐	☒

1. Referring to your **knees** only how much **pain** do you experience when

	None 0	1	2	3	Severe 4
a. Walking on a flat surface	☐	☐	☐	☐	☐
b. Going up and down stairs	☐	☐	☐	☐	☐
c. At night while in bed	☐	☐	☐	☐	☐
d. Sitting or lying	☐	☐	☐	☐	☐
e. Standing upright	☐	☐	☐	☐	☐

2. Referring to your **knees** only how much **stiffness** do you experience

	None 0	1	2	3	Severe 4
a. After first awakening	☐	☐	☐	☐	☐
b. Later in the day	☐	☐	☐	☐	☐

ID Number: | | | | | | |

Section 1 – WOMAC for knee (continued)

3. Referring to your **knees** only how much **functional deficit** do you experience when

	None 0	1	2	3	Severe 4
a. Descending stairs	☐	☐	☐	☐	☐
b. Ascending stairs	☐	☐	☐	☐	☐
c. Rising from bed	☐	☐	☐	☐	☐
d. Rising from sitting	☐	☐	☐	☐	☐
e. Putting on socks	☐	☐	☐	☐	☐
f. Taking off socks	☐	☐	☐	☐	☐
g. Bending to the floor	☐	☐	☐	☐	☐
h. Lying in bed	☐	☐	☐	☐	☐
i. Walking on flat surface	☐	☐	☐	☐	☐
j. Getting in/out of the bath	☐	☐	☐	☐	☐
k. Standing	☐	☐	☐	☐	☐
l. Sitting	☐	☐	☐	☐	☐
m. Getting in/out of the car	☐	☐	☐	☐	☐
n. Getting on/off the toilet	☐	☐	☐	☐	☐
o. Heavy domestic chores	☐	☐	☐	☐	☐
p. Light domestic chores	☐	☐	☐	☐	☐
q. Shopping	☐	☐	☐	☐	☐

ID Number: | | | | | | |

Section 2 – WOMAC for hip

4. Referring to your **hips** only how much **pain** do you experience when

	None 0	1	2	3	Severe 4
a. Walking on a flat surface	☐	☐	☐	☐	☐
b. Going up and down stairs	☐	☐	☐	☐	☐
c. At night while in bed	☐	☐	☐	☐	☐
d. Sitting or lying	☐	☐	☐	☐	☐
e. Standing upright	☐	☐	☐	☐	☐

5. Referring to your **hips** only how much **stiffness** do you experience

	None 0	1	2	3	Severe 4
a. After first awakening	☐	☐	☐	☐	☐
b. Later in the day	☐	☐	☐	☐	☐

ID Number: | | | | | | |

Section 2 – WOMAC for hip (continued)

6. Referring to your **hips** only how much **functional deficit** do you experience when

	None 0	1	2	3	Severe 4
a. Descending stairs	☐	☐	☐	☐	☐
b. Ascending stairs	☐	☐	☐	☐	☐
c. Rising from bed	☐	☐	☐	☐	☐
d. Rising from sitting	☐	☐	☐	☐	☐
e. Putting on socks	☐	☐	☐	☐	☐
f. Taking off socks	☐	☐	☐	☐	☐
g. Bending to the floor	☐	☐	☐	☐	☐
h. Lying in bed	☐	☐	☐	☐	☐
i. Walking on flat surface	☐	☐	☐	☐	☐
j. Getting in/out of the bath	☐	☐	☐	☐	☐
k. Standing	☐	☐	☐	☐	☐
l. Sitting	☐	☐	☐	☐	☐
m. Getting in/out of the car	☐	☐	☐	☐	☐
n. Getting on/off the toilet	☐	☐	☐	☐	☐
o. Heavy domestic chores	☐	☐	☐	☐	☐
p. Light domestic chores	☐	☐	☐	☐	☐
q. Shopping	☐	☐	☐	☐	☐