Gluten-Related Disorders in a Population of Canadian Adults

by

Joseph J Jamnik

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Nutritional Sciences
University of Toronto

© Copyright by Joseph Jamnik 2017
Gluten-Related Disorders in a Population of Canadian Adults

Joseph Jamnik

Doctor of Philosophy

Department of Nutritional Sciences
University of Toronto

2017

Abstract

**Background:** The prevalence of celiac disease (CD) in Canadian adults is unknown, and the extent of adverse health effects associated with undiagnosed CD remains unclear. Gluten sensitivity in the absence of CD is also thought to exist; however, the physiological effects of gluten intake in those without CD remain poorly understood.

**Objectives:** To determine the prevalence of CD seropositivity in Canadian adults, investigate whether undiagnosed CD is associated with altered biomarkers of health, and examine associations of gluten intake with proteomic biomarkers and genetic variants in those without CD using a multi-omics approach.

**Methods:** Subjects were 2,832 men and women from the Toronto Nutrigenomics and Health and Toronto Healthy Diet studies. Fasting blood samples were collected for CD screening, biomarker assessment and DNA isolation. CD was assessed using self-reported diagnoses and positive tissue transglutaminase (tTG) serology. Dietary gluten intake was estimated using a 196-item food-frequency questionnaire.

**Results:** The prevalence of CD seropositivity was 1 in 114 (0.88%; 95% CI, 0.57-1.30%), with the majority of cases being undiagnosed (87%). The prevalence of CD seropositivity, predisposing human leukocyte antigen risk variants, and average gluten intake was highest in Caucasians. Undiagnosed CD was associated with lower circulating HDL-cholesterol (p=0.008),
apolipoprotein-AI (p=0.02), and retinol (p=0.006), but not widespread nutrient deficiencies. In individuals without CD, gluten intake was positively associated with plasma α2-macroglobulin (p=0.004), a protein with various immune-related functions. α2-Macroglobulin was associated with variation in *ICOSLG* (p=1.67x10^-7), a T-cell stimulatory gene involved in cytokine release and gastrointestinal immune-related disorders; however, there was no significant interaction between *ICOSLG* and gluten intake on α2-macroglobulin.

**Conclusions:** Approximately 1% of Canadian adults likely have CD. Undiagnosed CD is associated with less favourable lipid profiles and lower concentrations of certain fat-soluble vitamins. In individuals without CD, gluten intake may be associated with plasma proteomic biomarkers of immune function.
Pursuing a PhD is a major endeavor which would not have been possible without the support of many individuals. I am forever grateful to those individuals, both in my professional and personal life, who have encouraged me to succeed and fostered my personal grown over the course of my studies.

First and foremost, I would like to thank my supervisor, Dr. Ahmed El-Sohemy. I am immensely appreciative of the initial opportunity I was given as a third-year undergraduate student to conduct a research project in the El-Sohemy lab. As an undergraduate, Master’s and ultimately PhD student, Ahmed has been my primary mentor and role model. Over the course of the past six years, there has never been a moment when I have not felt wholeheartedly supported. This has, without a doubt, positively influenced many aspects of my life and enabled me to aspire to great heights. Thank you, Ahmed.

I am additionally grateful for the other mentors I have had throughout my PhD. To the members of my Advisory Committee, including Dr. David Jenkins, Dr. Richard Bazinet and Dr. Ken Croitoru, and all external appraisers of my thesis, I am most appreciative of your time and valuable feedback. I would also like to thank the other faculty members and administrative staff in the Department of Nutritional Sciences for your mentorship and counsel over the years. To my former lab mates, Dr. Bibiana García-Bailo and Dr. Daiva Nielsen, thank you for your mentorship, time and patience. I will always be grateful for your guidance.

I would also like to acknowledge the funding agencies which have supported me over the course of my graduate studies. This includes the Natural Sciences and Engineering Research Council of Canada, the Ontario Graduate Scholarship Program, the Peterborough K.M. Hunter Charitable Foundation, and the Canadian Institutes of Health Research.

To all those in my personal life, I could not be more appreciative of the support network you have provided. I am grateful for my partner, Madeline Della Mora, and her family, for their love, encouragement and support. Your caring and companionship have meant the world to me. For putting up with me on a day-to-day basis, I must acknowledge my roommate and best friend, Ryan Giuricich. I am also thankful for the support I have received and friendships I have made.
with past and current lab mates, members of student council, and other students in the Department of Nutritional Sciences. You have truly made the last four years a most enjoyable experience.

I reserve my deepest thanks for my family. To all members of my extended family, thank you for being a continuous source of joy, love and optimism. To my younger brothers, Andrew and Matthew, thanks for all your brotherly love throughout the years. Finally, I would like to thank my parents, Josie and John Jamnik, for all the sacrifices they have made throughout their lives to ensure that I have had every opportunity to succeed. Anything I accomplish is truly the result of your love and support. I could not be more appreciative and grateful for everything you have given me.
# Table of Contents

Acknowledgments ........................................................................................................... iv  
Table of Contents ........................................................................................................... vi  
List of Tables .................................................................................................................... x  
List of Figures ................................................................................................................... xi  
List of Appendices .......................................................................................................... xii  
Abbreviations .................................................................................................................. xiii

Chapter 1: Literature Review ......................................................................................... 1  
1.1 Introduction ................................................................................................................. 1  
1.2 Gluten .......................................................................................................................... 2  
1.3 Celiac Disease .............................................................................................................. 4  
1.3.1 Pathophysiology ...................................................................................................... 5  
1.3.2 Genetics .................................................................................................................... 10  
1.3.3 Diagnosis .................................................................................................................. 13  
1.3.4 Treatment ................................................................................................................ 20  
1.3.5 Classification of Disease .......................................................................................... 22  
1.3.6 Complications of Untreated Disease ...................................................................... 25  
1.3.7 Epidemiology ............................................................................................................ 29  
1.4 Non-Celiac Gluten Sensitivity ...................................................................................... 35  
1.4.1 Diagnosis .................................................................................................................. 37  
1.4.2 Prevalence ............................................................................................................... 40  
1.4.3 Potential Mechanisms .............................................................................................. 41  
1.5 Omics Technologies ..................................................................................................... 44  
1.5.1 Proteomics ............................................................................................................... 44
3.3.3 Biomarker Assessment ................................................................. 81
3.3.4 Statistical Analysis ................................................................. 82
3.4 Results ......................................................................................... 84
3.5 Discussion .................................................................................. 88

Chapter 4: Gluten Intake and Plasma Proteomic Biomarkers in Young Adults Without Celiac Disease ................................................................. 95

4.1 Abstract ....................................................................................... 95
4.2 Introduction .................................................................................. 97
4.3 Methods ....................................................................................... 99
  4.3.1 Study Population ...................................................................... 99
  4.3.2 Dietary Assessment .................................................................... 99
  4.3.3 Proteomics Measurements ...................................................... 100
  4.3.4 Statistical Analysis .................................................................. 100
4.4 Results ......................................................................................... 101
4.5 Discussion .................................................................................... 107

Chapter 5: A Genome-Wide Association Study of Plasma Proteomic Biomarkers of Gluten Intake in Adults Without Celiac Disease ......................................................... 113

5.1 Abstract ....................................................................................... 113
5.2 Introduction .................................................................................. 115
5.3 Methods ....................................................................................... 116
  5.3.1 GWAS Cohort ......................................................................... 116
  5.3.2 α2-Macroglobulin Measurements ............................................ 117
  5.3.3 Gluten Assessment ................................................................. 117
  5.3.4 Genotyping ............................................................................. 118
  5.3.5 Statistical Analysis .................................................................. 120
5.4 Results ......................................................................................... 121
5.5 Discussion .................................................................................... 129
Chapter 6: General Discussion

6.1 Overview and Discussion

6.2 Limitations

6.2.1 Celiac Disease Assessment

6.2.2 Biochemical Measurements

6.2.3 Dietary Assessment

6.2.4 Nature of Study Populations

6.3 Future Research

6.3.1 Celiac Disease

6.3.2 Non-Celiac Gluten Sensitivity

6.4 Implications

6.5 Summary of Thesis

References

Appendices
List of Tables

Table 1.1: Summary of diagnostic performance for CD-specific serologic tests .......................... 17

Table 2.1: Subject characteristics for TNH and THD Study Populations................................. 63

Table 2.2: Prevalence of positive CD serology and previous diagnoses across ethnocultural groups .......................................................................................................................... 64

Table 3.1: Subject characteristics by CD serology status ............................................................. 85

Table 3.2: Lipid profiles by CD serology status ........................................................................... 86

Table 3.3: Biomarkers of nutritional status by CD serology status ............................................. 87

Table 4.1: Subject characteristics, serum biomarkers and dietary patterns by tertile of gluten intake ........................................................................................................................................ 103

Table 4.2: Plasma protein concentration by tertile of gluten intake in young adults without celiac disease .................................................................................................................................. 105

Table 5.1: Subject characteristics for TNH GWAS Study Population ........................................ 123

Table 5.2: Summary of variants associated with plasma α2-macrogobulin with the highest genome-wide significance ........................................................................................................ 124

Table 5.3: Association between gluten and α2-macrogobulin stratified by rs7278004 genotype ........................................................................................................................................ 125
List of Figures

**Figure 1.1:** Peptide sequence of a prominent 33mer present in digested gluten fragments........ 3

**Figure 1.2:** Celiac disease diagnostic cascade ................................................................. 19

**Figure 1.3:** Worldwide prevalence of celiac disease......................................................... 32

**Figure 1.4:** Non-celiac gluten sensitivity diagnostic cascade .............................................. 39

**Figure 2.1:** Prevalence of celiac disease-associated elevated-risk HLA genotypes across ethnocultural groups ........................................................................................................... 65

**Figure 2.2:** Prevalence of HLA-DQ genotypes across ethnocultural groups................. 66

**Figure 2.3:** Prevalence of elevated tTG antibodies by celiac disease-associated HLA-DQ risk category ............................................................................................................ 67

**Figure 2.4:** Average energy-adjusted gluten intake by ethnocultural group...................... 68

**Figure 5.1:** Manhattan plot for α2-macrogobulin GWAS .................................................... 126

**Figure 5.2:** Quantile-quantile plot α2-macrogobulin GWAS ............................................. 127

**Figure 5.3:** Regional association plot on chromosome 21 .................................................. 128
List of Appendices

**Appendix A1**: Gluten content of gluten-rich food items listed in Toronto-modified Willett FFQ

........................................................................................................................................... 190

**Appendix A2**: Population structure analyses for Toronto Nutrigenomics and Health GWAS Cohort
........................................................................................................................................... 191
Abbreviations

25OHD 25-hydroxyvitamin D
ACG American College of Gastroenterology
ANCOVA Analysis of covariance
ANOVA Analysis of variance
APC Antigen presenting cell
ATI Amylase trypsin inhibitor
BMI Body mass index
CD Celiac disease
CD4/CD8 Cluster of differentiation 4/cluster of differentiation 8
CVD Cardiovascular disease
ELISA Enzyme-linked immunosorbent assay
EMA Anti-endomysial antibodies
ESPGHAN European Society for Pediatric Gastroenterology, Hepatology and Nutrition =
FDA Food and Drug Administration
FFQ Food frequency questionnaire
FODMAP Fermentable, oligosaccharides, disaccharides, monosaccharides, and polyols
GHLQ General health and lifestyle questionnaire
GLM General linear model
GLMM General linear mixed model
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSRS</td>
<td>Gastrointestinal Symptom Rating Scale</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>HDL</td>
<td>High-density lipoprotein</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leukocyte antigen</td>
</tr>
<tr>
<td>HOMA-Beta</td>
<td>Homeostasis model of beta cell function</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>Homeostasis model of insulin resistance</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>hs-CRP</td>
<td>High-sensitivity C-reactive protein</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>IBS</td>
<td>Irritable bowel syndrome</td>
</tr>
<tr>
<td>ICOSLG</td>
<td>Inducible T-cell costimulatory ligand</td>
</tr>
<tr>
<td>IEL</td>
<td>Intraepithelial lymphocytes</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IgA</td>
<td>Immunoglobulin A</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-density lipoprotein</td>
</tr>
<tr>
<td>METs</td>
<td>Metabolic equivalent hours per week</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>NCGS</td>
<td>Non-celiac gluten sensitivity</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health and Nutrition Examination Survey</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis Systems</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Th</td>
<td>T helper</td>
</tr>
<tr>
<td>THD</td>
<td>Toronto Healthy Diet</td>
</tr>
<tr>
<td>TNH</td>
<td>Toronto Nutrigenomics and Health</td>
</tr>
<tr>
<td>tTG</td>
<td>Tissue transglutaminase</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>WGO</td>
<td>World Gastroenterology Organization</td>
</tr>
</tbody>
</table>
Chapter 1 : Literature Review

1.1 Introduction

Gluten is a protein found in dietary cereal grains. Sensitivity to gluten is thought to cause a number of adverse health effects, including gastrointestinal discomfort, skin rashes, and respiratory problems (Kagnoff 2007). Celiac disease (CD) represents the most severe form of gluten intolerance, and individuals with CD must adhere to a gluten-free diet. Adverse autoimmune reactions associated with CD occur in genetically susceptible individuals upon the ingestion of gluten-containing foods (Kagnoff 2007). Variation in the human leukocyte antigen (HLA) – DQ locus is known to confer susceptibility to CD, with virtually all of those diagnosed possessing specific HLA-DQ2 or DQ8 haplotypes (Pietzak et al. 2009). Although studying the prevalence of CD remains challenging due to the high number of undiagnosed cases, screening-based studies have shown that approximately 1% of individuals in North America and many European countries are affected by the condition (Kagnoff 2007; Gujral, Freeman, and Thomson 2012). CD is known to occur less frequently in other populations; however, the exact prevalence of CD in individuals of different ethnic backgrounds is unclear (Cummins and Roberts-Thomson 2009). Furthermore, the prevalence of CD in Canada remains unknown. It is also unclear whether those with undiagnosed CD identified through screening studies experience the same adverse health effects as those who have been diagnosed with CD.

Non-celiac gluten sensitivity (NCGS) represents a distinct form of gluten sensitivity which is thought to be less severe than frank CD. Unlike CD, many aspects of the pathogenesis of NCGS are not well understood (Sapone et al. 2012). To date, there are no validated biomarkers for
assessing NCGS, and a diagnosis is only made when the reintroduction of gluten to the diet causes a recurrence of symptoms previously absent on a gluten-free diet, and CD has been definitively ruled out (Catassi et al. 2015). Furthermore, the pathophysiological mechanisms by which gluten ingestion might lead to gastrointestinal and extra-gastrointestinal symptoms in individuals with NCGS remain unknown (Sapone et al. 2012).

1.2 Gluten

Gluten is the term used to define the disease-activating proteins in wheat, barley, and rye. In wheat, these proteins are known as gliadins and glutenins. The analogous proteins in barley and rye are the hordeins and secalins, respectively (Kagnoff 2005). Despite speculation that the gluten content of these grains has been increasing due to selective breeding (Sapone et al. 2012), analysis of the protein content of wheat over the past century suggests that the relative amount of gluten has remained constant (Kasarda 2013). Major dietary sources of gluten include breads, pastas, cold breakfast cereals, and baked goods. Gluten can also be found at lower levels in many soups, sauces, and meats prepared with bread crumbs (van Overbeek et al. 1997; Hopman et al. 2012). In addition to these common dietary sources, gluten is present in a number of hidden sources, including vegetarian meat substitutes, frozen meals, ready-to-eat fruit and vegetable mixes, and flavoured dairy products (Hlywiak 2008). The average intake of gluten for individuals in Western countries is thought to range from 10 to 15 g per day (van Overbeek et al. 1997; Catassi et al. 2015).

Gluten peptides are known for their high content of proline and glutamine amino acids (Figure 1.1). These residues inhibit the complete proteolytic digestion of gluten by enzymes secreted in
the stomach, pancreas, and by enzymes that are present on the brush border membrane of the small intestine (Kagnoff 2007). These enzymes cannot cleave proteins at proline residues as they lack prolyl endopeptidase activity (Shan et al. 2002; Hausch et al. 2002). This results in the accumulation of large, semi-digested gluten peptide fragments in the small intestine. Such fragments, along with other digested proteins, may accumulate to higher levels in the small intestine of individuals with gluten-related disorders (Kagnoff 2007).

**Figure 1.1:** Peptide sequence of a prominent 33mer present in digested gluten fragments.

![Peptide sequence](image)

Proline (P) and glutamine (Q) amino acid residues highlighted. Peptide sequence adapted from (Shan et al. 2002).

The increased awareness and diagnosis of gluten-related disorders has led to the rise of the gluten-free diet. The gluten-free diet is now recognized as one of the most popular diets in North America, currently surpassing the low-carbohydrate diet and fat-free diet in popularity (Sapone et al. 2012). The market for gluten-free food products was valued at $3 billion (USD) in 2012 (Sapone et al. 2012), and estimates continue to rise. The US Food and Drug Administration (FDA) recently set out strict guidelines for the labeling of ‘gluten-free’ food products to assist individuals adhering to gluten-free diets (Food and Drug Administration 2013). According to this
mandate, foods labelled as ‘gluten-free’ must contain no more than 20 ppm (parts per million) gluten. Although individuals diagnosed with CD must adhere to a gluten-free diet, the health benefits associated with such diets remain unclear for larger subsets of the population.

### 1.3 Celiac Disease

Celiac disease (CD) is an autoimmune disorder characterized by injury to the intestinal mucosa and subsequent nutrient malabsorption. In CD, gluten-derived peptides in the small intestine trigger an adverse immune response which results in villous atrophy and flattening of the mucosal surface in severe cases (Kagnoff 2007). Individuals with CD may experience a wide range of symptoms, such as weight loss, malnutrition, steatorrhea, diarrhoea, abdominal pain, and skin rashes; however, many individuals may be relatively asymptomatic (Anderson et al. 2013). CD has a strong genetic component, where specific variants in the human leukocyte antigen (HLA)-DQ locus are necessary, but not sufficient for the development of disease (Dieli-Crimi, Cenit, and Nunez 2015).

While there has been a considerable increase in awareness of CD in recent years, CD has likely affected humans since the introduction of wheat and other gluten-containing grains to the diet approximately 10,000 years ago (Sapone et al. 2012). Indeed, evidence of CD has been found in human remains dating back nearly 2000 years (Scorrano et al. 2014). However, it was not until 1888 when Dr. Samuel Gee gave the first clinical description of CD (Gee 1888), and the mid-1900s when the link between gluten and CD was finally recognized. Observations and studies by Dr. Willem-Karel Dicke, a pediatrician from the Netherlands, led to the understanding that certain proteins in wheat were indeed the environmental culprit in CD (formerly known as Gee-
Heter disease) in the early 1940s (van Berge-Henegouwen and Mulder 1993). Prior to Dr. Dicke’s discovery, dietary intervention was known to lead to symptomatic improvement in individuals with CD, but such effects were not yet attributed to gluten. Prominent diets prescribed to individuals with CD included diets consisting exclusively of beef, various fruits, bananas, milk and even oysters (Haas 1924; Yan and Holt 2009; van Berge-Henegouwen and Mulder 1993). Currently, the management of CD-associated symptoms is almost always achieved by eliminating all sources of gluten from the diet (Kagnoff 2007).

1.3.1 Pathophysiology

Individuals with CD experience adverse physiological events shortly after the ingestion of gluten. In the gut lumen, gluten is partially digested into smaller gluten fragments; however, most fragments are relatively large due to the high frequency of proline and glutamine residues in the primary amino acid sequence of gluten peptides (Kagnoff 2007). Under normal circumstances, such peptides would remain in the gut lumen and not be able to gain access through the epithelial cell layers. However, in CD, increased intestinal permeability and impaired tight junction function are thought to enable the movement of partially-digested gluten fragments through the epithelial cell layers and into the lamina propria of the intestinal mucosa (Visser et al. 2009). Indeed, dysfunction of intestinal tight junctions and subsequently compromised barrier function are prominent in CD (Schumann et al. 2012; Szakal et al. 2010; Madara and Trier 1980). Zonulin, a protein known to regulate tight junction function, has been shown to be up-regulated in CD causing dysregulation of tight junctions and increased intestinal permeability (Fasano et al. 2000; Visser et al. 2009). Gluten peptides could also gain access to the lamina propria through a process known as retrotranscytosis (Gujral, Freeman, and Thomson 2012).
individuals with CD, antibodies in the immunoglobulin A (IgA) class secreted into the intestinal lumen can bind gluten peptides, and the resulting gluten-IgA complexes can be paracellularly transported through the epithelial cells after binding CD71 receptors on the apical surface (Matysiak-Budnik et al. 2008; Heyman and Menard 2009). Regardless of the mechanism, the loss of intestinal barrier function is crucial in CD as it enables a number of downstream immunological events in response to partially-digested gluten peptide fragments traveling into the lamina propria.

Once through the epithelial cell layers and into the lamina propria, gluten peptides encounter tissue transglutaminase (tTG), an enzyme of major importance in the pathogenesis of CD (Kagnoff 2007). tTG is a multifunctional calcium-dependent enzyme which acts as a transglutaminase, GTP-/ATP-ase, protein disulfide isomerase and protein kinase (Nurminskaya and Belkin 2012). Of crucial importance in its role in the pathogenesis of CD, tTG is responsible for the extracellular deamidation of gluten peptides once they enter the lamina propria (Kupfer and Jabri 2012). Glutamine residues, particularly ones surrounded by a large number of proline residues, are the preferred target of deamidation by tTG (Fleckenstein et al. 2002). Such glutamine residues, which are abundant in digested gluten fragments (see Figure 1.1), are deamidated to form negatively charged glutamic acid (Molberg et al. 1998). Before deamination, such gluten peptides have a neutral charge and a low binding affinity for specific HLA-restricted heterodimers (DQ2 and DQ8) on antigen presenting cells (APCs) in the lamina propria. Deamidation by tTG gives fragmented gluten peptides an overall negative charge, which substantially increases their binding affinity for DQ2 and DQ8 heterodimers (Molberg et al. 1998; van de Wal et al. 1998).
Through pathways that have not been completely elucidated, the body begins to produce anti-tTG autoantibodies in CD. In addition to facilitating the deamidation of gluten peptides, tTG is known to form cross-linkages with gluten peptide fragments (Lindfors, Maki, and Kaukinen 2010). It is thought that this cross-linking between gluten peptides and tTG could facilitate the production of tTG autoantibodies (Solld et al. 1997; Solld and Jabri 2005; Fleckenstein et al. 2004). Indeed, B cells for self-antigens (eg. tTG) are present under normal conditions; however, differentiation into plasma cells and (auto)antibody production in such cells is only initiated in the presence of specific helper T cells which bind to antigens present on the cell surface (Zinkernagel et al. 1991; Solld et al. 1997). Such B cells specific for tTG could bind tTG-gluten complexes (Fleckenstein et al. 2004), consume such complexes through endocytosis, and ultimately present both gluten (ie. foreign) and tTG (ie. self) peptides on their cell surfaces (Solld et al. 1997). It is understood that helper T cells which recognize gluten peptides are present in the lamina propria of individuals with CD (Lundin et al. 1993; Lundin et al. 1994; Kagnoff 2007), and it is possible that such T cells provide help for these tTG-specific B which also present gluten peptides on their cell surface to differentiate into plasma cells and begin tTG autoantibody production and secretion (Solld et al. 1997). Such a mechanism would explain why the production of tTG autoantibodies does not persist in individuals with CD after the initiation of a gluten-free diet, as the lack of tTG-gluten complexes would preclude the presentation of gluten peptide fragments on tTG B cells and subsequent activation by gluten-specific helper T cells (Solld et al. 1997; Solld and Jabri 2005).

While the presence of autoantibodies against tTG are a hallmark of CD, it is not entirely clear whether such antibodies play an active role in the pathogenesis of CD (Lindfors, Maki, and Kaukinen 2010; Kagnoff 2007). Some studies have shown that anti-tTG antibodies are capable of actually enhancing tTG function in vitro (Kiraly et al. 2006; Caja et al. 2010; Myrsky, Caja, et
al. 2009). This would perpetuate the deamidation of gluten peptides and presentation to gluten-specific T cells through HLA-DQ2 and DQ8 heterodimers in individuals with CD. However, such effects have not been demonstrated \textit{in vivo} and other studies have shown that anti-tTG antibodies can inhibit tTG activity (Barone et al. 2007; Esposito et al. 2002; Dieterich et al. 2003; Lindfors, Maki, and Kaukinen 2010). Dysfunction of the mucosal vasculature in the small intestine represents another potential role for anti-tTG antibodies in the pathogenesis of CD (Lindfors, Maki, and Kaukinen 2010). Disorganized vasculature structure has been shown to occur in individuals with untreated CD (Myrsky, Syrjanen, et al. 2009), and \textit{in vitro} studies have demonstrated inhibitory effects of anti-tTG antibodies on angiogenesis (Caja et al. 2010; Myrsky et al. 2008). However, the significance of mucosal vascular disorganization in the pathogenesis of CD and the ultimate development of villous atrophy has yet to be determined (Lindfors, Maki, and Kaukinen 2010). Finally, the presence of tTG autoantibodies may be important in understanding some of the CD-specific effects on extra-intestinal organs. Indeed, tTG is present in a number of body tissues, including the brain, liver, kidney, lymph nodes and skeletal muscle (Lindfors, Maki, and Kaukinen 2010; Gujral, Freeman, and Thomson 2012). Accumulation of anti-tTG autoantibodies suggests that such autoantibodies interfere with the proper function of tTG in each of these respective tissues and potentially lead to adverse effects in individuals with CD.

After deamidation by tTG, gluten-derived peptides in the lamina propria have a high binding affinity for HLA-DQ2 and DQ8 major histocompatibility (MHC) class II heterodimers (Wolters and Wijmenga 2008). These heterodimers, which are present on activated dendritic cells or macrophages (APCs) in the lamina propria, have binding sites which preferentially bind proline-surrounded glutamic acid residues (Kim et al. 2004; Bergseng et al. 2005; Qiao et al. 2005; Kagnoff 2007). The α- and β-chains of HLA-DQ heterodimers are expressed on the cell surface
of APCs and are encoded by the DQ locus in the HLA gene region (Kagnoff 2005; Wolters and Wijmenga 2008). The presence of specific HLA-DQ2 or DQ8 heterodimers are necessary for gluten-peptide binding to APCs. Other HLA-DQ-heterodimers preclude the binding and presentation of gluten peptides on APCs, and HLA-DQ2 or DQ8 are, therefore, considered necessary for the development of CD (Kagnoff 2007; Gujral, Freeman, and Thomson 2012; Wolters and Wijmenga 2008). Once bound to HLA-DQ2 or DQ8-expressing APCs, gluten-derived peptides are exposed to HLA-restricted pathogenic cluster of differentiation (CD)4+ T cells (Lundin et al. 1993). Upon activation, these CD4+ T cells go on to proliferate and release a number of T helper 1 (Th1) pro-inflammatory cytokines, including interferon (IFN)-γ, which ultimately results in damage to the epithelial cells of the intestinal mucosa (Kagnoff 2007; Gujral, Freeman, and Thomson 2012; Kupfer and Jabri 2012).

Although the activation HLA-restricted CD4+ T cells specific to gluten peptides is an important adaptive immune response in the pathogenesis of CD, the interaction between this pathway and innate immune mechanisms are necessary for the development of mucosal damage that is characteristic of CD (Kupfer and Jabri 2012). Indeed, Th1 cytokines produced by gluten-specific CD4+ T cells simulate cytotoxicity of CD8+ intraepithelial lymphocytes (IELs) and natural killer cells within the epithelial cell layer (Hue et al. 2004; Gujral, Freeman, and Thomson 2012; Jabri and Sollid 2006). These CD8+ IELs and natural killer cells, in turn, cause apoptosis of the epithelial cells in the intestinal mucosa in a T cell receptor-independent manner (Meresse et al. 2004; Gujral, Freeman, and Thomson 2012; Jabri and Sollid 2006). Additionally, certain gluten-derived peptides have been shown to stimulate interleukin (IL)-15 production in the epithelial cells (Di Sabatino et al. 2006; Maiuri et al. 2003). IL-15 may then act to increase intestinal permeability, promote IFN-γ production by IELs, and increase cytotoxicity of select natural killer cells within the intestinal epithelium through various pathways (Gujral, Freeman, and
Thomson 2012). Such pathways work synergistically with the CD4+-mediated adaptive immune response to gluten peptides to perpetuate damage to the intestinal mucosa that is seen in patients with CD (Jabri and Sollid 2006; Gujral, Freeman, and Thomson 2012).

1.3.2 Genetics

CD is known to have a strong genetic component. Twin studies and those involving first-degree relatives of individuals with CD have shown that the development of CD is, indeed, highly heritable (Greco et al. 2002; Hervonen et al. 2000; Fasano et al. 2003). Variants in the human leukocyte antigen (HLA) gene region, located in the major histocompatibility complex (MHC) on chromosome 6, are major risk factors for CD (Wolters and Wijmenga 2008). Genes in this region are known to play a role in many autoimmune disorders, including rheumatoid arthritis, type I diabetes, and multiple sclerosis (Jones et al. 2006; Lie and Thorsby 2005). Alleles in the MHC class II HLA-DQ locus that encode specific DQ2 and DQ8 heterodimers are present in virtually all individuals with CD (Pietzak et al. 2009; Dieli-Crimi, Cenit, and Nunez 2015). Approximately 85-95% of individuals with CD are thought to possess the HLA-DQ2 heterodimer, with the remaining fraction of those with CD possessing the DQ8 molecule (Sollid 2000; Barakauskas, Lam, and Estey 2014). Cases of HLA-DQ2/8 negative CD are extremely rare, and the occurrence of such cases is thought to potentially be the result of incorrect diagnoses (Anderson et al. 2013; Barakauskas, Lam, and Estey 2014). HLA-DQ risk molecules present on the surface of (APCs) have a high affinity for gluten-derived peptides that have been deamidated by tTG. When gluten is presented to these DQ molecules on APCs, it binds and triggers a number of downstream immunological effects (Kagnoff 2007; Gujral, Freeman, and
Thomson 2012). While HLA-DQ2/8 alleles appear necessary for the development of CD, they are not sufficient as up to 40% of the general population (ie. individuals without CD) possess such alleles (Dieli-Crimi, Cenit, and Nunez 2015).

HLA-DQ heterodimers possess both DQα and DQβ-chains, which are genetically encoded by HLA-DQA1 and HLA-DQB1, respectively. DQ2.5 (DQA1*05/DQB1*02) and DQ8 (DQA1*03/DQBQ*0302) haplotypes directly confer risk to CD. Other risk haplotypes include DQ2.2 (DQA1*0201/DQB1*0202) and DQ7 (DQA1*0505/DQB1*0301); however, these haplotypes must be present together, or along with DQ2.5 in order to confer risk to developing CD (Monsuur et al. 2008). Unlike DQ2.5 heterodimers, DQ2.2 is rarely found in individuals with CD due to subtle differences in the binding sites for gluten-derived peptides of these heterodimers (Fallang et al. 2009). Individuals with a greater number of risk alleles (eg. DQ2.5 homozygotes) are more likely to develop CD than individuals with only one copy (eg. DQ2.5 heterozygotes) (Megiorni et al. 2009; Pietzak et al. 2009), and comprehensive HLA-DQ-based genetic risk gradients for CD have been established (Vader et al. 2003; Karell et al. 2003; Pietzak et al. 2009).

Genetic screening for the presence of HLA-DQ2 or DQ8 is a useful negative predictor for CD since all affected individuals possess these risk haplotypes (Kagnoff 2007; Barakauskas, Lam, and Estey 2014). Individuals who test negative for HLA-DQ2 or DQ8 can virtually rule out CD for life. Although most individuals with HLA-DQ2 or DQ8 will not ultimately develop CD (Kagnoff 2007), knowledge of possessing these risk genotypes can be beneficial. It is likely that those who knowingly possess such alleles would have a greater awareness for the signs and symptoms of CD and be more likely to seek a diagnosis if such symptoms are present. Indeed, an increase in genetic screening is expected to reduce the number of undiagnosed CD cases.
(Anderson et al. 2013). This is important because undiagnosed CD can be associated with a number of severe complications (see Chapter 1.3.5) and a decreased quality of life (Zingone et al. 2015; Goddard and Gillett 2006).

While HLA-DQ2/8 variants are necessary for the development of CD, such alleles are only thought to account for ~22% of the heritability of CD (Gutierrez-Achury et al. 2015). Genome-wide association studies (GWAS) have recently identified a number of non-HLA variants that may confer additional susceptibility to CD when present with HLA-DQ2 or DQ8 (van Heel et al. 2007; Hunt et al. 2008; Trynka et al. 2009; Dubois et al. 2010; Dieli-Crimi, Cenit, and Nunez 2015). An initial GWAS aimed at identifying extra-HLA alleles which confer susceptibility for CD in 778 cases and 1,422 controls identified genotype status at rs13119723 as being significantly associated with CD (van Heel et al. 2007). This single nucleotide polymorphism (SNP) is in linkage disequilibrium with the protein-coding *IL2* and *IL21* genes. Both genes encode cytokines that are important factors in T cell activation and proliferation (van Heel et al. 2007), making them prime candidates for causal CD susceptibility regions. Additional GWAS and candidate gene studies have identified up to 39 extra-HLA loci, many of which are involved in immune function, that may be involved in CD (Hunt et al. 2008; Dubois et al. 2010; Trynka et al. 2009). Combining a number of these loci along with HLA-DQ variants in screening panels has been shown to increase the predictive power of genetic tests for CD (Romanos et al. 2009). However, all extra-HLA variants identified to date only explain ~7% of the additional heritability of CD, suggesting the potential involvement of rare genetic variants with small effects and possible gene-gene and gene-environment interactions (Dieli-Crimi, Cenit, and Nunez 2015).
1.3.3 Diagnosis

A diagnosis of CD is made using a combination of serologic tests and a confirmatory biopsy. In the past, diagnosing CD relied heavily on multiple, highly-invasive biopsies while patients were on both gluten-containing and gluten-free diets (Gujral, Freeman, and Thomson 2012). The discovery of CD-specific serologic markers has enabled minimally-invasive serologic screening as the first-line test in individuals with suspected CD (Barakauskas, Lam, and Estey 2014). Individuals seeking serologic testing are encouraged to maintain a normal gluten-containing diet leading up to testing as the premature elimination of gluten from the diet can result in false negative test results (Rashid and Lee 2016). While there have historically been several laboratory tests for the screening of CD, two prominent methods are currently recommended in CD testing guidelines (Husby et al. 2012; Rubio-Tapia et al. 2013). The assessment of either anti-endomysial antibodies (EMA) or anti-tTG antibodies in the IgA class have been shown to be highly sensitive and specific markers for CD.

Since its discovery in the 1980s, screening for EMA has been prominent in the diagnosis of CD (Chorzelski et al. 1984; Barakauskas, Lam, and Estey 2014). EMA antibodies are detected by indirect immunofluorescence. Characteristic staining patterns on the smooth muscle endomysium of monkey esophageal or bladder tissue (Chorzelski et al. 1984; Reeves et al. 2006) as well as human umbilical cord tissue (Ladinser, Rossipal, and Pittschieler 1994; Volta et al. 1995) is indicative of the presence of EMA in human serum samples. The presence of such antibodies is a highly sensitive and specific marker for CD (see Table 1.1), and titers of EMA correlate with the degree of villous atrophy in individuals with CD (Barakauskas, Lam, and Estey 2014). While EMA assessment has been proven to be extremely valuable in diagnosing CD, there are a
number of drawbacks to this technique. First, common forms of EMA testing require esophageal or bladder tissue from endangered monkey species (Volta et al. 1995). Furthermore, indirect immunofluorescence is highly subjective and requires experienced personnel to correctly interpret the staining patterns produced (Leffler and Schuppan 2010; Nandiwada and Tebo 2013). Nevertheless, EMA testing remains prominent among laboratories that routinely screen for CD (Barakauskas, Lam, and Estey 2014).

The discovery of tTG as the autoantigen detected by EMA has led to the development of many commercial assays to directly quantify tTG antibody levels as a marker of CD (Dieterich et al. 1997; Barakauskas, Lam, and Estey 2014). Measurement of tTG is done using enzyme linked immunosorbent assays (ELISA). While initial tTG assays used tTG isolated from guinea pig liver as the substrate, modern assays often utilize human-recombinant tTG to achieve higher test accuracy (Fabiani et al. 2004; Wong et al. 2002; Wolters et al. 2002). The assessment of tTG antibodies using human recombinant tTG has been shown to be a highly sensitive and specific tool for CD screening (see Table 1.1). Like EMA, extremely high levels of tTG antibodies detected is often indicative of a high degree of villous atrophy (Barakauskas, Lam, and Estey 2014). Furthermore, levels of tTG decrease shortly after implementation of a gluten-free diet and routine tTG testing is often employed to monitor adherence to a gluten-free diet in adults (Rubio-Tapia et al. 2013; Nachman et al. 2011). Compared to EMA, the accuracy of tTG testing by ELISA does not depend on the experience of the personnel interpreting the results (Nandiwada and Tebo 2013; Barakauskas, Lam, and Estey 2014). Testing for tTG antibodies has also been shown to be slightly more sensitive than EMA screening (Lewis and Scott 2006). Additional advantages of tTG testing include its quantitative nature and relative ease of performance (Barakauskas, Lam, and Estey 2014; Nandiwada and Tebo 2013). For these reasons, screening for tTG is now widely recommended as the first-line test in individuals with suspected CD.
(Rubio-Tapia et al. 2013; Rashid and Lee 2016). Recently, point-of-care finger prick tTG tests have become available for quick assessment of tTG levels (Leffler and Schuppan 2010), although such tests are slightly less accurate than traditional ELISA-based assays (Raivio et al. 2007).

While EMA and tTG antibodies in the IgA class are highly sensitive and specific markers for CD, both tests have the potential of producing false negative results in the context of selective IgA deficiency (Leffler and Schuppan 2010; Barakauskas, Lam, and Estey 2014). This is particularly problematic as selective IgA deficiency (total serum IgA levels < 7 dg/ml) has been shown to be particularly prevalent in CD (Bonilla et al. 2015). Indeed, approximately 3-5% of individuals with CD have been shown to be IgA deficient (Halfdanarson, Litzow, and Murray 2007; Chow et al. 2012), while only 0.05-0.5% of the general population is affected by this immunoglobulin deficiency (Cunningham-Rundles 2001). In cases of IgA deficiency, EMA or tTG antibodies in the immunoglobulin G (IgG) class are often assessed (Leffler and Schuppan 2010; Barakauskas, Lam, and Estey 2014). Unlike secretory IgA antibodies which are prominent in mucosal secretions (Yel 2010), IgG antibodies are dominant in the systemic circulation (Papadea and Check 1989). While tTG IgG antibodies are known to be significantly less sensitive and specific for CD in individuals with sufficient IgA levels (Reeves et al. 2006; Barakauskas, Lam, and Estey 2014), the accuracy of IgG tests is markedly increased in the context of IgA deficiency. For this reason, screening for total IgA levels and appropriate follow-up with IgG-based tests if positive for deficiency is recommended when traditional IgA-based tests are negative yet the clinical suspicion of CD remains high (Wang et al. 2014; Barakauskas, Lam, and Estey 2014).
There are a number of other serologic tests for CD in addition to EMA and tTG. Before the discovery of EMA, specific reticulin-like staining patterns in rodent tissues were observed upon indirect immunofluorescence with serum from patients with CD (Eade et al. 1977). Such staining patterns are indicative of the presence of ‘anti-reticulin’ antibodies. Due to its significantly lower sensitivity for CD (see Table 1.1), screening for reticulin antibodies was replaced by EMA assays and the assessment of anti-reticulin antibodies is no longer recommended for CD-screening purposes (Barakauskas, Lam, and Estey 2014; Nandiwada and Tebo 2013). Similarly, antibodies against gliadin, a gluten-derived protein fragment, have also been found to be prominent in individuals with CD (Unsworth et al. 1981). However, like anti-reticulin antibodies, anti-gliadin antibodies in both the IgA and IgG class are no longer recommended for CD screening purposes due to their inferior sensitivity and specificity compared to both EMA and tTG assays (Leffler and Schuppan 2010). Recent assays that can detect the presence of antibodies against deamidated-gliadin peptides in both the IgA and IgG class have also been developed (Aleanzi et al. 2001; Schwertz et al. 2004; Barakauskas, Lam, and Estey 2014). Deamidated gliadin peptides are produced by the modification of gluten-derived gliadin peptides by tTG. Although such assays are generally less sensitive/specific and significantly more expensive than traditional tTG assays (Volta et al. 2010; Leffler and Schuppan 2010), they have been shown to potentially demonstrate higher diagnostic accuracy in children (Barakauskas, Lam, and Estey 2014; Leffler and Schuppan 2010).
### Table 1.1: Summary of diagnostic performance for CD-specific serologic tests.

<table>
<thead>
<tr>
<th>Serologic Test</th>
<th>Sensitivity (%) (Range)</th>
<th>Specificity (%) (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgA</td>
<td>98 (78-100)</td>
<td>98 (90-100)</td>
</tr>
<tr>
<td>IgG</td>
<td>70 (45-95)</td>
<td>95 (94-100)</td>
</tr>
<tr>
<td>EMA (IgA)</td>
<td>95 (86-100)</td>
<td>99 (97-100)</td>
</tr>
<tr>
<td>Deamidated gliadin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgA</td>
<td>88 (74-100)</td>
<td>95 (90-99)</td>
</tr>
<tr>
<td>IgG</td>
<td>80 (63-95)</td>
<td>98 (90-99)</td>
</tr>
<tr>
<td>Gliadin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgA</td>
<td>85 (57-100)</td>
<td>90 (47-94)</td>
</tr>
<tr>
<td>IgG</td>
<td>82 (42-100)</td>
<td>80 (50-94)</td>
</tr>
<tr>
<td>Reticulin (IgA)</td>
<td>72 (29-100)</td>
<td>99 (95-100)</td>
</tr>
</tbody>
</table>

Adapted from Leffler and Schuppan 2010, and Nandiwada and Tebo 2013.

While the presence of elevated EMA or tTG autoantibodies is highly specific for CD, a confirmatory upper endoscopic biopsy remains the gold standard for diagnosing CD and should be pursued in cases of positive CD serology (Barakauskas, Lam, and Estey 2014). Histological signs of CD upon biopsy include villous atrophy, crypt hyperplasia, and increased IEL infiltration in the mucosal cells (Marsh and Crowe 1995; Fry et al. 1972), although villous atrophy and inflammation may even be evident during endoscopy in some cases (Oxentenko et al. 2002). Multiple biopsy samples from the duodenum (Pais et al. 2008), including at least one from the duodenal bulb (Evans et al. 2011), are recommended considering the patchy and potentially erratic nature of histological lesions in individuals with varying degrees of mucosal damage (Weir et al. 2010; Ravelli et al. 2010). According to guidelines set by the American College of Gastroenterology (ACG), an intestinal biopsy should be performed in all cases of positive CD serology in order to make a definitive diagnosis (Rubio-Tapia et al. 2013). However, considering the highly invasive nature of endoscopic biopsy, the European Society for Pediatric
Gastroenterology, Hepatology and Nutrition (ESPGHAN) and World Gastroenterology Organization (WGO) guidelines allow for a diagnosis of CD to be inferred without biopsy in certain cases of positive CD serology (Husby et al. 2012; Bai et al. 2013). When the clinical suspicion of CD is high but serologic tests are negative, biopsy can still be pursued as some cases of true CD may have negative serological findings (Gujral, Freeman, and Thomson 2012). A comprehensive diagnostic cascade for CD proposed by the Mayo Clinic (Mayo Medical Laboratories) is summarized in Figure 1.2 (Gujral, Freeman, and Thomson 2012). While ACG and WGO guidelines recommend genetic screening for the presence of HLA-DQ2/DQ8 genotypes as a way to rule out CD in cases of equivocal serology or biopsy results (Rubio-Tapia et al. 2013; Bai et al. 2013), ESPGHAN guidelines encourage HLA screening as a first-line test in certain asymptomatic patients and allow for a diagnosis of CD to be made without biopsy in symptomatic patients with positive serology and positive HLA-DQ2/8 status (Husby et al. 2012).
Figure 1.2: Celiac disease diagnostic cascade.

Adapted from Gujral et al. 2013.
1.3.4 Treatment

Management of the clinical symptoms associated with CD is achieved through strict adherence to a gluten-free diet (Kagnoff 2007). Symptoms, damage to the intestinal mucosa, and levels of tTG autoantibodies often normalize within 6-12 months of commencing a gluten-free diet (Murray et al. 2004; Lewis and Scott 2006; See and Murray 2006). It is crucial that all sources of gluten, including all foods made with wheat, barley and rye, are eliminated from the diet. Oats are generally considered to be safe for consumption in individuals with CD as the peptide sequences of proteins present in oats are significantly different from those in wheat, barley and rye. Such peptide sequences in oats are unable to trigger immune responses in the majority of individuals with CD (Kagnoff 2005; Janatuinen et al. 1995; Janatuinen et al. 2000). However, many foods containing oats can be cross-contaminated with other grains, including wheat, barley or rye. Cross-contamination with such gluten-containing grains is problematic for individuals with CD and caution should be taken to ensure that only pure oats are consumed. Additionally, it must be noted that pure oats may be capable of triggering symptoms and mucosal damage in a small subset of patients with CD (Ludvigsson et al. 2014). Individuals with CD should also be conscious of potential cross-contamination of other food items with sources of gluten as even minute amounts of gluten can be problematic. Indeed, a recent systematic review found that as little as 10 mg of gluten can trigger symptoms and mucosal damage in individuals with CD (Akobeng and Thomas 2008). This equates to approximately 0.1% of the average daily amount of gluten consumed in Western populations (~10 g) (van Overbeek et al. 1997; Catassi et al. 2015). Considering that such trace amounts of gluten can trigger symptoms of CD and mucosal damage in affected individuals, the FDA has mandated that products labeled as “gluten-free” contain no more than 20 parts per million (ppm) gluten (Food and Drug Administration 2013).
This means that there would be a maximum of 20 mg of gluten in every 1 kg of such foods, and that one would need to consume at least 500 g to reach levels of gluten intake that may be problematic.

Although adherence to a gluten-free diet is an extremely effective treatment for the vast majority of individuals with CD, there are some drawbacks to adopting such a diet. The restrictive nature of a gluten-free diet can lead to decreased quality of life for some patients (Ludvigsson et al. 2014; Samasca et al. 2014). Such effects can be partially attributed to anxiety triggered by eating in social settings and at restaurants (Ludvigsson et al. 2014). The widespread availability of gluten-free foods may also be limited (Lee et al. 2007; Singh and Whelan 2011), especially in rural areas (Ludvigsson et al. 2014). Furthermore, even when available, gluten-free food options are significantly more expensive than their gluten-containing counterparts (Stevens and Rashid 2008; Pinto-Sanchez et al. 2015). Finally, the elimination of all gluten-containing foods can result in an increased risk of nutritional deficiencies (Saturni, Ferretti, and Bacchetti 2010), and special care must be taken to ensure adequate intakes of all macronutrients, micronutrients and minerals, and dietary fiber, especially in children (Penagini et al. 2013).

A number of novel potential treatments for CD have recently emerged (Makharia 2014). Assisted digestion approaches aim to either breakdown the toxic and otherwise indigestible proline-and glutamine-rich gluten fragments in the intestinal lumen using orally ingested endopeptidases or during food processing using bacterial-derived peptidases (Pyle et al. 2005; Tye-Din et al. 2010; Lahdeaho et al. 2014). Reducing paracellular passage of gluten through tight junctions into the lamina propria using antagonists of zonulin, a tight junction regulator, has also been explored (Paterson et al. 2007; Leffler et al. 2012). Other potential approaches include blocking the deadimation of partially-digested gluten fragments by tTG (Molberg et al. 2001; Maiuri et al.
2005), and blocking the binding of such gluten fragments to HLA-DQ2/8 heterodimers on APCs (Makharia 2014). Such approaches, if successful, would ultimately block the presentation of gluten to specific CD4+ T-cells and prevent the adverse immune reactions that lead to mucosal damage in CD. Finally, similar to approaches employed to control many IgE-mediated allergic reactions (Berin and Mayer 2013), the subcutaneous injection of gluten peptides in patients with CD has been investigated (Brown and Daveson; Makharia 2014). The goal of such an approach is to slowly increase tolerance to small amounts of dietary gluten. Despite their appeal as a potential replacement for a gluten-free diet, the novel treatments discussed above are likely to be most effective in conjunction with a gluten-free diet, rather than an alternative to it (Makharia 2014).

1.3.5 Classification of Disease

The clinical presentation of CD can be markedly different between individuals. Indeed, CD is considered a spectrum with symptoms ranging from mild extra-intestinal symptoms to severe gastrointestinal discomfort and nutrient deficiency (Sapone et al. 2012; Ferguson, Arranz, and O'Mahony 1993; Gujral, Freeman, and Thomson 2012). While CD has historically been thought of as a disorder diagnosed in childhood, an increasing amount of adult-onset cases are being identified (Dewar and Ciclitira 2005; Telega, Bennet, and Werlin 2008). There are substantial differences in the clinical presentation of childhood and adult CD, and there exists a broad range of symptoms with varying severities within each category (Ludvigsson et al. 2014; Rampertab et al. 2006). Prominent classifications of CD include the typical, atypical, silent, potential and refractory forms (Guandalini and Assiri 2014).
The typical form of CD is characterized by the presence of classical gastrointestinal symptoms including abdominal pain, diarrhea, steatorrhea, constipation, vomiting as well as unexplained weight loss and failure to thrive (Guandalini and Assiri 2014; Green and Cellier 2007). Such typical presentations of CD are often diagnosed in early childhood due to the severity of symptoms experienced by affected individuals (Rashid et al. 2005). The presence of tTG antibodies as well as a high degree of damage to the intestinal mucosa is observed in typical CD upon upper endoscopic biopsy (Guandalini and Assiri 2014). Mucosal damage is assessed by the Marsh scale, which ranges from 0 to III (Marsh 1992). A completely healthy intestinal mucosa would be considered Marsh 0. Marsh I is associated with the first signs of CD-associated abnormalities, and is characterized by the increased presence of intraepithelial lymphocytes (IELs) in the epithelial cell layer. Marsh II is characterized by an increase in IELs as well as crypt hyperplasia (ie. blunting of the usually long, finger-like mucosal villi). In Marsh III, partial or complete villous atrophy is present in addition to increased IELs and crypt hyperplasia (Oberhuber 2000; Marsh 1992). The typical manifestation of CD is often associated with Marsh stages II or III (Guandalini and Assiri 2014).

Other forms of CD are often associated with less acute symptoms which usually manifest after early childhood (Dewar and Ciclitira 2005). The atypical form of CD is characterized by predominantly extra-gastrointestinal symptoms (Telega, Bennet, and Werlin 2008; Guandalini and Assiri 2014). Such symptoms include (but are not limited to): fatigue, headaches, skin rashes (dermatitis herpetiformis), iron deficiency anemia, osteopenia/osteoporosis, bone fractures, arthritis, elevated liver enzymes, unexplained infertility in females, miscarriages, as well as various neurological and psychiatric disorders (Bottaro et al. 1999; Lo et al. 2003; Harper et al. 2007; Guandalini and Assiri 2014). Atypical CD is commonly diagnosed in adults, but it is sometimes found in older children as well. Individuals with atypical CD have positive tTG
antibodies and present with mucosal damage ranging from Marsh I-III (Guandalini and Assiri 2014). The silent form of CD mirrors the atypical form with a few notable exceptions. Individuals with silent CD are asymptomatic by definition, despite testing positive for tTG antibodies and presenting with mild to severe damage to the intestinal mucosa (Marsh I-III) (Ludvigsson, Leffler, et al. 2013; Guandalini and Assiri 2014). Furthermore, silent CD is most frequently diagnosed in adulthood (Guandalini and Assiri 2014). The potential form of CD is characterized by mild to no intestinal- and extra-intestinal symptoms, positive tTG antibodies in most cases, and zero to mild mucosal irregularities (Marsh 0 or I) (Ludvigsson, Leffler, et al. 2013; Husby et al. 2012; Guandalini and Assiri 2014).

Refractory CD occurs when mucosal damage and symptoms of CD persist despite the elimination of gluten from the diet for 6-12 months. This severe manifestation of CD can occur in both individuals with newly diagnosed CD as well as those who previously responded positively to a gluten-free diet (Rubio-Tapia and Murray 2010). The vast majority of individuals diagnosed with refractory CD are >50 years old (Rubio-Tapia, Kelly, et al. 2009; Rubio-Tapia and Murray 2010). It is crucial that the initial diagnosis of CD is correct before a definitive diagnosis of refractory CD is made, as conditions which appear similar to CD, such as various sprue-like intestinal disorders (Daum, Cellier, and Mulder 2005), may be responsible for the lack of response to a gluten-free diet in some cases (Rubio-Tapia and Murray 2010; Biagi and Corazza 2001; Gujral, Freeman, and Thomson 2012). Furthermore, it is imperative that individuals diagnosed with refractory CD are, in fact, adhering to a strict gluten-free diet (Abdulkarim et al. 2002; Leffler et al. 2007). Up to 5% of individuals with CD are thought to represent true cases of refractory CD (Green and Cellier 2007; Gujral, Freeman, and Thomson 2012). These individuals can be further divided into two classes based on their histological findings. Type 1 refractory CD is associated normal IEL counts and patients often respond well
to a combination of stringent gluten restriction, advance nutritional support (eg. consumption of an elemental diet), and select alternative therapies (eg. treatment with various steroids or other immunosuppressive drugs) (Rubio-Tapia and Murray 2010). Type 2 refractory CD is characterized by abnormal IEL presentation and is associated with a less favourable prognosis. Individuals with type 2 refractory CD often do not respond well to the treatments noted above, and enteropathy-associated T-cell lymphoma (EATL) has been shown to develop in many cases (Rubio-Tapia and Murray 2010; Malamut et al. 2009). The risk of developing refractory CD and EATL appears to be greater in those with extended untreated (ie. undiagnosed) CD (Catassi, Bearzi, and Holmes 2005; Gujral, Freeman, and Thomson 2012; Lebwohl, Granath, et al. 2013).

1.3.6 Complications of Untreated Disease

The continued ingestion of gluten in those with CD is associated with a number of adverse health effects. This is due to the sustained presence of an immune response against gluten-derived proteins in the lamina propria which can ultimately result in damage to the small villi of the intestinal mucosa (Kagnoff 2007). The presence of autoantibodies against tTG, a hallmark of CD, may also be independently associated with adverse health effects (Lindfors, Maki, and Kaukinen 2010). Those with undiagnosed CD, which is thought to comprise the vast majority of individuals with CD (Guandalini and Assiri 2014), may be particularly at risk for complications associated with the continued ingestion of gluten. Such complications can range from mild signs of altered absorption of select vitamins and minerals to an increased risk of overall mortality (Goddard and Gillett 2006). Prominent adverse health effects associated with untreated CD
include lipid malabsorption, deficiencies in select micronutrients and minerals, and an increased risk for a number of comorbidities.

Fat malabsorption is prominent in CD as sustained damage to large portions of the intestinal mucosa precludes the absorption of micelles formed in the intestinal lumen (Barker and Liu 2008; Iqbal and Hussain 2009). This can have a number of effects on serum lipid profiles. Indeed, individuals with untreated CD have been shown to have lower levels of total serum cholesterol (Lewis et al. 2009; Ciacci et al. 1999; West et al. 2003). This led researchers to speculate that untreated CD may, in fact, be associated with lower risk for cardiovascular disease (CVD). However, this has not proven to be the case, and hypocholesterolemia in individuals with untreated CD appears to be largely driven by a decrease in HDL-cholesterol but not LDL-cholesterol (Brar et al. 2006). Prospective studies support these findings and have shown that, when individuals with untreated CD begin treatment with a gluten-free diet, HDL- and total cholesterol levels increase, while the ratio of LDL-to-HDL-cholesterol decreases (Brar et al. 2006; De Marchi et al. 2013; Pillan et al. 1994; Capristo et al. 2009). Overall, these studies suggest that untreated CD can lead to unfavorable lipid profiles which may result in increased risk for CVD.

Sustained damage to the intestinal mucosa in individuals with untreated CD can also lead to malabsorption and deficiencies in various vitamins and minerals. Deficiencies in fat-soluble vitamins are common in individuals with newly diagnosed (ie. previously untreated) CD as the absorption of such vitamins is dependent on their incorporation into micelles and subsequent transport through the brush border membrane of enterocytes in the small intestine (Barker and Liu 2008). Vitamin A is an essential fat-soluble vitamin which plays an important role in vision, immune function and fetal development (Gilbert 2013). Vitamin A deficiency has been identified
in case reports of individuals with CD (Alwitry 2000). Furthermore, a recent study of individuals with newly diagnosed CD also found that, on average, those with CD have lower levels of vitamin A as well as an increased risk of vitamin A deficiency compared to healthy controls (Wierdsma et al. 2013). Vitamin D, an important compound in maintaining bone health among other increasingly recognized immune functions (DeLuca 2004), represents another candidate for malabsorption. Indeed, vitamin D deficiency has been shown to be prevalent in certain populations with CD (Lerner et al. 2012). Finally, carotenoids and tocopherols represent important dietary antioxidants with pro-vitamin A and vitamin E activity, respectively. While individuals with CD have been shown to have lower levels of circulating vitamin E (Hozyasz, Chelchowska, and Laskowska-Klita 2003), the association between untreated CD and serum levels of various carotenoids remains unknown.

Deficiencies in a number of water-soluble vitamins and minerals are also prominent in CD. Individuals with CD-associated damage to the intestinal mucosa are likely to present with malabsorption of minerals absorbed in the proximal small intestine, as this represents the primary site of damage in CD (Wierdsma et al. 2013; Barakauskas, Lam, and Estey 2014). Indeed, untreated CD is associated with a higher prevalence of both folate and iron deficiency (Kemppainen et al. 1998; Freeman 2015). The co-occurrence of these deficiencies likely plays a role in the increased risk for anemia with untreated CD, as sufficient levels of both iron and folic acid are required for red blood cell turnover (Dahele and Ghosh 2001). Vitamin B₁₂ is also necessary for red blood cell proliferation (Koury and Ponka 2004), and vitamin B₁₂ deficiency has also been shown to be prominent in untreated CD (Dickey 2002; Dahele and Ghosh 2001). Considering the ileum is the primary site of vitamin B₁₂ absorption, this suggests that ileal villous atrophy may also occur in CD (Dahele and Ghosh 2001). Alternatively, damage to the
proximal small intestine may interfere with pancreatic secretions ultimately required for vitamin B\textsubscript{12} absorption (Hjelt 1991; Dahele and Ghosh 2001).

In addition to altered lipid profiles and an increased risk of deficiency for select vitamin and minerals, untreated CD has been associated with a number of adverse health outcomes. A recent systematic review and meta-analysis found that individuals with CD are at an increased risk of overall CVD (OR 1.10; 95% CI 1.03-1.28). The risk of stroke was significantly higher with CD (OR 1.11; 95% CI 1.02-1.20), and while estimates for myocardial infarction (OR 1.12; 95% CI 0.83-1.40) and cardiovascular death (OR 1.12; 95% CI 0.96-1.29) were in the same direction, they did not reach statistical significance (Emilsson et al. 2015). Risk of osteoporosis may also be elevated in individuals with untreated CD (Stenson et al. 2005). A systematic review and meta-analysis of bone fractures in those with CD found an increased risk of overall fractures (OR 1.30; 95% CI 1.14-1.50), driven specifically by an increased risk of hip fractures (OR 1.69; 95% CI 1.10-2.59), in individuals with CD (Heikkila et al. 2015). In women, undiagnosed CD is considered a risk factor for infertility (OR 3.09; 95% CI 1.74-5.49) (Lasa, Zubiaurre, and Soifer 2014) and miscarriages (RR 1.39; 95% CI 1.15-1.67) in addition to other reproductive disorders (Tersigni et al. 2014). This is likely the result of mucosal damage-induced nutrient malabsorption and potential direct adverse effects of tTG autoantibodies on the placenta and uterus, as no association between CD and any reproductive disorders were observed in those treated with a gluten-free diet (Tersigni et al. 2014). CD has also been shown to be associated with an increased risk of gastrointestinal malignancies (OR 1.60; 95% CI 1.39-1.84) and overall malignancies (OR 1.25; 95% CI 1.09-1.44), while treatment with a gluten-free diet is thought to attenuate such risks (Han et al. 2015). Finally, undiagnosed CD has been associated with an increased risk of all-cause mortality (Rubio-Tapia, Kyle, et al. 2009; Metzger et al. 2006), although not all studies have supported this association (Canavan et al. 2011; Lohi et al. 2009; Godfrey et al. 2010).
1.3.7 Epidemiology

In the past, CD was thought of as a rare disease which primarily affected Caucasian children of European descent (Gujral, Freeman, and Thomson 2012; Fasano and Catassi 2001). Today, CD is recognized as a relatively common autoimmune disorder which affects most populations around the world at a prevalence ranging from 0.5% to 1% (Gujral, Freeman, and Thomson 2012). The recent increased precision in the estimates of worldwide CD prevalence can be attributed to the identification of highly sensitive and specific serologic tests which enable screening in large populations (Barakauskas, Lam, and Estey 2014). Such screening studies have been instrumental in defining the prevalence of CD in various countries, identifying high risk groups, estimating the ratio of diagnosed to undiagnosed cases, and even tracking changes in the prevalence of CD over time.

In North America, just under 1% of individuals in the general population are thought to be affected by CD. In a landmark multicenter screening study conducted in the United States, a total of 13,145 individuals were screened for CD-associated antibodies (Fasano et al. 2003). In the general population (ie. those not considered to be at risk for CD based on symptoms or family history), 1 in 133 (0.75%) individuals were determined to have positive CD serology. A subset of these patients underwent intestinal biopsy and all of these individuals had histological findings consistent with CD. This was the first study to suggest that CD is, in fact, relatively common in the general U.S population and that positive CD-associated antibodies is a reliable indicator of intestinal damage (Fasano et al. 2003). A number of subsequent screening studies in the U.S. have since corroborated these findings (Rubio-Tapia, Kyle, et al. 2009; Rubio-Tapia et al. 2012). Among participants in the nationally representative National Health and Nutrition Examination Survey (NHANES) 2009-2014 cycles, the prevalence of CD was found to be 0.69% (95% CI
0.53-0.84%) (Kim et al. 2016). Differences across the major ethnocultural groups were observed among NHANES participants, with the highest prevalence of CD found in non-Hispanic Caucasians (Mardini, Westgate, and Grigorian 2015; Kim et al. 2016). To date, there have been no screening studies in North America outside of the U.S., and the prevalence of CD in Canada, therefore, remains unknown (Kang et al. 2013).

The prevalence of CD in other regions of the world has been extensively studied. In Europe, screening studies have shown that approximately 1% of the population is likely affected by CD (Kang et al. 2013; Gujral, Freeman, and Thomson 2012). For reasons that are not entirely clear, there appears to be a high degree of variability in the prevalence of CD between European nations, with prevalence estimates ranging from 0.3% (0.1-0.4%) in Germany to 2.4% (95% CI 2.0-2.8%) in Finland (Mustalahti et al. 2010). Generally, the prevalence of CD in Europe seems to follow a latitudinal gradient, with higher estimates reported in northern countries (Gujral, Freeman, and Thomson 2012). There is also considerable variation in the prevalence of CD in Asia. In East Asia (China and Japan), the prevalence of CD is unknown, but it is thought to be extremely rare as only 24 case reports of CD have been identified in these regions (Cummins and Roberts-Thomson 2009; Kang et al. 2013). However, CD is thought to be relatively common in India, with prevalence estimates approaching 1% of the population (Gujral, Freeman, and Thomson 2012). A screening study of 23,331 adults in India illustrated a similar latitudinal gradient in the prevalence of CD as that observed in Europe (Ramakrishna et al. 2016). In northern India, 1.23% of the general population was found to have positive CD antibodies, while the prevalence of CD was found to be 0.87% and 0.10% in northeastern and southern India, respectively. In Australia/New Zealand as well many Middle Eastern nations, a limited number of screening studies have suggested that the prevalence of CD is likely ~1% (Kang et al. 2013; Gujral, Freeman, and Thomson 2012). CD is generally considered rare in South America, and
screening studies in healthy blood donors suggest a range in prevalence from ~0.15% in Brazil to 0.6% in Argentina (Gandolfi et al. 2000; Gomez et al. 2001; Gujral, Freeman, and Thomson 2012). Interestingly, the highest prevalence of CD worldwide has been observed in Saharawi children in North Africa, where 5.6% of the population has been shown to have CD (Catassi et al. 1999). This is in contrast to other African populations, where CD has been shown to affect 0.3-0.6% of individuals (Gujral, Freeman, and Thomson 2012). Figure 1.3 illustrates the approximate prevalence of CD across geographical regions.

In addition to increasing the accuracy of prevalence estimates for CD, screening studies have enabled the identification of individuals with previously undiagnosed CD. Among NHANES participants in the U.S. between 2009 and 2010, 29 out of 35 (83%) individuals with CD were identified as being undiagnosed (Rubio-Tapia et al. 2012). In agreement with these trends in the U.S., a high prevalence of undiagnosed CD has also been reported in European countries (Kang et al. 2013; Mustalahti et al. 2010; Kolho, Farkkila, and Savilahti 1998). For reasons that are currently unknown, the prevalence of undiagnosed CD appears to have been increasing over the past few decades. A recent landmark study in the U.S. identified a 4-fold increase in the prevalence of undiagnosed CD by comparing sera collected in the 1950s to similar present day cohorts (Rubio-Tapia, Kyle, et al. 2009). This trend has also been shown in a number of independent study populations around the world (Ludvigsson, Rubio-Tapia, et al. 2013; Lohi et al. 2007; Kang et al. 2013). Although cases of undiagnosed CD are more likely to present with mild symptoms of CD and mild damage to the intestinal mucosa, the lack of adherence to a gluten-free diet in these cases likely puts such individuals at risk for developing CD-associated complications over the course of their life (see Section 1.3.5). Nevertheless, individuals with undiagnosed CD represent an understudied group and it is unclear whether the extent of their risk for CD-associated complications is similar to those who are ultimately diagnosed with CD.
Figure 1.3: Worldwide prevalence of celiac disease.

Adapted from Gujral et al. 2012 and Dieli-Crimi et al. 2015.

The considerable variation in the prevalence of CD between different populations around the world is likely the result of an interplay between a number of factors. Indeed, like most complex autoimmune diseases, CD is influenced by both genetic predisposition and an environmental trigger (Green and Cellier 2007). Variation in the prevalence of HLA-DQ2/8 alleles closely mirrors that of CD in various populations. In the U.S. and many European nations, the prevalence of HLA-DQ2 is approximately 20% or greater. An additional 20% of individuals have been shown to carry the HLA-DQ8 heterodimer (Dieli-Crimi, Cenit, and Nunez 2015; Gujral, Freeman, and Thomson 2012). In contrast, the prevalence of HLA-DQ2 and -DQ8 alleles in China and Japan has been shown to be approximately <10% and <5%, respectively (Gujral,
In India, up to 40% of individuals carry HLA-DQ2/8 alleles, which is approximately equal to the prevalence of such alleles in the U.S. and Europe (Gujral, Freeman, and Thomson 2012). Data on per capita wheat consumption between nations indicates that, in addition to genetic predisposition, average consumption of gluten is associated with the population prevalence of CD. Compared to average wheat consumption in the U.S. and Europe, wheat consumption in East Asia (China and Japan) is relatively low as rice is the staple grain in these nations (Cummins and Roberts-Thomson 2009; Kang et al. 2013). Contrastingly, per capita wheat consumption in India is relatively high in certain regions, likely explaining why estimates of the prevalence of CD in northern India closely resemble those observed in the U.S. and Europe (Gujral, Freeman, and Thomson 2012). Remarkably, regional variation in wheat consumption within India has even been associated with local differences in the prevalence of CD between northern (high wheat consumption and high prevalence of CD) and southern (low wheat consumption and low prevalence of CD) regions (Ramakrishna et al. 2016). It is clear that the interaction between HLA risk genotypes and gluten consumption has a major influence on the population prevalence of CD, but there are a number of other factors which are also involved.

There has been considerable interest in identifying environmental factors which may trigger the development of CD in genetically susceptible individuals and further explain differences in the prevalence of CD observed between populations. Differences in the composition of the gut microbiota have been observed in those with CD. Individuals with both treated and untreated CD have been shown to have more Bacteroides species and less Bifidobacterium species than healthy controls (Sarno et al. 2015; Marasco et al. 2016). Such changes in the gut microbiota has been postulated to activate inflammatory pathways and potentially play a role in the pathogenesis of CD (Marasco et al. 2016). While it is unclear whether such changes represent a cause or effect
of CD-associated mucosal damage in humans, animal studies have suggested that changes in the intestinal microbiota may indeed be a risk factor for the development of CD (Galipeau et al. 2015; Caminero et al. 2016). Furthermore, supplementation with certain probiotics has been shown to alleviate symptoms of CD in affected individuals (Smecuol et al. 2013), likely through effects on the innate immune system (Pinto-Sanchez, Smecuol, et al. 2016). The presence of certain viral infections during the time of initial gluten exposure in at-risk infants may also be a risk factor for CD (Kagnoff 2007; Sarno et al. 2015). Indeed, a high frequency of rotavirus infections in early childhood was associated with an increased risk of developing CD in a U.S.-based prospective cohort study (Stene et al. 2006). It is possible that such viral infections could increase intestinal permeability and prime the immune system to mount an adverse physiological response to gluten (Kagnoff 2007). Finally, although observational studies initially suggested that the delayed termination of breastfeeding (Akobeng et al. 2006) and first introduction of gluten between 4 and 6 months of age would reduce the risk of developing CD (Norris et al. 2005), recent randomized-controlled trials have suggested that these factors do not influence the risk of developing CD in genetically susceptible infants (Lionetti et al. 2014; Vriezinga et al. 2014). A recent meta-analysis of both observational studies and randomized-controlled trials concluded that the late introduction of gluten (>6 months) may be a risk factor for the development of CD, while no significant effect of breastfeeding on risk for CD was found (Pinto-Sanchez, Verdu, et al. 2016).
1.4 Non-Celiac Gluten Sensitivity

Non-celiac gluten sensitivity (NCGS) is defined as adverse reactions to gluten in the absence of villous atrophy and the hallmark IgA antibodies associated with CD. Individuals with NCGS are thought to experience many of the same symptoms upon the ingestion of gluten as patients with CD, such as diarrhea, abdominal pain, fatigue and headaches (Catassi et al. 2013; Lundin and Alaedini 2012). Although case reports of adverse reactions to gluten in the absence of CD have appeared as early as the late 1970s (Ellis and Linaker 1978; Cooper et al. 1980), considerable interest in such potential adverse effects in individuals without CD has only recently developed. This increased interest coincides with the rise in popularity of the gluten-free diet among the general public (Sapone et al. 2012). Emerging research aims to identify whether gluten is, indeed, responsible for causing gastrointestinal and extra-gastrointestinal symptoms in a subset of individuals without CD and, if so, the pathophysiological mechanisms that are responsible for these symptoms.

A number of randomized-controlled trials have investigated whether gluten is capable of triggering gastrointestinal symptoms in individuals without CD. An initial parallel design study in 34 patients with irritable bowel syndrome (IBS) for which CD had been conclusively ruled out found that, after 6 weeks, those receiving gluten had significantly worse gastrointestinal symptoms, including pain, bloating, poor stool consistency and tiredness compared to those receiving a placebo (Biesiekierski et al. 2011). A crossover trial in adults without CD (n=61) who reported symptomatic improvement on a gluten-free diet found that gluten led to a significant increase in both gastrointestinal (abdominal bloating and pain) and extra-gastrointestinal symptoms (foggy mind, depression and aphthous stomatitis). However, these
adverse responses to gluten were largely driven by a limited number of subjects and were not experienced by the majority of individuals enrolled in the study (Di Sabatino et al. 2015). In line with these findings, two additional randomized crossover trials of gluten reintroduction in those believed to have NCGS found that only a subset of patients (34% and 14%) had a worsening of gastrointestinal symptoms when gluten was reintroduced to the diet (Zanini et al. 2015; Elli et al. 2016). While these findings suggest that some individuals without CD may experience adverse symptomatic responses to gluten, other components of the diet may also play a role in the development of gastrointestinal symptoms.

It is possible that other components of gluten-containing foods are partially responsible for causing symptoms in individuals with apparent gluten sensitivity. Adverse effects of gluten intake in those with perceived gluten-related symptoms were not observed in a trial where subjects were on a low fermentable, oligo-, di-, monosaccharides, and polyols (FODMAPs) diet, suggesting that other dietary factors may explain clinical symptoms in individuals with perceived NCGS (Biesiekierski et al. 2013). Other components of wheat and related grains thought to possibly play a role in the development of symptoms in individuals without CD include amylase trypsin inhibitors (ATIs) (Huebener et al. 2015; Lebwohl, Ludvigsson, and Green 2015). ATIs have been shown to activate innate immune responses in cells from biopsies of CD patients as well as individuals without CD, and it is possible that these non-gluten components of wheat, barley and rye could explain symptoms in some individuals with perceived sensitivity to gluten (Junker et al. 2012). Further research is necessary to better understand how gluten interacts with other components of the diet and ultimately leads to the development of various symptoms in individuals with potential NCGS.
1.4.1 Diagnosis

The lack of any known biomarkers of NCGS has made the diagnostic process for this condition difficult to standardize. It is paramount that individuals with suspected NCGS undergo screening for other disorders which may resemble this condition (Catassi et al. 2015). CD and wheat allergy represent two well-defined conditions which must be ruled out before a diagnosis of NCGS is made. To rule out CD, a combination of serologic and genetic tests, as well as upper endoscopy and biopsy of the intestinal mucosa should be performed (see CD diagnostic cascade in Figure 1.2). Wheat allergy is an IgE-mediated food allergy, and this condition can be ruled out based on the results of a negative skin prick test and negative serologic IgE markers (Sapone et al. 2012). In both cases, patients must be on a typical gluten-containing diet prior to testing in order to reduce the likelihood of a false negative test result. Once CD and wheat allergy have been conclusively ruled out, a diagnosis of NCGS can be explored based on the disappearance and recurrence of symptoms when on a gluten-free and gluten-containing diet, respectively (Tonutti and Bizzaro 2014).

Recent guidelines have been developed to offer a standardized approach in the diagnosis of NCGS after the exclusion of CD and wheat allergy have been made. These guidelines advise that the diagnosis of NCGS be carried out in two general steps (Catassi et al. 2015). The first step simply requires patients to show symptomatic improvement on a gluten-free diet. Before initiating a gluten-free diet, patients must be on a standard gluten-containing diet for a minimum of 6 weeks. At this point, a diagnostic questionnaire is administered to establish baseline symptom scores. The questionnaire is a modified version of the Gastrointestinal Symptom Rating Scale (GSRS), an instrument commonly used in the assessment of gastrointestinal symptoms of many disorders (Kulich et al. 2008). Modifications to the standard GSRS allow for
common extra-gastrointestinal symptoms associated with NCGS to also be evaluated (Catassi et al. 2015). After consultation with a registered dietitian, a strict gluten-free diet is initiated for a minimum of 6 weeks. The modified GSRS is then administered in weekly intervals throughout the gluten-free diet to track changes in patients’ symptom scores. If one of any three main symptoms identified at baseline (on a gluten-containing diet) decrease by at least 30% in at least 3 of the weekly evaluations, a patient is considered to respond favourably to a gluten-free diet and can progress to the second step. Step two is a double-blind placebo-controlled crossover gluten challenge. Before initiating the challenge, patients must again adhere to a gluten-free diet for a minimum of 4 weeks. At this point, patients will begin the first phase of the gluten challenge. In a random and blinded fashion, they will receive either gluten (8g/day) or placebo in a whole food vehicle for one week. This is followed by a one week gluten-free washout period. Finally, the patient will then receive either gluten or placebo (whichever they did not receive initially) for the final week. Patients will fill out the modified GSRS at baseline and daily throughout the gluten challenge, and an increase in any one of three main symptoms in the gluten arm of the challenge compared to the placebo arm would lead to a diagnosis of NCGS (Catassi et al. 2015). A summary of the NCGS diagnostic cascade is shown in Figure 1.4.
**Figure 1.4:** Non-celiac gluten sensitivity diagnostic cascade.

Adapted from Sapone et al. 2012 and Catassi et al. 2015.

The current approach to diagnosing NCGS relies exclusively on the demonstration of symptomatic relief on a gluten-free diet and the recurrence of symptoms upon gluten reintroduction. This represents a rather complicated diagnosis process which has proven to be difficult to standardize. The use of specific biomarkers to both assess the risk of NCGS while on a gluten-containing diet and monitor the effects of gluten elimination and reintroduction would be useful in the clinical evaluation of NCGS; however, to date, there are no validated biomarkers for assessing the condition. A number of studies have shown that approximately 50% of individuals that fulfill the diagnostic criteria for NCGS will test positive for anti-gliadin antibodies in the IgG class (Sapone et al. 2011; Catassi et al. 2015; Volta et al. 2012; Caio et al. 2014). While this may be indicative of specific gluten-restricted effects in a subset of individuals...
fulfilling the diagnostic criteria for NCGS, the low sensitivity of such antibodies for NCGS (~50% of individuals with possible NCGS test negative) precludes its use as a laboratory biomarker for diagnosis. The identification of novel biomarkers for NCGS is currently an area of active research.

1.4.2 Prevalence

The lack of standardized diagnostic criteria for NCGS has made studying the prevalence of such a condition particularly challenging. This had led to considerable uncertainty in estimates of the prevalence of NCGS. While the popularity of gluten-free foods has greatly risen over the past decade (Sapone et al. 2012; Fasano et al. 2015), it is unlikely that all individuals consuming such foods fulfil the diagnostic criteria for NCGS. Indeed, self-reported positive responses to a gluten-free diet appears to be a poor predictor for a diagnosis of NCGS. In a prospective multicenter study in Italy which screened over 12,000 individuals who experienced self-reported adverse responses to gluten, only 4% of these individuals ultimately received a diagnosis of NCGS (Volta et al. 2014). Another prospective study in Italy found that the diagnostic criteria for NCGS was fulfilled in 27 out of 392 patients (~7%) who reported gluten-related symptoms (Capannolo et al. 2015). Out of 5,896 patients seen at the Center for Celiac Research and the University of Maryland, USA between 2004 and 2010, 6% fulfilled the diagnostic criteria for NCGS (Sapone et al. 2012). However, this estimate likely represents an overestimation compared to the general population considering all individuals screened were patients of a tertiary care center for gluten-related disorders.
More accurate estimates of the true prevalence of NCGS may come from comparing the diagnosis rate of NCGS to that of CD. In the two aforementioned prospective studies conducted in Italy, both found that the ratio of patients diagnosed with NCGS to CD was approximately 1:1, suggesting that NCGS and CD may have a similar prevalence of about 1% in the general population (Volta et al. 2014; Capannolo et al. 2015). Whether the prevalence of NCGS follows the same geographic distribution as CD remains unknown. Furthermore, while NCGS seems to primarily affect women and middle aged adults (Fasano et al. 2015), risk factors for developing NCGS have not been well established. Clearly, more research is necessary to improve our understanding of NCGS and what fraction of the general population may be affected by such a condition.

1.4.3 Potential Mechanisms

The mechanisms by which gluten may cause adverse gastrointestinal and extra-gastrointestinal symptoms in individuals without CD remain elusive. Unlike CD, individuals with potential NCGS do not develop villous atrophy as a result of gluten ingestion. There appears to be an increase in IELs in the mucosal cell layers in some patients with NCGS compared to healthy controls and those with well managed CD (Brottveit et al. 2013; Sapone et al. 2011; Sapone et al. 2010); however, it is unclear why such a response might develop. The HLA-DQ-mediated adaptive immune response which characterizes CD is absent in NCGS (Sapone et al. 2011), suggesting the potential involvement of other pathways. The identification of such pathways remains an active area of research, with many studies focusing on intestinal permeability and the
innate branch of the immune system to gain a better understanding of how gluten may ultimately affect those with possible NCGS (Lebwohl, Ludvigsson, and Green 2015).

It is thought that an increase in intestinal permeability and activation of the innate immune system may be more prominent in individuals with NCGS. Altered expression of certain genes involved in innate immunity, including an upregulation of Toll-like receptor 2, have been observed in mucosal biopsy specimens of individuals with NCGS (Sapone et al. 2011). While an early study of patients with potential gluten sensitivity observed a decrease in intestinal permeability in those with potential NCGS (Sapone et al. 2011), subsequent studies have suggested otherwise. In patients with diarrhea-prominent IBS, a gluten-containing diet has been shown to result in higher intestinal permeability than a gluten-free diet (Vazquez-Roque et al. 2013). An ex-vivo study of human duodenal biopsy samples found that individuals with potential NCGS had a greater increase in intestinal permeability in response to exposure to gluten-derived peptides than those with controlled CD, but not greater than biopsies from controls without CD (Hollon et al. 2015). Furthermore, there was an increase in IL-10, an anti-inflammatory cytokine and important regulator of innate immunity (Moore et al. 2001), in response to gluten exposure in the control samples, but not those with NCGS. This suggests that gluten may cause increased intestinal permeability in those with NCGS which could potentially be caused and/or accentuated by an altered innate immune response (Hollon et al. 2015).

A recent study in those with NCGS investigated the possible link between immunogenic products produced by gut microbes, intestinal permeability and immune activation (Uhde et al. 2016). They found that gluten caused a significant increase in serum levels of soluble CD14 and lipopolysaccharide (LPS)-binding protein in addition to antibodies against LPS and flagellin, two microbial antigens, in those with NCGS. Furthermore, they found that fatty acid-binding protein
2, a marker of compromised epithelial cell integrity, was significantly elevated in response to a gluten challenge, but normalized when patients were on a gluten-free diet. Overall, these results suggest that increased intestinal permeability in response to gluten likely results in the translocation of microbes and microbial products through epithelial cells and into the systemic circulation where they elicit an immune response in NCGS (Uhde et al. 2016). While the molecular mechanisms behind this increase in intestinal permeability remain unknown, this represents one of the first studies to identify markers of immune activation and damage to the intestinal epithelial barrier in NCGS which appear to normalize with the resolution of symptoms on a gluten-free diet. Further studies are necessary to determine whether such markers could be useful in the diagnosis of NCGS.

Many aspects of the pathophysiology and patient population affected by NCGS remain unknown. Unlike CD, it is unclear whether lifelong adherence to a strict gluten-free diet is required in NCGS (Fasano et al. 2015; Catassi et al. 2015). It is possible that those with NCGS experience a transient sensitivity to gluten, and the periodic re-introduction of gluten back into the diet of these patients is currently recommended (Catassi et al. 2015). Also, while gut dysmotility is prominent in CD, the presence of motility alterations and the potential mechanisms behind them are currently unknown in NCGS (Pinto-Sanchez, Bercik, and Verdu 2015). Finally, it is unclear whether individuals with NCGS are at risk for similar long-term complications with the continued ingestion of gluten as those with CD (Fasano et al. 2015). Clear diagnostic criteria for NCGS and well-designed observational studies are required to answer such questions. A better understanding of the physiological effects of gluten intake in those without CD may also help elucidate potential processes that may become dysregulated in NCGS. Recent advances in omics technologies allow for the identification of novel markers linking specific dietary
exposures to various biological processes, and the utilization of an omics approach may be useful in improving our understanding of the effects of gluten intake in individuals without CD.

1.5 Omics Technologies

1.5.1 Proteomics

Plasma contains many diverse proteins and is considered to be the largest version of the human proteome (Anderson 2005). Although upwards of 3000 different proteins have been identified in human plasma, a relatively small number of them account for the majority of the total protein mass (Hortin, Sviridov, and Anderson 2008). As few as 20 proteins make up over 99% of the total protein mass in circulating plasma (Lee et al. 2006), and concentrations of the 150 most abundant plasma proteins span several orders of magnitude (Hortin, Sviridov, and Anderson 2008).

Highly-abundant plasma proteins include various physiologically important molecules and are useful biomarkers of disease and dietary exposures (Garcia-Bailo et al. 2012). Certain plasma proteins have been identified as biomarkers for systemic inflammation and have been used as indicators of risk for CVD (Hemingway et al. 2010; Rifai and Ridker 2003). The composition of plasma proteins is thought to differ under various physiological conditions, and it is thought that most metabolic abnormalities in the body are reflected in the plasma proteome (Anderson 2005). Assessing a panel of highly abundant plasma proteins may, therefore, be useful in evaluating disease risk and health responses to certain dietary exposures like gluten. Indeed, it has previously been shown that the plasma proteome can be used to identify physiological pathways
associated with circulating concentrations of ascorbic acid (vitamin C) (Da Costa et al. 2013b), α-tocopherol (vitamin E) (Da Costa et al. 2013a), 25-OH-vitamin D (Garcia-Bailo, Jamnik, Da Costa, Borchers, et al. 2013), dietary caffeine (Tian et al. 2013), and hormonal contraceptive use (Josse et al. 2012).

In the past, accurately measuring the concentration of more than a few proteins has been difficult. However, high throughput analytical platforms now enable the simultaneous measurement of multiple proteins along diverse physiological pathways (Kuzyk et al. 2009). This allows for the combination of multiple biomarkers and increases their predictive power (Garcia-Bailo et al. 2012). A high throughput multiple reaction monitoring (MRM), mass spectrometry-based proteomic assay has been used to measure circulating levels of 54 plasma proteins involved in inflammation, endothelial cell function and other pathways (Kuzyk et al. 2009). It is, therefore, possible that physiological effects of gluten intake in individuals without CD are reflected in the concentrations of these proteins.

1.5.2 Genomics

Improvements in genome sequencing techniques and SNP genotyping have made large-scale genome-wide association studies (GWAS) possible (Hirschhorn and Daly 2005). SNPs represent substitution mutations of single base pairs in the DNA duplex which comprise the majority of observed inter-individual genetic variation (Sherry et al. 2001). In order to use SNPs as a measure of genetic variation across the entire genome, GWAS take advantage of linkage disequilibrium (LD) between various SNPs and, by extension, their associated genes. Genes in LD are likely close in proximity along a chromosome and usually segregate together due to a
lack of recombination during meiosis. LD results in common gene groupings, or haplotypes, that fragment the genome and remain intact across multiple generations. Therefore, by sequencing SNPs across haplotypes, genetic information about the rest of the genome can be inferred. It has been suggested that a few hundred thousand SNPs can sufficiently represent common genetic variation across a genome (Hirschhorn 2005).

GWAS compare genome-wide genetic variants in multiple subjects to a particular phenotype of interest in order to identify variants that are significantly associated with the phenotypic trait. Unlike candidate gene studies, which investigate a limited number of genes with purported links to the trait of interest, GWAS search for associations between a phenotype and genes across the entire genome in an unbiased manner. Thus, GWAS approaches are able to identify novel genes that are associated with certain traits (Hirschhorn and Daly 2005). To date, GWAS have identified many novel genes associated with common diseases including type 1 and type 2 diabetes, inflammatory bowel disease, various cancers, Alzheimer’s disease, asthma, CVD, atrial fibrillation and CD (Iles 2008; McCarthy et al. 2008; Dieli-Crimi, Cenit, and Nunez 2015). Although most studies to date have directly focused on identifying genetic determinants of disease, it has been argued that such GWAS designs lack the power to detect certain causal variants of common diseases (Iles 2008).

Certain GWAS approaches have indirectly studied disease-associated variants by investigating pathways that may become dysregulated during disease progression. Using an approach that investigates physiological pathways rather than directly observing disease phenotypes increases the power of GWAS analyses as intermediate phenotypes are accounted for. Such approaches also allow for the disease-causing mechanisms to be better understood (Gieger et al. 2008). For example, a GWAS of human urine metabolites identified a number of genes involved in
pathways that become dysregulated during the progression of various detoxification-related disorders (Suhre et al. 2011). Other GWAS on serum metabolite concentrations have identified multiple genetic variants associated with altered metabolic states (Gieger et al. 2008; Illig et al. 2010). Such an approach could be useful for identifying genes involved in physiological pathways associated with response to dietary exposures like gluten.

1.6 Summary of Literature Review

Gluten is a protein found in wheat, barley and rye. Celiac disease (CD) is characterized by adverse autoimmune reactions triggered by the ingestion of gluten, and individuals with CD must adhere to a strict gluten-free diet. Variation in the human leukocyte antigen (HLA)-DQ locus is known to confer susceptibility to CD, with virtually all of those diagnosed possessing specific HLA-DQ2 or DQ8 gene variants. Studying the prevalence of CD remains challenging due to the high number of undiagnosed cases. While there are various comorbidities associated with continued gluten intake in those with diagnosed CD, it is unclear whether individuals with undiagnosed CD are at the same risk for impaired nutritional status and other adverse health effects associated with mucosal damage and malabsorption. Furthermore, while the prevalence of CD has been defined in many regions, differences in the prevalence of CD across certain ethnic groups remains unknown, and no screening-based studies to date have assessed the prevalence of CD in Canada.

Gluten sensitivity in the absence of CD is also thought to exist, and such a condition has been termed non-celiac gluten sensitivity (NCGS). Many aspects of NCGS remain unclear, including its pathogenesis, prevalence and even the diagnostic approach to such a condition. This is
partially due to the lack of biomarkers for assessing the physiological effects of gluten intake in those without CD. Advances in omics technologies enable the identification of biomarkers that are associated with nutritional exposures, and allow for the understanding of pathways that are associated with response to such exposures. Highly-abundant plasma proteins include various physiologically important molecules, and assessing a panel of these proteins may be useful in evaluating biological responses to dietary gluten intake in those without CD.

1.7 Hypothesis and Objectives

1.7.1 Hypothesis

The overall aims of this thesis are to investigate both the prevalence of CD seropositivity and predisposing risk factors in a population of ethnoculturally diverse Canadian adults, as well as to examine physiological pathways associated with gluten intake in those without CD, which may become dysregulated in individuals with NCGS.

Regarding CD, the hypothesis is that the prevalence of positive serology in Canada will be similar to the approximate prevalence of 1% observed in the U.S. and Europe, with the majority of likely CD cases being undiagnosed. Furthermore, the prevalence of CD, predisposing HLA-DQ2/8 high risk alleles, and average gluten intake will differ between ethnocultural groups in a Canadian population. Finally, individuals with likely undiagnosed CD will present with less favourable cardiometabolic risk profiles and have lower concentrations of circulating markers of nutritional status than those without CD.
1.7.2 Objectives

1. To determine the prevalence of positive CD serology in a population of Canadian adults, and determine whether the prevalence of CD seropositivity, HLA-DQ2/8 high risk alleles and average gluten intake differ between major ethnocultural groups.

2. To determine whether biomarkers of cardiometabolic health and nutritional status differ between Canadian adults with positive and negative CD serology.

3. To examine the association between gluten intake and plasma proteomic biomarkers in a population of young adults without CD.

4. To identify genetic variants that are associated with plasma proteomic biomarkers of gluten intake using a genome-wide approach and determine whether any such variants modify the association between gluten intake and plasma protein concentration.
CHAPTER 2:

PREVALENCE OF POSITIVE CELIAC DISEASE SEROLOGY, HLA RISK GENOTYPES AND AVERAGE GLUTEN INTAKE IN A MULTIETHNIC POPULATION OF ADULTS IN CANADA

Adapted from:

Chapter 2: Prevalence of Positive Celiac Disease Serology, HLA Risk Genotypes and Average Gluten Intake in a Multiethnic Population of Adults in Canada

2.1 Abstract

**Background:** Celiac disease (CD) is a complex autoimmune disorder with known genetic risk factors. Approximately 1% of individuals of European ancestry have CD, but the prevalence among different ethnicities living in Canada remains unknown.

**Objective:** The objective was to determine the prevalence of positive CD serology in a population of Canadian adults living in Toronto, and to determine whether the prevalence of CD, predisposing HLA-DQ2/DQ8 risk genotypes, and average gluten intake differ between major ethnocultural groups.

**Methods:** Subjects (n=2,832) were participants from the Toronto Nutrigenomics and Health Study and the Toronto Healthy Diet Study. Prevalence of CD was determined by positive serology testing for anti-tissue transglutaminase (tTG) antibodies. HLA genotypes were determined using 6 single nucleotide polymorphisms in the HLA gene region. Gluten intake was estimated using a 196-item semi-quantitative food frequency questionnaire.

**Results:** Of the 2,832 individuals screened, a total of 25 (0.88%; 95% CI, 0.57-1.30%) were determined to have positive CD serology. The majority of likely CD cases were undiagnosed (87%). Prevalence was highest among Caucasians (1.48%; 95% CI, 0.93-2.23%), but this did not differ from those of ‘Other’ (0.74%; 95% CI, 0.09-2.63%) or ‘Unknown’ (0.43; 95% CI 0.01-2.36%) ethnicity. No cases of positive CD serology were identified among East Asian or South Asian individuals. East Asians had a lower prevalence of high risk HLA genotypes and lower average gluten intake than Caucasians and South Asians (p<0.05).
**Conclusions:** The prevalence of positive CD serology among Canadian adults living in Toronto is approximately 1%, with 87% of likely CD cases being undiagnosed. Caucasians comprise the majority of likely CD cases, and differences in HLA-DQ2/8 risk genotypes and gluten intake may explain a portion of the variation in CD prevalence across ethnocultural groups. These findings suggest the need for better screening in high genetic risk groups.
2.2 Introduction

Celiac disease (CD) is an autoimmune disorder with defined genetic risk factors. Human leukocyte antigen (HLA)-DQ2 or -DQ8 alleles are considered necessary for the development of CD as virtually all affected individuals possess these genetic variants (Dieli-Crimi, Cenit, and Nunez 2015; Karell et al. 2003; Pietzak et al. 2009). Dietary exposure to gluten, a protein found in wheat, barley and rye, triggers adverse autoimmune reactions in affected individuals. Damage to the intestinal mucosa that is characteristic of CD can ultimately result in nutrient malabsorption, and the only effective treatment to date is strict adherence to a gluten-free diet (Green and Cellier 2007). Symptoms of CD include gastrointestinal discomfort, malnutrition, steatorrhea, irritability and depression, although some individuals may be relatively asymptomatic (Kagnoff 2007). If untreated, individuals with CD may be at an increased risk for various nutrient deficiencies (Caruso et al. 2013), osteoporosis (Vasquez et al. 2000), infertility (Tersigni et al. 2014), certain gastrointestinal lymphomas (Catassi, Bearzi, and Holmes 2005) and overall mortality (Rubio-Tapia, Kyle, et al. 2009).

Diagnosis of CD is done using a combination of serologic tests and biopsy. Initially, individuals reporting symptoms consistent with the clinical manifestation of CD will undergo serologic screening for anti-tissue transglutaminase (tTG) IgA or anti-endomysial IgA (EMA) antibodies (Nandiwada and Tebo 2013). The autoantigen for both anti-tTG and EMA antibodies is the tissue transglutaminase (Dieterich et al. 1997), an important enzyme in the pathogenesis of CD. tTG and EMA IgA antibodies are highly sensitive and specific for CD; however, selective IgA deficiency, which affects up to 5% of individuals with CD, can lead to false negative serology (Halfdanarson, Litzow, and Murray 2007). As a result, screening for total IgA levels recommended, and tTG IgG is assessed in individuals with selective IgA deficiency (Nandiwada
A positive result for any of the above antibodies would result in a recommendation for biopsy, which remains the gold standard for diagnosing CD. Genetic screening for the presence of HLA-DQ2/DQ8 can be used to rule out cases of CD (Gujral, Freeman, and Thomson 2012).

Approximately 1% of individuals in the United States and many European populations are affected by CD (Fasano et al. 2003; Mardini, Westgate, and Grigorian 2015; Mustalahti et al. 2010; Kang et al. 2013; Rubio-Tapia, Kyle, et al. 2009; Kim et al. 2016). Of particular concern is that the prevalence of CD has been shown to be on the rise (Lohi et al. 2007; Ludvigsson, Rubio-Tapia, et al. 2013; Kang et al. 2013; Rubio-Tapia, Kyle, et al. 2009). The prevalence of CD in East Asian populations is much lower than in Caucasians with only 24 case-reports identified in China and Japan (Kang et al. 2013), but CD is more common in individuals of South Asian descent (Ramakrishna et al. 2016; Gujral, Freeman, and Thomson 2012; Cummins and Roberts-Thomson 2009). Variation in gluten intake and the prevalence of HLA-DQ2/DQ8 risk alleles are thought to explain some of the regional variation in CD prevalence (Cummins and Roberts-Thomson 2009; Gujral, Freeman, and Thomson 2012); however, the extent to which such factors influences the prevalence of CD in immigrant populations is unclear. Furthermore, the prevalence of CD among Canadian adults, including those of various ethnocultural backgrounds, remains unknown. The objective of this study was to determine the prevalence of CD seropositivity in a population of Canadian adults, and to determine whether the prevalence of CD seropositivity, predisposing HLA-DQ2/DQ8 risk genotypes and average gluten intake differ between major ethnocultural groups.
2.3 Methods

2.3.1 Study Populations

2.3.1.1 Toronto Nutrigenomics and Health Study

The Toronto Nutrigenomics and Health (TNH) Study is a cross-sectional cohort of young adults aged 20-29 years living in Toronto, Canada. Subjects (n=1,634) were recruited from the University of Toronto Campus between October 2004 and December 2010. Pregnant or breastfeeding women were excluded from the study. All individuals completed a general health and lifestyle questionnaire (GHLQ), had anthropometric measurements taken, and provided a fasting blood sample for various biomarker assessments and for DNA isolation. The ethnocultural background of each subject was ascertained by self-report. Individuals were classified as either Caucasian (European, Middle Eastern, or Hispanic), East Asian (Chinese, Japanese, Korean, Filipino, Vietnamese, Thai, or Cambodian), South Asian (Bangladeshi, Indian, Pakistani, or Sri Lankan) or Other (Aboriginal Canadians, Afro-Caribbean, or individuals belonging to two or more distinct ethnocultural groups) (Garcia-Bailo et al. 2012). A total of 1,620 subjects (1,103 women and 517 men) with available plasma samples were included in the study, which was approved by the Research Ethics Board at the University of Toronto.

2.3.1.2 Toronto Healthy Diet Study

The Toronto Healthy Diet (THD) Study is a randomized, controlled trial designed to investigate whether increased consumption of fruits, vegetables and whole grains would lead to a reduction in body weight and improvement in biomarkers of obesity-related chronic disease (Jenkins et al. 2016). The study protocol was registered with ClinicalTrials.gov (NCT00516620) and was
approved by the Research Ethics Boards at the University of Toronto and St. Michael’s Hospital. Households in the Toronto region with at least one person having a body mass index (BMI) $\geq 25$ kg/m$^2$ were recruited into the study. Individuals with chronic conditions, including diagnosed CD, were excluded from the study. A total of 1,245 adults 18-82 years of age provided baseline data, which included anthropometric measurements and a fasting blood sample. The ethnocultural background of participants was assessed by self-report and individuals were classified into the same groups described above for the TNH study. An additional ‘Unknown’ ethnocultural group was included for the THD Study where data on ethnicity was missing (n=234). All 1,212 individuals (901 women and 311 men) with baseline plasma samples available were included in the study.

2.3.2 Genotyping

HLA genotyping for subjects in both study populations was performed using TaqMan™ allelic discrimination assays (Applied Biosystems®) at the Analytical Genetics Technology Centre at the Princess Margaet Hospital, University Health Network, Toronto, Canada. Genotyping was performed for 6 single nucleotide polymorphisms (SNPs) in the HLA region (rs7454108, rs2395182, rs7775228, rs4713586, rs2187668, and rs4639334), which collectively determine the presence of HLA-DQ2/DQ8 alleles associated with CD (Monsuur et al. 2008). Subjects were categorized into risk groups for CD based on an HLA gradient that has previously been established (Pietzak et al. 2009). Subjects with DQA1*0501-DQB1*0201 (DQ2.5), DQA1*0301-DQB1*0302 (DQ8) or DQA1*0505-DQB1*0301/DQA1*0201-DQB1*0202 (DQ2.2/DQ7) were considered to be at ‘elevated risk’ for CD compared to the general
population (Pietzak et al. 2009; Dieli-Crini, Cenit, and Nunez 2015). A total of 2,832 subjects underwent genotyping and were included in all analyses.

2.3.3 Celiac Disease Assessment

Anti-tissue transglutaminase (tTG) antibodies from plasma samples stored at -80°C were measured using human-recombinant tTG enzyme-linked immunosorbent assays (ELISAs). Only subjects classified as ‘elevated risk’ for CD based on their HLA-DQ genotype plus all those with at least one copy of DQA1*0201-DQB1*0202 (DQ2.2) had their plasma analyzed (n=1,555), since virtually all individuals with CD possess such alleles (Pietzak et al. 2009). Initially, samples were analyzed using dual-isotope transglutaminase kits screening for tTG antibodies in both the IgA- and IgG-class (product no. ORG540S; Orgentec Diagnostika, Main, Germany). This was done to account for individuals with selective IgA deficiency (Reeves et al. 2006), the primary cause of false negative test results with tTG IgA screening (Leffler and Schuppan 2010). The threshold for a positive test result was ≥ 15 U/ml. All samples that tested positive were then analyzed using a separate kit for tTG IgA (product no. ORG540A, Orgentec Diagnostika; Main, Germany). The threshold for a positive test result for tTG IgA was ≥ 10 U/ml. Individuals who tested positive for tTG IgA antibodies were considered to have positive CD serology. Individuals who tested negative for tTG IgA were screened for IgA deficiency using ELISA (product no. CSB-E07985h, Cusabio Biotech Co., Wuhan, China). Individuals with IgA deficiency (total IgA < 7 mg/dl) (Bonilla et al. 2015) were then screened using a separate kit for tTG IgG (product no. ORG540G; Orgentec Diagnostika, Main, Germany). Those with IgA deficiency and positive tTG IgG (≥ 10 U/ml) were considered to have positive CD serology. Equivocal results were assigned
if the dual-isotope screen was positive, but individuals subsequently tested negative for tTG IgA antibodies and were IgA-sufficient. Individuals in the TNH population who reported having been diagnosed with CD on the general health and lifestyle questionnaire were also considered to have positive CD serology.

2.3.4 Dietary Assessment

Individuals in both study populations completed the Toronto-modified Willett 196-item semi-quantitative FFQ. Subjects were provided with instructions on how to complete the FFQ using visual aids of portion sizes. Gluten intake was estimated by adding subjects’ responses to gluten-rich food items. Gluten-rich foods were identified as foods that fell into one of the following categories: bread, baked goods, cereal, and pasta (van Overbeek et al. 1997). The vegetable protein fraction per serving of each gluten-rich food item was estimated using the United States Department of Agriculture National Nutrient Database for Standard Reference and Harvard FFQ Databases (USDA 2014). This value was then multiplied by a factor of 0.8 to obtain the approximate amount of gluten in each item as described previously (van Overbeek et al. 1997; Hopman et al. 2007; Hopman et al. 2012). For FFQ items that contained multiple foods (eg. ‘English muffins, bagels, or rolls), the gluten content of all sub-items was added and divided by the total number of sub-items. Intermediate serving sizes were selected when multiple options were present. Food items were considered gluten-containing if all associated sub-items were solely made with wheat, barley, or rye. Items that included foods made from other cereal grains in addition to those (eg. ‘Other grain products’) were not considered gluten-containing. See Appendix A1 for a list of gluten-rich foods on the Toronto-modified FFQ and their approximate
amount of gluten. Gluten intake for all dietary analyses was adjusted for total energy intake and expressed as grams of gluten per 1000 kcal. Individuals with incomplete dietary data, and those who were likely energy over-reporters (>3,500 kcal/day for women or >4,500 kcal/day for men) or under-reporters (<800 kcal/day for both women and men) were excluded from all dietary analyses (n=222). Individuals who self-reported being on a gluten/wheat-free or reduced diet were also excluded from dietary analyses (n=10) due to the lack of gluten-free food items on the FFQ.

2.3.5 Statistical Analysis

Statistical analyses were conducted using Statistical Analysis Software version 9.2 (SAS Institute). For all analyses, the α-error was set at 0.05 and reported P values are 2-sided. Continuous variables were loge or square root- transformed to improve normality where necessary. Although reported P values are from models using transformed variables, means and measures of spread for such variables are reported in their untransformed state to facilitate interpretability. Binomial 95% confidence intervals (Clopper-Pearson) were calculated for all dichotomous variables.

Subject characteristics between the two study populations were assessed using ANCOVA and chi-square tests for continuous and categorical variables, respectively. Fisher’s exact test was used to determine whether there were differences in the frequency of the HLA-DQ2/DQ8 elevated risk genotypes between ethnocultural groups. All possible pairs of ethnocultural groups were compared. The Bonferroni adjustment for multiple testing was used to account for the multiple pairwise comparisons (p<0.005, based on 10 tests and α=0.05). Differences in the
prevalence of positive CD serology between the two study populations were assessed using exact logistic regression adjusted for sex and ethnicity. Exact logistic regression adjusted for ethnicity and study population was also used to assess differences in the proportion of males and females with elevated tTG antibodies. Differences in average energy-adjusted gluten intake between those with positive and negative CD serology were assessed with ANCOVA adjusted for age, sex, BMI, ethnicity and study population. Differences in gluten intake between ethnocultural groups were assessed using ANCOVA adjusted for age, sex, BMI, total energy intake and study population. Duncan’s multiple range post hoc test was then used to correct for multiple comparisons between groups.

### 2.4 Results

Subject characteristics for both study populations are shown in Table 2.1. Differences (p<0.05) in age, sex, ethnocultural status, anthropometric measurements, blood pressure, blood glucose and lipid profiles were observed between study populations.

#### 2.4.1 HLA-DQ Genotypes

The overall prevalence of elevated risk HLA-DQ2/DQ8 genotypes across both study populations was 38.4% (95% CI 36.6–40.2%). The proportion of individuals with elevated-risk HLA-DQ genotypes in each ethnocultural group are shown in Figure 2.1. Caucasians and South Asians had a higher proportion of elevated-risk HLA genotypes than East Asians (p<0.005), while
individuals in the ‘Other’ and ‘Unknown’ groups had an intermediate prevalence of such alleles. A breakdown of all HLA-DQ sub-categories by ethnocultural group is shown in Figure 2.2. Generally, the presence of higher risk HLA-DQ risk alleles was increasingly rare as their associated CD risk increased across all ethnocultural groups.

2.4.2 Anti-Tissue Transglutaminase Antibodies

A total of 46 individuals tested positive on the initial dual-isotope tTG-IgA/IgG screen. Of these individuals, 22 tested positive for tTG IgA antibodies and were considered to have positive CD serology. Of the individuals who tested positive on the dual-isotope screen but negative for tTG IgA, one individual was found to be IgA deficient (total IgA < 7 mg/dl) and tested positive for tTG IgG antibodies. This individual was also considered to have positive CD serology. Two additional individuals who reported having been diagnosed with CD were also considered to have positive CD serology. In total, 25 individuals were identified with positive CD serology, which translates to an overall prevalence of 1:114 (0.88%; 95% CI, 0.57%-1.30%). In the TNH study population, 13 out of 15 (87%) cases were identified exclusively through serology and did not report having been diagnosed with CD in the health questionnaire. These likely represent cases of undiagnosed CD. The other two cases reported having been diagnosed with CD. Estimates on the ratio of undiagnosed to diagnosed cases of CD were not available for the THD study population since a diagnosis of CD would have excluded individuals from the study. Estimates for the prevalence of CD across ethnocultural groups are shown in Table 2.2. The majority of individuals with positive CD serology were Caucasian. No cases of positive serology were identified among East Asian or South Asian individuals, while a limited number of cases
were identified among those of ‘Other’ and ‘Unknown’ ancestry. The prevalence did not differ between the TNH population (0.93%; 95% CI, 0.52%-1.52%) and the THD population (0.83%; 95% CI, 0.40%-1.51%) after adjusting for sex and ethnicity (p=0.32). There was no difference in the prevalence between males (0.72%; 95% CI, 0.27%-1.57%) and females (0.95%; 95% CI, 0.57%-1.48%) after adjusting for study population and ethnicity (p=0.74). The proportion of individuals with positive CD serology by sub-category of HLA risk are shown in Figure 2.3. The proportion of individuals with positive serology tended to increase with increasing predefined HLA risk status (Pietzak et al. 2009), although differences between individual risk groups were not statistically significant (p>0.05).

2.4.3 Gluten Intake

Average gluten intake across both study populations was 10.1±0.13g/day (4.83±0.05g/1000 kcal/day). Energy-adjusted total gluten intake across ethnocultural groups are shown in Figure 2.4. Caucasians and South Asians had higher levels of gluten intake than East Asians, ‘Others’ and individuals of unknown descent (p<0.05). Energy-adjusted gluten intake did not significantly differ between those with elevated tTG antibodies (5.38±0.46 g/1000 kcal/day) and those without such antibodies (4.83±0.05 g/1000 kcal/day) after adjusting for age, sex, BMI, ethnicity and study population (p=0.5).
Table 2.1: Subject characteristics for TNH and THD Study Populations.

<table>
<thead>
<tr>
<th></th>
<th>TNH Study</th>
<th>THD Study</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (n)</td>
<td>1,620</td>
<td>1,212</td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>22.7 ± 0.06&lt;sup&gt;1&lt;/sup&gt;</td>
<td>44.7 ± 0.36</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Sex (n[%])</td>
<td></td>
<td></td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Female</td>
<td>1,103 (68)</td>
<td>901 (74)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>517 (32)</td>
<td>311 (26)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity (n[%])</td>
<td></td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Caucasian</td>
<td>769 (47)</td>
<td>721 (59)</td>
<td></td>
</tr>
<tr>
<td>East Asian</td>
<td>560 (35)</td>
<td>54 (4)</td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>173 (11)</td>
<td>49 (4)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>118 (7)</td>
<td>154 (13)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>234 (19)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>22.9 ± 0.09</td>
<td>32.5 ± 0.17</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>74.4 ± 0.23</td>
<td>102.1 ± 0.42</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>114 ± 0.29</td>
<td>115 ± 0.36</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>69 ± 0.20</td>
<td>73 ± 0.26</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>4.79 ± 0.01</td>
<td>4.91 ± 0.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total(mmol/L)</td>
<td>4.26 ± 0.02</td>
<td>5.07 ± 0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>HDL (mmol/L)</td>
<td>1.54 ± 0.01</td>
<td>1.27 ± 0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>LDL (mmol/L)</td>
<td>2.28 ± 0.02</td>
<td>3.23 ± 0.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Total:HDL cholesterol</td>
<td>2.91 ± 0.02</td>
<td>4.23 ± 0.04</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

<sup>1</sup>Mean ± SE (all such values). Differences between study populations were assessed using X<sup>2</sup> tests for categorical variables and ANCOVA adjusted for age, sex, BMI and ethnicity for continuous variables.
Table 2.2: Prevalence of positive CD serology and previous diagnoses across ethnocultural groups.  

<table>
<thead>
<tr>
<th>Ethnocultural Group</th>
<th>Cases (n)</th>
<th>Prevalence (%[95% CI])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>25</td>
<td>0.88 (0.57-1.30)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>22</td>
<td>1.48 (0.93-2.23)</td>
</tr>
<tr>
<td>East Asian</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>South Asian</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>0.74 (0.09-2.63)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>0.43 (0.01-2.36)</td>
</tr>
</tbody>
</table>

Cases of CD were determined by measuring levels of anti-tissue transglutaminase (tTG) IgA antibodies and previous diagnosis. tTG IgG antibodies were assessed in cases of IgA deficiency. CI, confidence interval.
Figure 2.1: Prevalence of celiac disease-associated elevated-risk HLA genotypes across ethnocultural groups. Elevated-risk genotypes include DQA1*0501-DQB1*0201 (DQ2.5), DQA1*0301-DQB1*0302 (DQ8) or DQA1*0505-DQB1*0301/DQA1*0201-DQB1*0202 (DQ2.2/DQ7). Differences in the prevalence of elevated-risk genotypes between groups compared using Fisher’s exact test. All pairs of ethnocultural groups were compared, and a Bonferroni correction was applied to account for the multiple pairwise tests. Different letters indicate estimates that significantly differed (p<0.005). Error bars represent 95% confidence intervals.
Figure 2.2: Prevalence of HLA-DQ genotypes across ethnocultural groups. Celiac disease HLA risk groups (increasing risk from left to right) modified from previously established risk gradient (3). Error bars represent 95% confidence intervals. Caucasians, n=1,492; East Asians, n=615; South Asians, n=223; Other, n=272; Unknown, n=234. DQX, non-DQ2/8 type; DQY, DQ2.2 or non-DQ2/8 type; DQ8, DQ8 hetero- and homozygotes; DQ2*, DQ2.2 or DQ2.5 type.
Figure 2.3: Prevalence of elevated tTG antibodies by celiac disease-associated HLA-DQ risk category. HLA risk groups (increasing risk from left to right) modified from previously established risk gradient (3). Error bars represent 95% confidence intervals. DQY, DQ2.2 or non-DQ2/8 type; DQ8, DQ8 hetero- and homozygotes; DQ2*, DQ2.2 or DQ2.5 type.
**Figure 2.4:** Average energy-adjusted gluten intake by ethnocultural group. Differences between groups assessed using ANCOVA adjusted for age, sex, BMI and study population. Different letters indicate means that significantly differed (p<0.05). Duncan’s multiple range post hoc test was used to account for multiple comparisons between means. Error bars represent SEs.
2.5 Discussion

CD represents a common autoimmune condition affecting individuals in North America and Europe (Guandalini and Assiri 2014). The prevalence of CD is known to vary globally and is thought to be particularly rare in certain regions, notably in Asia (Cummins and Roberts-Thomson 2009; Gujral, Freeman, and Thomson 2012; Kang et al. 2013). Differences in dietary gluten intake and the prevalence of HLA variants conferring susceptibility to CD across regions is thought to partially explain the variability in the prevalence of CD observed between nations. However, it is unclear if such factors play a role in the development of CD in individuals from different ethnocultural backgrounds living in North America. Furthermore, there is a notable lack of screening studies assessing the prevalence of CD in individuals from diverse ethnocultural backgrounds living in Canada. To our knowledge, the present study is the first to screen for CD-associated antibodies in a population of Canadian adults. We found the prevalence of positive CD serology among ethnoculturally diverse Canadians living in Toronto to be 0.88% (95% CI, 0.57-1.30%), which is similar to estimates in Europe and the United States (Rubio-Tapia, Kyle, et al. 2009; Fasano et al. 2003; Mardini, Westgate, and Grigorian 2015; Mustalahti et al. 2010; Kang et al. 2013; Kim et al. 2016). Furthermore, 87% of likely CD cases were undiagnosed, which is consistent with commonly reported figures of about 85% (Kagnoff 2007; Guandalini and Assiri 2014; Rubio-Tapia et al. 2012) and consistent with findings from the Canadian Celiac Association Health Survey indicating an average delay in CD diagnosis of ~12 years after the initial onset of symptoms (Cranney et al. 2007). Caucasians comprised the majority of CD cases, while no individuals of East Asian or South Asian descent had positive serology or a previous diagnosis of CD. Compared to Caucasians, East Asians consumed less gluten and had a lower
prevalence of HLA-DQ high risk genotypes, while South Asians and Caucasians had similar intakes of gluten and a similar prevalence of HLA-DQ high risk genotypes.

CD is known to vary by ethnicity and has historically been considered a condition that primarily affects Caucasians (Cataldo and Montalto 2007; Gujral, Freeman, and Thomson 2012). Our findings indicate that Caucasians are among the highest risk ethnocultural groups for CD in Canada. It is noteworthy that one individual in the ‘Other’ ethnocultural group with positive CD serology was of mixed descent, with one parent being Caucasian. In contrast, CD is rare in East Asia (Kang et al. 2013), although case reports have been previously documented in Canada (Freeman 2003). This is consistent with our findings that CD is rare in East Asians, and this is partly due to the lower prevalence of HLA-DQ2/8 risk alleles and lower gluten intake.

Conversely, in South Asians we found that the prevalence of HLA risk alleles and average gluten intake was similar to Caucasians. This suggests that CD may occur in South Asians at a frequency that approaches that of Caucasians. Indeed, although we did not identify any cases with positive CD serology in the cohort of South Asians in the present study, screening studies in South Asian populations around the world suggest that CD may affect individuals at similar frequencies to those reported for Caucasians (Sher et al. 1993; Ramakrishna et al. 2016; Sood et al. 2006; Butterworth, Iqbal, and Cooper 2005).

We observed a pattern of elevated tTG antibodies by HLA risk sub-categories consistent with previously established HLA-DQ risk gradients (Pietzak et al. 2009). Differences between risk sub-categories were not as distinct as previously reported (Pietzak et al. 2009), however this is likely a result of smaller sample size and the fact that our study populations did not consist of individuals considered to be at high risk for CD. It has also been shown that the prevalence of certain HLA risk alleles in individuals of different ethnocultural backgrounds with CD differs
from the prevalence of such alleles in Caucasians with CD (Butterworth, Iqbal, and Cooper 2005). This suggests that other genetic regions that confer susceptibility to CD may be of increased importance in certain ethnocultural groups. Indeed, genome-wide studies have identified several other genetic variants that account for additional fractions of the genetic risk for CD (Dieli-Crimi, Cenit, and Nunez 2015). It is possible that such variants, coupled with HLA-DQ genotypes, may be particularly useful in establishing genetic risk gradients for CD in individuals of non-Caucasian ancestry.

Many factors in addition to genetic risk and gluten consumption are thought to contribute to differences in the prevalence of CD across ethnocultural groups. The timing of first gluten exposure was long thought to play a role in the development of CD (Norris et al. 2005); however, this has recently been called into question and it is unclear whether the initial introduction of gluten into the diet influences risk of developing CD (Lionetti et al. 2014). Exposure to certain viral infections during first introduction of gluten is thought to possibly prime the immune system and increase the risk of developing CD in at-risk infants (Myleus et al. 2012; Stene et al. 2006). Interactions between the gut microbiota and the intestinal mucosa can also affect the immune system (Belkaid and Hand 2014). Gut microbiota profiles have been shown to differ between patients with CD and healthy controls, and it is thought that bacteria-mediated inflammation in the mucosa may also play a role in the development of CD (Marasco et al. 2016). Dietary patterns are known to influence the composition of the gut microbiota (Wu et al. 2011). Furthermore, differences in dietary patterns between ethnicities have been identified, including within the TNH study population assessed here (Brenner et al. 2011). It is, therefore, possible that differences in broad dietary patterns between ethnocultural groups may explain some of the variation in the prevalence of CD between such groups.
Demographic characteristics of individuals with CD are of considerable interest. We observed no difference in the prevalence of positive CD serology between males and females. While it has been reported that more women are diagnosed with CD than men (Green and Cellier 2007), our results are in agreement with other antibody screening studies which report similar rates of CD between sexes (Rubio-Tapia et al. 2012; Fasano et al. 2003). Furthermore, while there are distinct peaks in diagnosis of CD in both early childhood and later in adulthood (Dewar and Ciclitira 2005; Green et al. 2001), screening studies have suggested that the prevalence of positive CD serology is similar across adult age groups (Rubio-Tapia, Kyle, et al. 2009; Rubio-Tapia et al. 2012; Kim et al. 2016; Fasano et al. 2003). In agreement with these findings, we observed no significant difference in the prevalence of positive CD serology between the TNH and THD cohorts, populations with average ages of 23 and 45 years, respectively. We identified no difference in gluten intake between those with positive and negative CD serology. This suggests that, on average, those with undiagnosed CD likely do not link gluten intake to any adverse symptoms and regulate their diet accordingly. Finally, while rare in the general population, IgA deficiency occurs more often in CD (Leffler and Schuppan 2010), affecting roughly 3-5% of individuals with this condition (Halfdanarson, Litzow, and Murray 2007). Consistent with these estimates, we found that 1 out of 23 (4.3%) individuals with positive CD serology was IgA deficient.

The present study has some limitations. There were low numbers of individuals in some ethnocultural groups examined, and estimates of the prevalence of CD in these groups should be interpreted with caution. Furthermore, while the populations included were from Toronto, Canada’s largest metropolis, the results may not be generalizable to all Canadian adults. While we were primarily interested in the prevalence of CD, subjects with positive serology did not undergo a confirmatory biopsy for a definitive diagnosis of CD. This may have resulted in an
overestimation of the prevalence of CD in the assessed study populations. Alternatively, the lack of histological assessment could have missed a small subset of individuals with seronegative CD (Volta et al. 2016). Nevertheless, positive CD serology is often indicative of the identification of villous atrophy upon biopsy (Carroccio et al. 2002). Additionally, we considered individuals to have positive serology suggestive of CD based on two positive human-recombinant tTG ELISA tests. Screening for CD is sometimes carried out using both tTG and EMA antibody assays. However, EMA relies on indirect immunofluorescence, and the results can be highly subjective and require extensive experience to interpret (Nandiwada and Tebo 2013; Leffler and Schuppan 2010). Measuring tTG antibodies is considered appropriate for screening purposes and tTG has been specifically recommended for screening asymptomatic individuals due to its greater sensitivity (Lewis and Scott 2006; Leffler and Schuppan 2010). Finally, while only 13% of likely CD cases in the TNH population were previously diagnosed, we were unable to assess cases of diagnosed CD in the THD cohort since such individuals would have been excluded from the study. However, if a similar ratio of diagnosed to undiagnosed cases between study populations is assumed, this would equate to an additional 2 individuals with clinically diagnosed CD in the THD cohort. This would increase the overall prevalence of positive CD serology to 27 out of 2,834 (0.95%; 95% CI 0.63-1.38%) in our multiethnic population of Canadian adults.

In summary, we conducted a screening study for positive CD antibodies in an ethnoculturally diverse population of Canadian adults in Toronto. We demonstrated that approximately 1% of Canadian adults have positive CD serology. We also report that 87% of likely CD cases are undiagnosed. Caucasians had the highest prevalence of positive CD serology, HLA-DQ2/8 elevated-risk alleles, and highest average gluten intake. While no individuals of East Asian or South Asian descent were found to have positive CD serology, average gluten intake and the prevalence of high risk HLA-DQ2/8 alleles were similar in South Asians and Caucasians, but
both were lower in East Asians. This study highlights the high prevalence of undiagnosed cases of CD in Canada, and emphasizes the importance of serologic screening in patients with high risk HLA genotypes.
CHAPTER 3:

BIOMARKERS OF CARDIOMETABOLIC HEALTH AND NUTRITIONAL STATUS IN INDIVIDUALS WITH POSITIVE CELIAC DISEASE SEROLOGY

Adapted from:

3.1 Abstract

**Background:** Celiac disease (CD) is an autoimmune disorder characterized by damage to the intestinal mucosa and nutrient malabsorption in many cases. Nevertheless, it remains unclear whether nutrient deficiencies and other adverse health effects are prevalent in individuals with positive CD serology identified through screening studies.

**Objective:** The objective was to determine whether biomarkers of cardiometabolic health and nutritional status differ between those with positive CD serology and those with negative serology identified in a screening study of Canadian adults.

**Methods:** Subjects (n=2,832) were participants from the Toronto Nutrigenomics and Health Study and the Toronto Healthy Diet Study. Individuals were previously screened for CD-specific anti-tissue transglutaminase IgA autoantibodies. Lipid profiles as well as concentrations of six carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, all-trans lycopene, and zeaxanthin), three tocopherols (α-tocopherol, δ-tocopherol, and γ-tocopherol), vitamin A (all-trans-retinol), vitamin C (ascorbic acid), and vitamin D (25-hydroxyvitamin D) were compared between those with positive and negative CD serology using general linear mixed models.

**Results:** Individuals with positive CD serology had significantly lower levels of HDL-cholesterol (p=0.008), lower levels of apolipoprotein-AI (p=0.02), a higher ratio of total cholesterol to HDL cholesterol (p=0.006) and a higher apolipoprotein-B/AI ratio (p=0.03) than those with negative CD serology. Positive CD serology was also associated with significantly lower concentrations of retinol (p=0.006) in fully adjusted models accounting for age, sex, BMI,
retinol intake and serum triglycerides. Those with positive CD serology had lower serum 25-hydroxyvitamin D in unadjusted models (p=0.01), but not in fully adjusted models (p=0.08). Circulating carotenoids, tocopherols, or ascorbic acid did not differ in those with positive CD serology.

**Conclusions:** Individuals with positive CD serology and not diagnosed with CD may have unfavorable lipid profiles and lower circulating levels of certain fat-soluble vitamins compared to those with negative CD serology. These findings suggest that individuals with undiagnosed CD may be at elevated risk for cardiometabolic disease and inadequacy of certain fat-soluble vitamins, but not widespread nutrient deficiencies.
3.2 Introduction

Celiac disease (CD) is an autoimmune disorder which affects approximately 1% of individuals in North America and Europe (Mardini, Westgate, and Grigorian 2015; Kim et al. 2016; Mustalahti et al. 2010). Individuals with CD experience adverse autoimmune reactions to dietary gluten intake, a protein found in wheat, barley and rye, and strict adherence to a gluten-free diet remains the only available treatment to date (Green and Cellier 2007). Recent studies have suggested that the prevalence of CD is on the rise (Rubio-Tapia, Kyle, et al. 2009; Ludvigsson, Rubio-Tapia, et al. 2013); however, the underlying reasons for this increase is unclear. Predisposing variants in the human leukocyte antigen (HLA)-DQ region are necessary, but not sufficient for the development of CD (Dieli-Crimi, Cenit, and Nunez 2015). Other risk factors associated with the development of CD include non-HLA genetic risk variants, early viral infections, and the composition of the gut microbiota (Sarno et al. 2015).

CD has various clinical manifestations and levels of severity. The ‘typical’ presentation CD is characterized by the development of autoantibodies against tissue transglutaminase (tTG), a high degree of villous atrophy, and classical gastrointestinal symptoms. Failure to thrive is an indicator for typical CD, and this form is most commonly diagnosed in early childhood (Guandalini and Assiri 2014). ‘Atypical’ CD is characterized by the presence of elevated tTG antibodies and extra-intestinal symptoms including iron deficiency anemia, fatigue, headaches, joint pain and even psychiatric disorders. This presentation of CD is often associated with a lower degree of damage to the intestinal mucosa and is most frequently diagnosed in adulthood (Guandalini and Assiri 2014; Gujral, Freeman, and Thomson 2012). Adults with positive CD serology and mild intestinal damage may also be relatively asymptomatic. This is referred to as the ‘silent’ form of CD (Fasano and Catassi 2001; Guandalini and Assiri 2014). ‘Latent’ CD is
characterized by the potential presence of CD autoimmunity, but no observable damage to the intestinal mucosa (Gujral, Freeman, and Thomson 2012; Guandalini and Assiri 2014). Finally, refractory CD is a rare and extremely severe form of CD where intestinal damage and symptoms of CD fail to improve on a gluten-free diet (Rubio-Tapia and Murray 2010).

Untreated CD is known to be associated with an increased risk of many adverse health outcomes related to impaired nutritional status. This is not surprising considering that damage to the intestinal mucosa and malabsorption are major hallmarks of the disorder. Studies of recently diagnosed CD patients have reported deficiencies in several vitamins and minerals compared to healthy controls (Wierdsma et al. 2013; Dahele and Ghosh 2001; Botero-Lopez et al. 2011). Individuals with newly diagnosed CD have also been shown to have altered blood lipid profiles compared to healthy controls (Brar et al. 2006; Capristo et al. 2000). However, it is unclear whether such nutrient deficiencies and other adverse health effects associated with diagnosed CD are present in individuals without a diagnosis of CD, but who test positive for CD-associated antibodies in screening studies. This is especially important considering that up to 85% of individuals with CD are undiagnosed (Rubio-Tapia et al. 2012). The objective of this study was to determine whether biomarkers of cardiometabolic health and nutritional status differ between Canadian adults with positive and negative CD serology identified in a screening study.

### 3.3 Methods

#### 3.3.1 Study Population

Subjects were Canadian adults from two study populations described previously (see Chapter
2.3.1) The Toronto Nutrigenomics and Health (TNH) Study is a cross-sectional examination of young adults aged 20-29 recruited from the University of Toronto Campus (n=1,620). The Toronto Healthy Diet (THD) Study is a randomized-controlled trial (clinicaltrials.gov registration: NCT00516620), which investigated the effectiveness of various types of dietary interventions on weight loss and biomarkers of chronic disease (n=1,212 at baseline). Individuals in the THD Study were between 18 and 82 years of age, and all subjects came from households in the Toronto region with at least one member possessing a BMI $\geq 25$ kg/m$^2$. In both study populations, individuals had anthropometric measurements taken, provided a fasting blood sample for biomarker assessment, and completed the Toronto-modified Willet 196-item semi-quantitative food frequency questionnaire (FFQ). Subjects provided a self-reported account of their ethnocultural background and were placed into one of five groups: Caucasian, East Asian, South Asian, Other, or Unknown as described previously (Garcia-Bailo et al. 2012).

3.3.2 Celiac Disease Serology

Serological screening for CD was previously carried out by measuring tissue transglutaminase (tTG) autoantibodies as described (see Chapter 2.3.3). Briefly, plasma samples from individuals possessing the HLA DQ2 and DQ8 risk variants necessary for the development of CD were screened initially using a dual-isotope (IgA+IgG) anti-tTG enzyme-linked immunosorbent assay (ELISA). Genotyping for the HLA DQ2 and DQ8 risk alleles was described previously (see Chapter 2.3.2). Individuals who tested positive for tTG antibodies on this initial assay were then tested using separate tTG IgA assay. Individuals who tested positive for anti-tTG IgA antibodies were considered to have positive CD serology. Individuals who tested negative for tTG IgA
antibodies were screened for selective IgA deficiency. Individuals identified as being IgA deficient were then screened using a separate assay for tTG IgG antibodies. Those with IgA deficiency and positive tTG IgG antibodies were considered to have positive CD serology. Those who were positive on the initial dual-isotope screen, but tTG IgA negative and IgA sufficient were considered to have equivocal serology and were excluded from the present study (n=25). Those with a self-reported diagnosis of CD and negative serology were also excluded (n=2).

Individuals who did not have the HLA-DQ2/8 variants necessary for the development of CD were not screened for tTG antibodies, and were all considered to have negative CD serology in the present study. East Asian (n=614) and South Asian (n=222) individuals were excluded from all analyses in the present study due to the absence of individuals in these groups with positive CD serology. Analyses were, therefore, restricted to individuals of Caucasian or Other ancestry.

3.3.3 Biomarker Assessment

Markers of cardiometabolic health included triglycerides, total cholesterol, HDL cholesterol, LDL cholesterol, the ratio of total to HDL cholesterol, apolipoprotein-A1 (apo-A1), apolipoprotein-B (apo-B) and the ratio of apo-B to apo-A1. Biomarkers of nutritional status included serum carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, all-trans lycopene, and zeaxanthin), tocopherols (α-tocopherol, δ-tocopherol, and γ-tocopherol), vitamin A (all-trans-retinol), vitamin C (ascorbic acid), and vitamin D (25-hydroxyvitamin D).

Fasting blood samples were collected from each participant for biomarker assessment. Serum cholesterol (total, HDL, and LDL) and triglycerides were assessed using a chromatographic enzymatic method described previously (Cahill, Corey, and El-Sohemy 2009). Apo-A1 and apo-
B concentrations were measured at the University of Victoria-Genome British Columbia Proteomics Centre using a multiple reaction monitoring assay described previously (Kuzyk et al. 2009; Domanski et al. 2012). Biomarkers of nutritional status were only available for individuals from the TNH Study. Concentrations of all six carotenoids, three tocopherols and all-trans retinol were measured from plasma samples using reverse-phase isocratic high-performance liquid chromatography (HPLC) with fluorescent detection as described previously (Liu et al. 2011). Serum ascorbic acid was measured using HPLC as described previously (Cahill, Corey, and El-Sohemy 2009). Circulating concentrations of 25-hydroxyvitamin D (25OHD) were measured using an HPLC-tandem mass spectrometry (MS/MS) assay described elsewhere (Wagner et al. 2011; Garcia-Bailo, Karmali, et al. 2013). Vitamin D values reported in the present study represent the sum of both 25-hydroxycholecalciferol (25-hydroxyvitamin D3) and 25-hydroxyergocalciferol (25-hydroxyvitamin D2).

3.3.4 Statistical Analysis

All statistical analyses were performed using Statistical Analysis Software (SAS) version 9.2 (SAS Institute Inc., Cary, N.C., USA). Continuous variables were loge- or square root-transformed to improve normality when required. Raw means and measures of variance are reported in order to facilitate interpretability, but reported P values are from models that use transformed variables where necessary. For all analyses, the α-error was set to 0.05 and reported P values are 2-sided. When means were compared between individuals with positive and negative CD serology, general linear mixed models (GLMM) were used in order to estimate variances separately between the two groups. The Satterthwaite approximation was used to
adjust the degrees of freedom for the unequal variances. This was done to account for potentially heterogeneous variances that can arise between groups of contrasting sample sizes.

Subject characteristics were compared between those with positive and negative CD serology using GLMMs adjusted for age, sex, BMI, ethnicity and study population (TNH or THD Study) for continuous variables, and Fisher’s exact test for categorical variables. Lipid profiles were compared between groups using GLMMs adjusted for age, sex, BMI, ethnicity and study population. Differences in circulating concentrations of nutritional biomarkers between groups were also compared using GLMMs. Only Caucasians from the TNH study were included as all individuals with positive CD serology and complete data on nutritional biomarkers were of Caucasian descent. Initially, the association between each biomarker and CD serology status was explored using unadjusted GLMMs (Model 1). Model 2 consisted of GLMs adjusted for age, sex and BMI. Finally, Model 3 included the same covariates as Model 2 plus dietary intake of each respective compound derived from subjects’ FFQ responses. This was done to identify differences between groups that were likely due to altered absorption of such nutritional biomarkers, rather than differences in dietary intake. Model 3 for retinol was further adjusted for serum triglycerides. For 25OHD, season of blood draw, physical activity and hormonal contraceptive use among women were also included in Model 3 since seasonal variation, sun exposure and hormonal contraceptives have been shown to significantly influence circulating levels of 25OHD (Garcia-Bailo, Karmali, et al. 2013). Since outcome variables were selected based on an a priori hypothesis that undiagnosed CD may be associated with altered levels of these biomarkers, analyses were not adjusted for multiple comparisons.
3.4 Results

Subject characteristics for those with positive and negative CD serology are shown in Table 3.1. There were no significant differences for any subject characteristics, including height, weight and BMI, between groups (p>0.05). Differences in blood lipid profiles between groups are shown in Table 3.2. Individuals with positive CD serology had lower levels of HDL-cholesterol and a higher total- to HDL-cholesterol ratio than those with negative CD serology (p<0.05). Circulating levels of apo-A1 were also significantly lower, and the ratio of apo-B to apo-A1 was higher in those with positive CD serology (p<0.05). Such differences were driven by HDL-cholesterol and apo-A1 since total cholesterol, LDL cholesterol, serum triglycerides and apo-B did not significantly differ between those with positive and negative CD serology.

Table 3.3 shows the association between various biomarkers of nutritional status and CD serology status. Individuals with positive CD serology had significantly (p<0.05) lower levels of circulating retinol than those with negative CD serology in all models. Positive CD serology was associated with lower levels of 25OHD in the unadjusted model (p=0.01) and after adjusting for age, sex and BMI (p=0.009); however, this association was no longer significant after adjusting further for vitamin D intake, season of blood draw, physical activity and hormonal contraceptive use among women (p=0.08). No other differences in circulating markers of nutritional status between those with positive and negative CD serology were identified in any models (p>0.05).
Table 3.1: Subject characteristics by CD serology status\(^1\).

<table>
<thead>
<tr>
<th></th>
<th>Negative CD Serology</th>
<th>Positive CD Serology</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (n)</td>
<td>1,929</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>35.35 ± 0.33(^1)</td>
<td>32.13 ± 2.73</td>
<td>0.75</td>
</tr>
<tr>
<td>Sex (n[%])</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1,374 (71)</td>
<td>17 (74)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Male</td>
<td>555 (29)</td>
<td>6 (26)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td>Caucasian</td>
<td>1,437 (74)</td>
<td>20 (87)</td>
<td></td>
</tr>
<tr>
<td>Non-Caucasian</td>
<td>492 (26)</td>
<td>3 (13)</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>167.37 ± 0.21</td>
<td>167.33 ± 1.62</td>
<td>0.88(^3)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>79.99 ± 0.44</td>
<td>82.50 ± 4.39</td>
<td>0.16(^3)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>28.60 ± 0.16</td>
<td>29.42 ± 1.45</td>
<td>0.11(^3)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>90.74 ± 0.42</td>
<td>91.03 ± 3.67</td>
<td>0.18(^3)</td>
</tr>
<tr>
<td>Physical activity (METs)</td>
<td>9.47 ± 0.28</td>
<td>7.48 ± 0.85</td>
<td>0.7</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>115.52 ± 0.27</td>
<td>115.39 ± 2.50</td>
<td>0.85</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>71.73 ± 0.20</td>
<td>73.39 ± 1.74</td>
<td>0.45</td>
</tr>
<tr>
<td>C-reactive protein (mg/L)</td>
<td>2.33 ± 0.12</td>
<td>2.87 ± 1.18</td>
<td>0.98</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>4.84 ± 0.01</td>
<td>4.88 ± 0.10</td>
<td>0.62</td>
</tr>
</tbody>
</table>

\(^1\)Differences between groups assess using general linear mixed models for continuous variables (adjusted for age, sex, ethnicity, BMI and study population) with variances estimated separately within each group. The Satterthwaite approximation was used to adjust the degrees of freedom for potentially unequal variances. Fisher’s exact test was used assess differences in categorical variables between groups. CD, celiac disease; METs, metabolic equivalent hours per week.

\(^2\)Means ± SEs (all such values).

\(^3\)Adjusted for age, sex, ethnicity and study population only.
Table 3.2: Lipid profiles by CD serology status

<table>
<thead>
<tr>
<th></th>
<th>Negative CD Serology (n=1,929)</th>
<th>Positive CD Serology (n=23)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (mmol/L)</td>
<td>$4.71 \pm 0.02$(^2)</td>
<td>$4.55 \pm 0.16$</td>
<td>0.61</td>
</tr>
<tr>
<td>HDL (mmol/L)</td>
<td>$1.40 \pm 0.01$</td>
<td>$1.19 \pm 0.05$</td>
<td>0.008</td>
</tr>
<tr>
<td>LDL (mmol/L)</td>
<td>$2.80 \pm 0.02$</td>
<td>$2.86 \pm 0.13$</td>
<td>0.25</td>
</tr>
<tr>
<td>Total:HDL cholesterol</td>
<td>$3.60 \pm 0.03$</td>
<td>$3.92 \pm 0.16$</td>
<td>0.006</td>
</tr>
<tr>
<td>Triglycerides (mmol/L)</td>
<td>$1.12 \pm 0.01$</td>
<td>$1.10 \pm 0.11$</td>
<td>0.99</td>
</tr>
<tr>
<td>Apolipoproteins(^3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apo-A1 (g/L)</td>
<td>$0.94 \pm 0.01$</td>
<td>$0.89 \pm 0.07$</td>
<td>0.02</td>
</tr>
<tr>
<td>Apo-B (g/L)</td>
<td>$0.23 \pm 0.004$</td>
<td>$0.30 \pm 0.04$</td>
<td>0.31</td>
</tr>
<tr>
<td>Apo-B:A1</td>
<td>$0.24 \pm 0.003$</td>
<td>$0.31 \pm 0.04$</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\(^1\)Differences between groups assess using general linear mixed models adjusted for age, sex, ethnicity, BMI and study population with variances estimated separately within each group. The Satterthwaite approximation was used to adjust the degrees of freedom for potentially unequal variances. CD, celiac disease.

\(^2\)Mean ± SE (all such values).

\(^3\)n=1,653 (negative CD serology), 20 (positive CD serology).
Table 3.3: Biomarkers of nutritional status by CD serology status\textsuperscript{1}.

<table>
<thead>
<tr>
<th></th>
<th>Negative CD Serology (n=565)</th>
<th>Positive CD Serology (n=10)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carotenoids (μmol/L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-carotene</td>
<td>0.21 ± 0.01</td>
<td>0.24 ± 0.08</td>
<td>0.94</td>
<td>0.99</td>
<td>0.89</td>
</tr>
<tr>
<td>β-carotene</td>
<td>0.65 ± 0.02</td>
<td>0.68 ± 0.16</td>
<td>0.79</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>β-cryptoxanthin</td>
<td>0.28 ± 0.01</td>
<td>0.22 ± 0.03</td>
<td>0.43</td>
<td>0.39</td>
<td>0.36</td>
</tr>
<tr>
<td>Lutein</td>
<td>0.40 ± 0.01</td>
<td>0.38 ± 0.07</td>
<td>0.64</td>
<td>0.52</td>
<td>0.41</td>
</tr>
<tr>
<td>Lycopene</td>
<td>0.92 ± 0.02</td>
<td>1.09 ± 0.24</td>
<td>0.77</td>
<td>0.70</td>
<td>0.64</td>
</tr>
<tr>
<td>Zeaxanthin</td>
<td>0.11 ± 0.003</td>
<td>0.09 ± 0.02</td>
<td>0.32</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td>Tocopherols (μmol/L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-tocopherol</td>
<td>30.56 ± 0.49</td>
<td>27.56 ± 2.86</td>
<td>0.40</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td>δ-tocopherol</td>
<td>0.29 ± 0.01</td>
<td>0.21 ± 0.03</td>
<td>0.37</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>γ-tocopherol</td>
<td>3.50 ± 0.08</td>
<td>3.29 ± 0.54</td>
<td>0.88</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>Other Vitamins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascorbic acid (μmol/L)\textsuperscript{2}</td>
<td>31.21 ± 0.68</td>
<td>36.45 ± 4.88</td>
<td>0.21</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>Retinol (μmol/L)</td>
<td>2.12 ± 0.02</td>
<td>1.74 ± 0.10</td>
<td>0.005</td>
<td>0.004</td>
<td>0.006\textsuperscript{4}</td>
</tr>
<tr>
<td>25-hydroxyvitamin D (nmol/L)\textsuperscript{3}</td>
<td>71.19 ± 1.12</td>
<td>53.98 ± 4.48</td>
<td>0.01</td>
<td>0.009</td>
<td>0.08\textsuperscript{5}</td>
</tr>
</tbody>
</table>

\textsuperscript{1}All values represent means ± SEs. Differences between groups assessed using general linear mixed models with variances estimated separately within each group. The Satterthwaite approximation was used to adjust the degrees of freedom for potentially unequal variances. Model 1 was unadjusted. Model 2 was adjusted for age, sex and BMI. Model 3 was adjusted for age, sex, BMI, as well as intake of each respective compound. CD, celiac disease.
\textsuperscript{2}n=612 (negative CD serology), n=11 (positive CD serology).
\textsuperscript{3}n=736 (positive CD serology), n=12 (negative CD serology).
\textsuperscript{4}Additionally adjusted for serum triglycerides.
\textsuperscript{5}Additionally adjusted for season of blood draw, physical activity, and hormonal contraceptive use among women.
3.5 Discussion

CD is an autoimmune condition which primarily affects the intestinal mucosa. In its most severe form, CD can lead to complete villous atrophy and result in impaired nutrient absorption (Guandalini and Assiri 2014). The prevalence of CD has been shown to be on the rise (Rubio-Tapia, Kyle, et al. 2009; Ludvigsson, Rubio-Tapia, et al. 2013), currently affecting approximately 1% of North Americans (Mardini, Westgate, and Grigorian 2015; Kim et al. 2016). Such estimates are based on serologic screening studies, and the number of individuals actually diagnosed with CD is considerably less (Rubio-Tapia et al. 2012). Although more severe manifestations of CD are known to be associated with failure to thrive, unexplained weight loss and various signs of malnutrition (Guandalini and Assiri 2014), it is unclear whether the estimated 85% of those with CD who remain undiagnosed are affected by similar adverse health effects. The objective of this study was to examine the association between positive CD serology in those without a diagnosis of CD and various biomarkers of cardiometabolic health and nutritional status. We found that individuals with positive CD serology had lower levels of HDL-cholesterol, lower apo-AI, a higher total-to HDL-cholesterol ratio, a higher apo-B/AI ratio, and lower levels of circulating retinol compared to individuals with negative CD serology or those lacking the HLA-DQ risk variants necessary for the development of CD.

Positive CD serology was not associated with any differences in height, weight or BMI compared to those with negative CD serology in our study population. It is well established that severe manifestations of CD, particularly those diagnosed in early childhood, are associated with reduced height and body weight due to the sustained presence of severe intestinal damage and resulting nutrient malabsorption (Tully 2008). However, recent studies have found that this classical presentation of CD may not be applicable to all cases, particularly to those that are
diagnosed in adulthood (Tucker et al. 2012; Dickey and Kearney 2006). Indeed, in the United States, almost 40% of those with biopsy-proven CD have been shown to be overweight or obese (Stein et al. 2016). This is likely due to the increased prevalence of the ‘atypical’ and ‘silent’ forms of CD observed in adulthood (Fasano and Catassi 2001). A significant portion of children newly diagnosed with CD have also been shown to be overweight or obese (Reilly et al. 2011; Venkatasubramani, Telega, and Werlin 2010). Interestingly, adherence to a gluten-free diet in both overweight or obese children and adults with CD has been shown to result in beneficial decreases in BMI (Reilly et al. 2011; Cheng et al. 2010). Our results support the notion that many adults with undiagnosed CD may not be underweight. Taken together with evidence which suggests it may take longer for overweight individuals to receive a diagnosis of CD (Tucker et al. 2012), this stresses the importance of screening in overweight and obese individuals with signs and symptoms of CD.

We observed significant differences in lipid profiles in those with positive CD serology compared to those with negative serology. In agreement with previous findings, we observed lower levels of both HDL-cholesterol and apo-A1 in individuals with positive CD serology (Capristo et al. 2000). We did not observe lower total cholesterol values in individuals with positive CD serology reported in many studies of patients with newly diagnosed CD (Lewis et al. 2009; Ciacci et al. 1999; West et al. 2003). However, recent studies suggest that such changes in total cholesterol in CD may be largely driven by HDL-cholesterol (Brar et al. 2006). It is possible that changes in the intestinal mucosa associated with CD could lead to reduced lipid absorption and lower levels of HDL-cholesterol. The link between apo-A1 and the intestinal mucosa could further explain the unique association with HDL-cholesterol. Apo-A1 is produced and secreted in the intestinal mucosa, and CD-associated damage to the mucosa has been shown to lead to a reduction in apo-A1 levels (Floren and Alm 1988; Capristo et al. 2000). Apo-A1 is
the major lipoprotein constituent of HDL-cholesterol (Frank and Marcel 2000), and a decrease in apo-A1 levels would lead to a reduction in HDL-cholesterol synthesis. In support of this proposed mechanism, we observed that individuals with positive CD serology had significantly lower levels of apo-A1 and a higher apo-B/A1 ratio compared to those with negative serology. A recent meta-analysis found that individuals with CD are at a greater risk for cardiovascular disease (CVD) than the general population (Emilsson et al. 2015). Our findings of lower HDL-cholesterol and apo-A1, and a higher ratio of total-to-HDL-cholesterol and apoB/apoA1 in individuals with positive CD serology suggest that unfavorable lipid profiles may explain a portion of this increased risk for CVD. Considering that adherence to a gluten-free diet has been shown to increase levels of HDL-cholesterol as well as decrease the ratio of total-to-HDL-cholesterol in newly diagnosed individuals (Brar et al. 2006; Lewis et al. 2009; De Marchi et al. 2013), this highlights the importance of timely diagnosis and treatment with a gluten-free diet in individuals with CD.

We observed significantly lower levels of retinol in individuals with positive CD serology. Although short-term dietary intake is understood to have little influence on circulating retinol levels, this association remained significant after adjustment for dietary intake of vitamin A, suggesting that any differences in intake do not explain the observed difference in circulating levels. Lower levels of serum retinol and a higher prevalence of retinol deficiency have been previously reported in individuals with CD (Wierdsma et al. 2013). Although no individuals with positive CD serology were retinol deficient (<0.70 μmol/L) (Sommer, Davidson, and Annecy 2002), our results support the notion of maldigestion and malabsorption of certain lipid-soluble compounds in individuals with undiagnosed CD. Alternatively, serum retinol concentrations are known to decrease during the acute phase response (Stephensen 2001), and it is possible that inflammation in the intestinal mucosa in individuals with positive CD serology could have
contributed to this observation. However, serum concentrations of C-reactive protein were not elevated in individuals with positive CD serology in the present study, suggesting the lack of a substantial acute phase response in such individuals.

We observed an association between positive CD serology and lower levels of circulating 25OHD in unadjusted models; however, this association was not significant in fully adjusted models accounting for intake of vitamin D, season of blood draw, physical activity, and hormonal contraceptive use among women. Nevertheless, lower levels of serum vitamin D have been previously reported in individuals with untreated CD (Corazza et al. 1996; Kemppainen et al. 1999). This could be due to malabsorption or a consequence of either reduced sun exposure or lower intake of vitamin D from dietary sources, which is particularly possible in CD considering the high prevalence of secondary lactose intolerance (Bode and Gudmand-Hoyer 1988).

Regardless of the mechanism, lower levels of 25OHD are of particular concern considering the adverse health outcomes that have been associated with low vitamin D (Wacker and Holick 2013). Indeed, individuals with untreated CD are known to be at elevated risk for osteoporosis (Kemppainen et al. 1999), and it is possible that lower levels of serum vitamin D might explain a portion of this increased risk.

We did not observe differences in various tocopherols or carotenoids in individuals with positive CD serology. Levels of α-, δ-, and γ-tocopherol did not significantly differ between those with positive and negative CD serology as previously reported (Hozyasz, Chelchowska, and Laskowska-Klita 2003); however, we did observe a trend where individuals with positive CD serology had lower levels of all three tocopherols. No significant associations or clear trend was observed for any of the carotenoids analyzed. This is in agreement with findings from patients with controlled CD (Ward, Zhao, and Bernstein 2008), and our results suggest that those with
positive CD serology but no diagnosis of CD may not experience severe malabsorption resulting in lower carotenoid concentrations.

Individuals with positive CD serology did not have markedly lower levels of most biomarkers of nutritional status assessed. Furthermore, we observed no significant increase in C-reactive protein, a marker of systemic inflammation, in those with positive CD serology as previously reported in some diagnosed cases of CD (Bayar et al. 2016). This suggests that individuals with positive CD serology identified through screening studies may fall into the mild end of the CD spectrum. Indeed, cases of CD diagnosed in adulthood are more likely to be of the ‘atypical’, ‘silent’ or even ‘latent’ form (Guandalini and Assiri 2014). It stands to reason that those with undiagnosed CD would most likely present with mild symptoms and a lesser degree of mucosal damage than those who ultimately seek a diagnosis. Furthermore, individuals in the present study population with potential undiagnosed CD may be particularly likely to represent mild cases of CD. Both studies had a certain degree of participant burden and it is possible that individuals with increasingly symptomatic forms of undiagnosed CD would avoid participation in such studies. Nevertheless, these findings suggest that many cases of undiagnosed CD identified through screening studies may present with mild mucosal damage and minor signs of malabsorption. This represents is an important consideration when assessing the benefits and drawbacks to population screening for CD.

The present study has some limitations. Although the total number of individuals in the study was relatively large, the number of individuals with positive CD serology was limited. This likely resulted in reduced statistical power to detect associations with the various biomarkers analyzed. Furthermore, biopsy samples were not available for individuals with positive CD serology and definitive diagnoses of CD can therefore not be made. Additionally, serum values
for select vitamins and minerals in which individuals with CD are commonly deficient, including vitamin B12 and iron (Dahele and Ghosh 2001; Halfdanarson, Litzow, and Murray 2007), were not available in our study populations. Finally, the findings reported were not adjusted for multiple testing despite the assessment of multiple outcome variables. However, outcome variables in the present study were selected based on an a priori hypothesis that undiagnosed CD may be associated with altered levels of these biomarkers. Furthermore, many of the biomarkers of health and nutritional status assessed are highly correlated, and accounting for multiple testing by treating these as independent tests would lead to an inflated type II error rate.

In summary, we assessed various biomarkers of health and nutritional status in those with positive CD serology identified through a screening study of Canadian adults. Individuals with positive CD serology but no diagnosis of CD had lower levels of HDL-cholesterol, apo-A1, and a higher ratio of total-to-HDL cholesterol and apoB/apoA1. Compared to individuals with negative CD serology, individuals with likely undiagnosed CD also had lower circulating levels of retinol. No significant differences in circulating levels of various carotenoids, tocopherols or ascorbic acid between those with positive and negative CD serology were observed. These results suggest that undiagnosed cases of CD identified through screening may be associated with altered lipid profiles and compromised absorption of certain fat-soluble vitamins. However, severe nutrient deficiencies were not evident in individuals with positive CD serology, suggesting more mild manifestations of CD in such cases.
CHAPTER 4:

GLUTEN INTAKE AND PLASMA PROTEOMIC BIOMARKERS IN YOUNG ADULTS WITHOUT CELIAC DISEASE

Adapted from:

Chapter 4: Gluten Intake and Plasma Proteomic Biomarkers in Young Adults Without Celiac Disease

4.1 Abstract

**Background:** Gluten-free foods have increased in popularity over the past decade and are now being consumed by individuals without celiac disease. However, the physiological effects of gluten intake in individuals without celiac disease remain unknown. High-abundance plasma proteins involved in inflammation, endothelial function and other physiological pathways may represent potential biomarkers of biological effects of gluten intake.

**Objective:** The objective was to examine the association between gluten intake and plasma proteomic biomarkers in a population of adults without celiac disease.

**Methods:** Subjects (n=998) were participants of the Toronto Nutrigenomics and Health Study, a cross-sectional examination of young adults aged 20-29 years. All individuals with a clinical diagnosis of celiac disease or positive celiac disease serology were excluded. Dietary gluten intake was estimated using a one month 196-item semi-quantitative food frequency questionnaire. The concentrations of 54 plasma proteins were measured simultaneously by liquid chromatography/multiple-reaction monitoring mass spectrometry. The association between gluten intake and each proteomic biomarker was examined using general linear models.

**Results:** Increased gluten intake was associated with increased concentrations of plasma α2-macroglobulin (p=0.004), a marker of inflammation and cytokine release. The association remained after adjusting for age, sex, BMI, ethnicity, physical activity, energy intake, fiber intake, and hormonal contraceptive use among women.

**Conclusion:** Gluten consumption is associated with increased plasma α2-macroglobulin in young
adults without celiac disease, suggesting possible effects of gluten on markers of immune function.
4.2 Introduction

Gluten is a protein found in wheat, barley, and rye. Sensitivity to gluten is thought to cause a number of adverse health effects, including gastrointestinal discomfort, skin rashes, and respiratory problems (Anderson et al. 2013). Celiac disease (CD) represents the most severe form of gluten sensitivity and is characterized by injury to the intestinal mucosa and nutrient malabsorption. The genetic basis and pathogenesis of CD have been well characterized. Genetic variants in the human leukocyte antigen (HLA)-DQ locus are necessary, but not sufficient to develop CD (Kagnoff 2007; Pietzak et al. 2009). Other factors, such as timing of initial exposure to gluten and duration of breastfeeding may be involved in the development of CD (Liu et al. 2014; Norris et al. 2005), but it remains unclear whether these factors permanently decrease risk of developing CD or simply delay the onset of symptoms (Lionetti et al. 2014; Pinto-Sanchez, Verdu, et al. 2016).

There is growing evidence to suggest that milder forms of gluten sensitivity may be responsible for the gluten-related symptoms seen in some individuals without clinically diagnosed CD (Sapone et al. 2011; Biesiekierski et al. 2011; Vazquez-Roque et al. 2013; Di Sabatino et al. 2015). Non-celiac gluten sensitivity (NCGS) is defined as gluten sensitivity in the absence of villous atrophy and the hallmark IgA antibodies associated with CD (Catassi et al. 2013). Individuals with NCGS experience many of the same symptoms upon the ingestion of gluten as those with CD, such as diarrhea, abdominal pain, fatigue, and headaches (Catassi et al. 2013; Lundin and Alaedini 2012). A diagnosis of NCGS is made when the re-introduction of gluten to the diet causes a recurrence of symptoms not previously present during a period of gluten exclusion (Sapone et al. 2012; Grace-Farfaglia 2014). NCGS is thought to affect up to 1% of
individuals in the general population (Volta et al. 2014; Capannolo et al. 2015), similar to estimates for the prevalence of CD (Kagnoff 2007; Gujral, Freeman, and Thomson 2012).

Although interest in NCGS has increased significantly over the past 10 years, the pathogenesis and possible genetic basis of the condition remain unknown (Sapone et al. 2012). This is largely due to the lack of biomarkers for NCGS as well as limited understanding of any physiological effects of gluten intake in those without CD. Advances in plasma proteomics have made it possible to quantify multiple proteins involved in inflammation and endothelial function, yielding potential markers of biological responses to certain dietary exposures like gluten. Indeed, highly-abundant plasma proteins include various physiologically important molecules and are useful biomarkers of disease and dietary exposures (Garcia-Bailo et al. 2012). The composition of the plasma proteome differs under various physiological conditions, and it is thought that most metabolic abnormalities in the body are reflected in the plasma proteome (Anderson 2005). A recently developed multiple reaction monitoring (MRM) proteomics assay has been used to measure levels of 54 highly abundant plasma proteins (Kuzyk et al. 2009), which we have previously shown to be associated with individual nutrients and distinct dietary patterns (Garcia-Bailo et al. 2012; Da Costa et al. 2013b, 2013a). It is possible that gluten-induced physiological effects are reflected in the concentrations of these proteins. The objective of this study was to explore the association between gluten intake and plasma proteomic biomarkers in an ethnically diverse population of young adults without CD.
4.3 Methods

4.3.1 Study Population

Subjects \((n=1,034)\) were participants of the Toronto Nutrigenomics and Health (TNH) study belonging to major ethnocultural groups with available proteomics data. The TNH study is a cross-sectional analysis of men and women of diverse ethnocultural backgrounds, aged 20-29 years, who were recruited from the University of Toronto campus between October 2004 and December 2010. The study was approved by the University of Toronto Ethics Review Board. Subjects completed a one month, 196-item semi-quantitative food frequency questionnaire (FFQ), a general health and lifestyle questionnaire, a physical activity questionnaire, and provided a fasting blood sample. Women who were pregnant or breastfeeding, and individuals who were fasting for religious reasons were excluded from the study. Subjects who were likely over-reporters (>3500 kcal per day for women or >4500 kcal per day for men) or under-reporters (<800 kcal per day) of total energy intake or had incomplete FFQ data were excluded from the analyses \((n=12)\). We also excluded individuals who self-reported having a medical diagnosis of CD \((n=2)\), wheat allergy, or following gluten-restricted diets \((n=5)\), as well as those with positive or equivocal tissue transglutaminase (tTG) serology indicative of undiagnosed CD \((n=24)\). See Chapter 2.3.3 for more details on tTG measurements. After exclusions, 998 subjects (299 men and 699 women) remained.

4.3.2 Dietary Assessment

Please refer to Chapter 2, section 2.3.4.
4.3.3 Proteomics Measurements

The concentrations of high-abundance proteins were measured at the University of Victoria – Genome British Columbia Proteomics Centre (Victoria, BC, Canada), using an MRM assay as described previously (Garcia-Bailo et al. 2012), initially for 45 proteins (Kuzyk et al. 2009), but later expanded to 63 proteins. Proteins with intra-assay coefficients of variation $\geq 15\%$ ($n=9$) were excluded from the analyses. The remaining 54 proteins were retained for analysis. The panel consisted of proteins involved in diverse physiological processes including inflammation and immune cell function.

4.3.4 Statistical Analysis

All statistical analyses were conducted using SAS Statistical Analysis Software v.9.2 (SAS Institute Inc, Cary, NC). The $\alpha$ error was set at 0.05 and $p$-values are two-sided. Variables that were not normally distributed were either loge- or square root- transformed prior to analysis to improve normality. The $p$-values from models using the transformed protein concentrations are reported, but untransformed means and measures of spread are reported to facilitate interpretability.

Subject characteristics across tertiles of gluten intake were compared using $\chi^2$ tests and general linear models (GLM) for categorical and continuous variables, respectively. For continuous variables, differences in means across tertiles of gluten intake were assessed using Duncan’s multiple range post-hoc test. The association between energy-adjusted gluten intake expressed as
a continuous variable and the 54 proteomics biomarkers was initially explored using unadjusted GLMs (Model 1). GLMs were then conducted using a multivariate model adjusted for age, sex, BMI, ethnicity, physical activity, energy intake, fiber intake, and hormonal contraceptive use among women (Model 2). The Benjamini-Yekutieli (B-Y) procedure \(\frac{\alpha}{\sum(1/i)}\), where \(i\) varies from 1 to the total number of tests conducted] was applied to account for multiple testing \((p<0.01\), calculated based on 54 proteins and \(\alpha=0.05\)). The B-Y method was selected because it allows for potential dependence between tests, and many of the proteins in the proteomics panel are biologically related (Narum 2006; Garcia-Bailo, Jamnik, Da Costa, Badawi, et al. 2013). The effect of other potential dietary confounders, including intake of food groups significantly associated with gluten and n-3 fatty acids, was also explored. Differences in mean protein concentration across tertiles of gluten intake were then assessed using Duncan’s multiple range post-hoc test for proteins significant at the B-Y threshold.

### 4.4 Results

Subject characteristics across tertiles of energy-adjusted gluten intake are shown in Table 4.1. Gluten intake differed significantly \((p<0.01)\) between ethnocultural groups, with more Caucasians consuming in the upper tertiles of gluten intake than any of the other groups. There were no significant differences in any of the biomarkers of cardiometabolic disease across tertiles of gluten intake. Gluten intake was positively associated with dietary fiber intake \((p<0.01)\). Gluten intake was also significantly associated with distinct dietary patterns. Intakes of grain products and sweets/baked goods increased with increasing tertile of gluten consumption.
(p<0.01), while intake of fruits/vegetables and meat/alternatives decreased as gluten consumption increased (p<0.01).

The associations between gluten intake and each plasma proteomics biomarker is shown in Table 4.2. The first analysis (Model 1) was adjusted for energy intake only. α2-Macroglobulin, α1B-glycoprotein, transferrin, inter-α-trypsin inhibitor heavy chain, complement C3, coagulation factor XIIa heavy chain, α1-anti-trypsin, apolipoprotein E, vitamin D-binding protein and angiotensinogen were significantly associated with gluten intake at the B-Y threshold (p<0.01, calculated based on 54 proteins and α=0.05) for this model. Of these proteins, only α2-macroglobulin remained significantly (p=0.004) associated with gluten intake after additional adjustments for age, sex, BMI, ethnicity, physical activity, fiber intake, and hormonal contraceptive use among women (Model 2). Plasma concentrations of α2-macroglobulin increased as gluten intake increased. The association between gluten and α2-macroglobulin remained significant when dietary patterns significantly associated with gluten intake (total servings of fruits/vegetables, grain products, meat/alternatives, and sweets/baked goods) as well as dietary n-3 fatty acids were included in Model 2 instead of dietary fiber. Furthermore, intakes of the food groups mentioned above were not significantly associated with α2-macroglobulin when gluten intake was adjusted for. No other significant associations were observed between gluten and any plasma proteins.
Table 4.1: Subject characteristics, serum biomarkers and dietary patterns by tertile of gluten intake.

<table>
<thead>
<tr>
<th></th>
<th>Tertile 1</th>
<th>Tertile 2</th>
<th>Tertile 3</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gluten intake range (g/1000 kcal per day)</td>
<td>0 - 3.6</td>
<td>3.7 - 5.6</td>
<td>5.6 - 14.5</td>
<td></td>
</tr>
<tr>
<td>Subjects (n)</td>
<td>333</td>
<td>333</td>
<td>332</td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>22.6 ± 0.14</td>
<td>22.7 ± 0.13</td>
<td>22.7 ± 0.13</td>
<td>0.84</td>
</tr>
<tr>
<td>Sex [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td>Female</td>
<td>230 (33)</td>
<td>245 (35)</td>
<td>224 (32)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>103 (34)</td>
<td>88 (29)</td>
<td>108 (36)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Caucasian</td>
<td>121 (24)</td>
<td>175 (35)</td>
<td>203 (41)</td>
<td></td>
</tr>
<tr>
<td>East Asian</td>
<td>171 (45)</td>
<td>122 (32)</td>
<td>91 (24)</td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>41 (36)</td>
<td>36 (31)</td>
<td>38 (33)</td>
<td></td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>22.62 ± 0.18</td>
<td>22.84 ± 0.20</td>
<td>22.52 ± 0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>73.38 ± 0.47</td>
<td>74.00 ± 0.51</td>
<td>73.67 ± 0.45</td>
<td>0.11</td>
</tr>
<tr>
<td>Physical Activity (MET-h)</td>
<td>7.42 ± 0.17</td>
<td>7.70 ± 0.17</td>
<td>7.56 ± 0.17</td>
<td>0.60</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>113.3 ± 0.63</td>
<td>113.5 ± 0.63</td>
<td>113.6 ± 0.60</td>
<td>0.79</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>68.68 ± 0.45</td>
<td>68.97 ± 0.44</td>
<td>68.89 ± 0.43</td>
<td>0.93</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>4.79 ± 0.02</td>
<td>4.77 ± 0.02</td>
<td>4.76 ± 0.02</td>
<td>0.93</td>
</tr>
<tr>
<td>Insulin (pmol/L)</td>
<td>47.37 ± 1.66</td>
<td>48.70 ± 1.59</td>
<td>46.21 ± 1.51</td>
<td>0.51</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>1.43 ± 0.05</td>
<td>1.45 ± 0.05</td>
<td>1.38 ± 0.05</td>
<td>0.53</td>
</tr>
<tr>
<td>HOMA-Beta</td>
<td>106.8 ± 4.04</td>
<td>112.5 ± 3.98</td>
<td>105.0 ± 3.21</td>
<td>0.61</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>4.25 ± 0.04</td>
<td>4.28 ± 0.04</td>
<td>4.15 ± 0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>HDL cholesterol (mmol/L)</td>
<td>1.57 ± 0.02</td>
<td>1.59 ± 0.02</td>
<td>1.55 ± 0.02</td>
<td>0.35</td>
</tr>
<tr>
<td>LDL cholesterol (mmol/L)</td>
<td>2.25 ± 0.03</td>
<td>2.27 ± 0.04</td>
<td>2.16 ± 0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>Total/HDL cholesterol (mmol/L)</td>
<td>2.82 ± 0.04</td>
<td>2.82 ± 0.04</td>
<td>2.81 ± 0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>Triglycerides (mmol/L)</td>
<td>0.95 ± 0.03</td>
<td>0.96 ± 0.02</td>
<td>0.97 ± 0.02</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Tertile 1</td>
<td>Tertile 2</td>
<td>Tertile 3</td>
<td>P-value</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>hs-CRP (mg/L)</td>
<td>1.20 ± 0.17</td>
<td>1.26 ± 0.13</td>
<td>1.17 ± 0.13</td>
<td>0.98</td>
</tr>
<tr>
<td>Free fatty acids (µmol/L)</td>
<td>508.3 ± 13.7</td>
<td>496.5 ± 13.5</td>
<td>476.0 ± 13.7</td>
<td>0.37</td>
</tr>
<tr>
<td>Fibre Intake (g/day)</td>
<td>22.52 ± 0.72a</td>
<td>22.77 ± 0.59a</td>
<td>25.05 ± 0.65b</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Fruits and Vegetables Intake (servings/day)</td>
<td>7.48 ± 0.25a</td>
<td>6.96 ± 0.21a</td>
<td>6.39 ± 0.21b</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Grain Products Intake (servings/day)</td>
<td>2.75 ± 0.08a</td>
<td>3.28 ± 0.08b</td>
<td>4.37 ± 0.11c</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dairy Intake (servings/day)</td>
<td>1.78 ± 0.08</td>
<td>2.15 ± 0.09</td>
<td>2.06 ± 0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>Meat and Alternatives Intake (servings/day)</td>
<td>2.07 ± 0.07a</td>
<td>1.68 ± 0.05b</td>
<td>1.56 ± 0.06b</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sweets and Baked Goods Intake (servings/day)</td>
<td>1.19 ± 0.06a</td>
<td>1.41 ± 0.06b</td>
<td>1.67 ± 0.08c</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

1Values shown are mean ± SE or n (%). Differences across tertiles of gluten intake were assessed using X² test for categorical variables and ANCOVA for continuous variables adjusting for energy intake, age, sex, BMI, ethnicity, physical activity, fibre intake, and hormonal contraceptive use among women. Different superscript letters indicate significant differences between tertiles of gluten intake (p<0.05). Duncan's multiple range test was used to adjust for multiple comparisons between groups. MET-h, metabolic equivalent task hours; hs-CRP, high sensitivity C-reactive protein; HLA, human leukocyte antigen.
Table 4.2: Plasma protein concentration by tertile of gluten intake in young adults without celiac disease\(^1\).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Tertile 1</th>
<th>Tertile 2</th>
<th>Tertile 3</th>
<th>Model 1 (p)</th>
<th>Model 2 (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_2)-macroglobulin</td>
<td>5.80 ± 0.09(^a)</td>
<td>5.81 ± 0.09(^a)</td>
<td>6.08 ± 0.10(^b)</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>(\alpha_1B)-glycoprotein</td>
<td>1.62 ± 0.03</td>
<td>1.68 ± 0.03</td>
<td>1.71 ± 0.03</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Transferrin</td>
<td>12.14 ± 0.17</td>
<td>12.64 ± 0.17</td>
<td>12.92 ± 0.18</td>
<td>0.003</td>
<td>0.07</td>
</tr>
<tr>
<td>Inter-(\alpha)-trypsin inhibitor HC</td>
<td>0.62 ± 0.01</td>
<td>0.62 ± 0.01</td>
<td>0.63 ± 0.01</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>(\beta_2)-glycoprotein I</td>
<td>2.76 ± 0.04</td>
<td>2.76 ± 0.04</td>
<td>2.86 ± 0.04</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Histidine-rich glycoprotein</td>
<td>1.35 ± 0.02</td>
<td>1.29 ± 0.02</td>
<td>1.36 ± 0.02</td>
<td>0.66</td>
<td>0.09</td>
</tr>
<tr>
<td>Hemopexin</td>
<td>9.93 ± 0.13</td>
<td>10.23 ± 0.11</td>
<td>10.26 ± 0.13</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>Complement C3</td>
<td>19.18 ± 0.27</td>
<td>19.73 ± 0.25</td>
<td>19.84 ± 0.26</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>Coagulation factor XIIa HC</td>
<td>0.23 ± 0.01</td>
<td>0.28 ± 0.01</td>
<td>0.29 ± 0.01</td>
<td>&lt;0.001</td>
<td>0.15</td>
</tr>
<tr>
<td>(\alpha_1)-Anti-trypsin</td>
<td>10.68 ± 0.16</td>
<td>11.26 ± 0.16</td>
<td>11.34 ± 0.17</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td>Vitamin D-binding protein</td>
<td>2.75 ± 0.04</td>
<td>2.86 ± 0.04</td>
<td>2.90 ± 0.04</td>
<td>0.004</td>
<td>0.24</td>
</tr>
<tr>
<td>Apolipoprotein E</td>
<td>0.53 ± 0.01</td>
<td>0.49 ± 0.01</td>
<td>0.49 ± 0.01</td>
<td>0.003</td>
<td>0.27</td>
</tr>
<tr>
<td>Gelsolin, isoform 1</td>
<td>1.20 ± 0.02</td>
<td>1.21 ± 0.02</td>
<td>1.23 ± 0.02</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>(\alpha_2)-HS-glycoprotein</td>
<td>8.75 ± 0.12</td>
<td>8.86 ± 0.11</td>
<td>8.91 ± 0.11</td>
<td>0.08</td>
<td>0.29</td>
</tr>
<tr>
<td>Heparin cofactor II</td>
<td>0.68 ± 0.01</td>
<td>0.70 ± 0.01</td>
<td>0.71 ± 0.01</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>(\alpha_1)-acid glycoprotein 1</td>
<td>1.74 ± 0.04</td>
<td>1.74 ± 0.03</td>
<td>1.78 ± 0.04</td>
<td>0.16</td>
<td>0.33</td>
</tr>
<tr>
<td>Apolipoprotein A-I</td>
<td>43.43 ± 0.58</td>
<td>44.21 ± 0.57</td>
<td>43.61 ± 0.55</td>
<td>0.59</td>
<td>0.35</td>
</tr>
<tr>
<td>Fibrinopeptide A</td>
<td>6.82 ± 0.12</td>
<td>7.24 ± 0.16</td>
<td>7.04 ± 0.16</td>
<td>0.54</td>
<td>0.37</td>
</tr>
<tr>
<td>Apolipoprotein C-III</td>
<td>2.37 ± 0.05</td>
<td>2.45 ± 0.04</td>
<td>2.41 ± 0.05</td>
<td>0.88</td>
<td>0.37</td>
</tr>
<tr>
<td>Fibrinogen (\beta) chain</td>
<td>9.16 ± 0.17</td>
<td>9.85 ± 0.22</td>
<td>9.50 ± 0.23</td>
<td>0.55</td>
<td>0.38</td>
</tr>
<tr>
<td>Vitronectin</td>
<td>3.61 ± 0.05</td>
<td>3.81 ± 0.05</td>
<td>3.75 ± 0.05</td>
<td>0.05</td>
<td>0.39</td>
</tr>
<tr>
<td>Angiotensinogen</td>
<td>0.85 ± 0.03</td>
<td>1.03 ± 0.04</td>
<td>1.01 ± 0.04</td>
<td>0.004</td>
<td>0.39</td>
</tr>
<tr>
<td>Afamin</td>
<td>0.25 ± 0.004</td>
<td>0.26 ± 0.003</td>
<td>0.26 ± 0.003</td>
<td>0.36</td>
<td>0.42</td>
</tr>
<tr>
<td>Prothrombin</td>
<td>0.58 ± 0.01</td>
<td>0.58 ± 0.01</td>
<td>0.58 ± 0.01</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>Zinc-(\alpha_2)-glycoprotein</td>
<td>1.03 ± 0.02</td>
<td>1.05 ± 0.02</td>
<td>1.06 ± 0.02</td>
<td>0.18</td>
<td>0.46</td>
</tr>
<tr>
<td>Complement C4 (\gamma) chain</td>
<td>1.60 ± 0.03</td>
<td>1.56 ± 0.03</td>
<td>1.57 ± 0.03</td>
<td>0.75</td>
<td>0.48</td>
</tr>
<tr>
<td>Kininogen-1</td>
<td>2.08 ± 0.03</td>
<td>2.19 ± 0.03</td>
<td>2.20 ± 0.03</td>
<td>0.01</td>
<td>0.48</td>
</tr>
<tr>
<td>Antithrombin-III</td>
<td>3.59 ± 0.03</td>
<td>3.54 ± 0.03</td>
<td>3.61 ± 0.04</td>
<td>0.59</td>
<td>0.49</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Fibrinogen γ chain</td>
<td>9.08 ± 0.17</td>
<td>9.79 ± 0.24</td>
<td>9.48 ± 0.25</td>
<td>0.61</td>
<td>0.51</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>0.51 ± 0.04</td>
<td>0.67 ± 0.06</td>
<td>0.61 ± 0.07</td>
<td>0.77</td>
<td>0.54</td>
</tr>
<tr>
<td>Albumin, serum</td>
<td>971.39 ± 8.63</td>
<td>952.25 ± 8.18</td>
<td>968.94 ± 8.79</td>
<td>0.91</td>
<td>0.55</td>
</tr>
<tr>
<td>Apolipoprotein A-II precursor</td>
<td>24.51 ± 0.31</td>
<td>25.69 ± 0.32</td>
<td>25.46 ± 0.31</td>
<td>0.08</td>
<td>0.56</td>
</tr>
<tr>
<td>Apolipoprotein A-IV</td>
<td>1.40 ± 0.02</td>
<td>1.45 ± 0.02</td>
<td>1.46 ± 0.03</td>
<td>0.20</td>
<td>0.57</td>
</tr>
<tr>
<td>Complement C1 inactivator</td>
<td>4.84 ± 0.06</td>
<td>4.60 ± 0.07</td>
<td>4.67 ± 0.07</td>
<td>0.37</td>
<td>0.57</td>
</tr>
<tr>
<td>α1-antichymotrypsin</td>
<td>3.37 ± 0.05</td>
<td>3.31 ± 0.04</td>
<td>3.43 ± 0.05</td>
<td>0.26</td>
<td>0.58</td>
</tr>
<tr>
<td>Complement factor H</td>
<td>0.59 ± 0.01</td>
<td>0.60 ± 0.01</td>
<td>0.60 ± 0.01</td>
<td>0.11</td>
<td>0.59</td>
</tr>
<tr>
<td>Plasminogen</td>
<td>1.20 ± 0.01</td>
<td>1.26 ± 0.02</td>
<td>1.24 ± 0.02</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>Apolipoprotein C-I lipoprotein</td>
<td>3.21 ± 0.05</td>
<td>3.27 ± 0.05</td>
<td>3.22 ± 0.05</td>
<td>0.72</td>
<td>0.60</td>
</tr>
<tr>
<td>Serum amyloid P-component</td>
<td>0.44 ± 0.01</td>
<td>0.44 ± 0.01</td>
<td>0.45 ± 0.01</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>α2-antiplasmin</td>
<td>1.92 ± 0.02</td>
<td>1.93 ± 0.02</td>
<td>1.94 ± 0.02</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Ceruloplasmin</td>
<td>2.19 ± 0.06</td>
<td>2.39 ± 0.06</td>
<td>2.34 ± 0.06</td>
<td>0.10</td>
<td>0.62</td>
</tr>
<tr>
<td>Plasma retinol-binding protein</td>
<td>0.90 ± 0.01</td>
<td>0.94 ± 0.02</td>
<td>0.94 ± 0.02</td>
<td>0.06</td>
<td>0.62</td>
</tr>
<tr>
<td>Complement factor B</td>
<td>1.42 ± 0.02</td>
<td>1.45 ± 0.02</td>
<td>1.44 ± 0.02</td>
<td>0.28</td>
<td>0.63</td>
</tr>
<tr>
<td>Fibrinogen α chain</td>
<td>11.56 ± 0.24</td>
<td>12.41 ± 0.34</td>
<td>11.93 ± 0.34</td>
<td>0.83</td>
<td>0.64</td>
</tr>
<tr>
<td>Complement C4 β chain</td>
<td>1.45 ± 0.03</td>
<td>1.43 ± 0.03</td>
<td>1.43 ± 0.03</td>
<td>0.92</td>
<td>0.65</td>
</tr>
<tr>
<td>Complement C9</td>
<td>2.71 ± 0.05</td>
<td>2.70 ± 0.05</td>
<td>2.71 ± 0.05</td>
<td>0.88</td>
<td>0.66</td>
</tr>
<tr>
<td>L-selectin</td>
<td>0.07 ± 0.001</td>
<td>0.07 ± 0.001</td>
<td>0.07 ± 0.001</td>
<td>0.46</td>
<td>0.72</td>
</tr>
<tr>
<td>Transthyretin</td>
<td>5.71 ± 0.07</td>
<td>5.74 ± 0.07</td>
<td>5.86 ± 0.07</td>
<td>0.23</td>
<td>0.76</td>
</tr>
<tr>
<td>Apolipoprotein B-100</td>
<td>0.79 ± 0.01</td>
<td>0.81 ± 0.01</td>
<td>0.79 ± 0.01</td>
<td>0.84</td>
<td>0.76</td>
</tr>
<tr>
<td>Apolipoprotein D</td>
<td>0.34 ± 0.005</td>
<td>0.34 ± 0.005</td>
<td>0.35 ± 0.005</td>
<td>0.62</td>
<td>0.88</td>
</tr>
<tr>
<td>Haptoglobin β chain</td>
<td>10.85 ± 0.31</td>
<td>10.30 ± 0.27</td>
<td>10.60 ± 0.30</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>Adiponectin</td>
<td>0.06 ± 0.002</td>
<td>0.07 ± 0.002</td>
<td>0.07 ± 0.002</td>
<td>0.26</td>
<td>0.94</td>
</tr>
<tr>
<td>Clusterin</td>
<td>1.51 ± 0.02</td>
<td>1.52 ± 0.02</td>
<td>1.53 ± 0.02</td>
<td>0.69</td>
<td>0.97</td>
</tr>
<tr>
<td>Apolipoprotein L1</td>
<td>0.40 ± 0.01</td>
<td>0.41 ± 0.01</td>
<td>0.42 ± 0.01</td>
<td>0.16</td>
<td>0.99</td>
</tr>
</tbody>
</table>

1 All values are crude means ± SEs. Results were obtained using general linear models with gluten intake as a continuous variable. Model 1 was adjusted for energy intake. Model 2 was adjusted for energy intake, age, sex, BMI, ethnicity, physical activity, fibre intake, and hormonal contraceptive use among women. Proteins are listed in order of statistical significance under Model 2. Different superscript letters indicate significant differences between tertiles of gluten intake (p<0.05). Duncan's Multiple Range post hoc was used to adjust for multiple comparisons between groups. HC, heavy chain; HS, Heremans-Schmid.
4.5 Discussion

There is growing interest in gluten sensitivity beyond CD (Catassi et al. 2013), and the last decade has seen a surge in the popularity of gluten-free foods far greater than the prevalence of CD would appear to warrant (Sapone et al. 2012). Recent studies assessing NCGS have been somewhat inconclusive (Biesiekierski et al. 2011; Biesiekierski et al. 2013), and this has led to uncertainty over the existence of such a condition. To date, no studies have identified any possible biomarkers or elucidated the mechanisms behind NCGS. Furthermore, the broad physiological effects of gluten intake in healthy populations have not been previously investigated. The present study aimed to assess the association between gluten intake and plasma proteomics biomarkers in an ethnically diverse population of young adults. A mass-spectrometry based MRM proteomics assay was used to determine the concentrations of 54 plasma proteins involved in various physiological processes. We identified a significant association between gluten intake and α2-macroglobulin. The observed association was not influenced by cases of CD as individuals with previously reported medical diagnoses of CD, as well as those with positive tTG serology indicative of undiagnosed CD were excluded from the present study. We also identified significant differences in patterns of gluten intake between Caucasians and other ethnocultural groups, as well as an association between gluten and distinct dietary patterns.

We observed a significant positive association between plasma concentrations of α2-macroglobulin and gluten consumption. Plasma α2-macroglobulin is an acute phase reactant that increases in response to inflammation. It is produced to protect against the toxic effects of the over-production of pro-inflammatory cytokines (Mocchegiani et al. 2006). Specifically, α2-macroglobulin binds and inactivates these cytokines and downregulates the T cell response of the immune system (Borth 1994). Alternatively, α2-macroglobulin has been shown to protect certain
cytokines from degradation and enable their delivery to immune cells (Rehman, Ahsan, and Khan 2013). $\alpha_2$-Macroglobulin is also an inhibitor of proteinases released during inflammation, such as matrix metalloproteinases (MMPs) (Mocchegiani et al. 2006). In CD, the antigen presenting cells (APC) of genetically susceptible individuals have a high affinity for gluten peptides in the gut. When gluten is presented to the APCs, it binds and triggers a number of downstream immunological effects. Gluten-presenting APCs encounter gluten-specific pathogenic CD4$^+$ T cells, which then proliferate and release interferon (IFN)-$\gamma$ and other TH1 cytokines (Kagnoff 2007). These pro-inflammatory molecules act to alter intestinal permeability and result in the activation and release of MMPs that damage the intestinal mucosa (Kagnoff 2007; Pender et al. 1997; Daum et al. 1999). Gluten consumption in individuals with CD is also known to lead to an increase in intestinal CD8$^+$ intraepithelial lymphocytes, which further contribute to the intestinal damage associated with this condition (Han et al. 2013). Our finding of higher $\alpha_2$-macroglobulin in healthy individuals with higher gluten consumption suggests that gluten may also have effects on the immune system in individuals without CD.

It is possible that individuals without CD experience an immune response after gluten consumption that is less severe than the response observed in individuals with frank CD. It has been shown that gluten triggers the release of zonulin and subsequent increases in intestinal permeability in individuals with CD and in healthy controls (Drago et al. 2006). Gluten has also been shown to cause the release of tumor necrosis factor (TNF)-$\alpha$, IFN-$\gamma$, interleukin (IL)-6, IL-8, IL-10 and IL-13 in peripheral blood mononuclear cells from both CD patients and healthy individuals (Lammers et al. 2011). Modulators of immune functions like $\alpha_2$-macroglobulin may interact with these cytokines associated with gluten exposure. Indeed, levels of $\alpha_2$-macroglobulin have been shown to increase in response to IFN-$\gamma$ (Fabrizi et al. 1994), although serum IFN-$\gamma$ was not significantly associated with $\alpha_2$-macroglobulin or gluten intake in the present study. $\alpha_2$-
Macroglobulin can also bind to a number of cytokines, including TNF-α, IL-6, IL-8 and IL-10 (Mocchegiani et al. 2006; Kurdowska et al. 1997; Garber, Gonias, and Webb 2000). This may facilitate the delivery of these cytokines to host immune cells as part of an immune response. The binding of such cytokines to α2-macroglobulin may also lead to their clearance from the circulation through receptor-mediated pathways (Rehman, Ahsan, and Khan 2013; Mocchegiani et al. 2006). It is possible that the dysregulation of such pathways could be associated with the development of adverse symptomatic responses to gluten intake in individuals without CD. Additionally, the observed association between gluten and α2-macroglobulin was demonstrated over a relatively large range of gluten intakes (0-15g/1000 kcal per day). If markers like α2-macroglobulin are indicative of normal biological responses to gluten intake in those without CD, this suggests that such responses are likely observable when dietary gluten intakes are much higher than the trace amounts of gluten (10-50mg) that can trigger symptoms in individuals with CD (Catassi et al. 2007; Akobeng and Thomas 2008).

Gluten consumption was also positively associated with fiber intake as well as distinct dietary patterns. Many gluten-rich foods are good sources of dietary fiber, and adherence to a gluten-free diet has previously been associated with inadequate intakes of many nutrients (Grace-Farfaglia 2014), including fiber (Wild et al. 2010). Fiber intake is known to be associated with a decreased risk for a number of chronic conditions (Anderson et al. 2009), suggesting that gluten intake may be associated with beneficial effects due to its association with dietary fiber. Further studies on the long-term effects of gluten-free and gluten-reduced diets are necessary to determine whether individuals adopting such diets are at an increased risk of developing various chronic conditions associated with suboptimal fiber intakes.
The present study has some limitations. The study population consisted of self-selected university students and may not be representative of the general population. Our FFQ was not designed specifically to assess dietary gluten intake. However, we identified food items as gluten-containing if all associated sub-items were solely made with wheat, barley, or rye. Items that included foods made from other cereal grains in addition to those (eg. ‘Other grain products’) were not considered gluten-containing. We also utilized previously reported methods to estimate the approximate gluten content of each gluten-containing food item (van Overbeek et al. 1997; Hopman et al. 2012; Hopman et al. 2007). Nevertheless, it remains possible that our method for estimating gluten content captured another compound present in these gluten-containing foods, such as fermentable, oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs), which have been shown to confound symptoms associated with gluten sensitivity (Biesiekierski et al. 2013). Additionally, our FFQ did not contain any items specifically designated as gluten-free since such products were not as widely available during the time subjects were recruited. However, we accounted for this by excluding individuals with CD or wheat allergy, as well as those who indicated that they were on special gluten-restricted diets. The long-term storage of plasma samples could have affected the proteomics results, although we observed no significant difference in average sample storage time across tertiles of gluten intake. It is possible that the associations observed in the present study are due to residual confounding. However, our final proteomics analysis was adjusted for energy intake, age, sex, BMI, ethnicity, physical activity, fiber intake, and hormonal contraceptive use among women. A number of these covariates have previously been shown to modify levels of the plasma proteins in our proteomics panel (Garcia-Bailo et al. 2012; Josse et al. 2012). Furthermore, it is not likely that gluten intake served as a marker for other dietary components already known to cause inflammation as it was not associated with hs-CRP, a marker of systemic inflammation that has
been used to identify inflammatory dietary patterns (Cavicchia et al. 2009). Nevertheless, it is possible that the results obtained were influenced by unidentified residual confounders. Finally, the cross-sectional nature of this study precludes the establishment of causality between gluten intake and plasma concentrations of α2-macroglobulin.

In summary, we examined the association between gluten intake and plasma proteomics biomarkers in young adults without CD and identified a positive association between gluten intake and α2-macroglobulin, a proteomic marker with various immune-related functions. Further studies are necessary to determine whether gluten, and not other components of gluten-containing foods, stimulates the production of α2-macroglobulin. If gluten is, indeed, responsible for the observed increase in α2-macroglobulin, studies will be required to determine whether changes in α2-macroglobulin are related to symptoms of any adverse health conditions associated with gluten.
CHAPTER 5:

A GENOME-WIDE ASSOCIATION STUDY OF PLASMA PROTEOMIC BIOMARKERS OF GLUTEN INTAKE IN ADULTS WITHOUT CELIAC DISEASE
Chapter 5: A Genome-Wide Association Study of Plasma Proteomic Biomarkers of Gluten Intake in Adults Without Celiac Disease

5.1 Abstract

**Background:** Gluten intake has been shown to be associated with circulating levels of α2-macroglobulin, a protein involved in the inflammatory response and cytokine release, in young adults without celiac disease. However, it is unknown whether genetic variation modifies the association between gluten intake and plasma α2-macroglobulin.

**Objective:** The objective was to identify genetic variants that are associated with α2-macroglobulin which may modify the association between gluten intake and α2-macroglobulin in a population of young adults without celiac disease using a genome-wide approach.

**Methods:** A genome-wide association study (GWAS) was conducted on circulating levels of α2-macroglobulin in a population of Caucasians without celiac disease from the cross-sectional Toronto Nutrigenomics and Healthy Study (n=389). Concentrations of α2-macroglobulin were measured in plasma using a multiple reaction monitoring HPLC-MS/MS assay. Genome-wide genotyping was conducted using the Affymetrix 6.0 chip, and a total of 641,509 genotyped and 5,399,846 imputed single nucleotide polymorphisms (SNPs) with minor allele frequencies > 5% were included in all analyses. The association between genome-wide genetic variants and plasma α2-macroglobulin was explored using linear regression with an additive mode of inheritance. General linear models were then used to determine whether variants with the most significant genome-wide association modified the association between gluten intake and α2-macroglobulin.

**Results:** Circulating α2-macroglobulin was associated with rs7278004 in ICOSGL (p = 1.67x10⁻⁷), a T-cell stimulatory gene involved in cytokine release and various gastrointestinal immune-
related disorders. The positive association between gluten intake and \( \alpha_2 \)-macroglobulin was strongest in minor allele (C) homozygotes at rs7278004; however, the interaction between rs7278004 and gluten intake on \( \alpha_2 \)-macroglobulin was not statistically significant (p=0.12).

**Conclusion:** These results suggest that genes involved in T-cell proliferation and cytokine release may play a role in the physiological effects of gluten intake in individuals without celiac disease.
5.2 Introduction

Gluten is the term used to define a subset of the proteins found in wheat, barley, and rye. Celiac disease (CD), an autoimmune disorder which affects approximately 1% of the population, is triggered by the ingestion of dietary gluten. Strict adherence to a gluten-free diet is the only suitable treatment for individuals with CD (Gujral, Freeman, and Thomson 2012). Recently, there has been an increase in the popularity of gluten-free food products with the promotion of such foods to individuals without CD (Sapone et al. 2012; Fasano et al. 2015). Gluten sensitivity in the absence of CD is thought to affect a subset of patients with irritable bowel syndrome (IBS)-like symptoms. Recent randomized-controlled trials have provided some evidence for gluten-specific adverse effects in these patients (Biesiekierski et al. 2011; Di Sabatino et al. 2015; Zanini et al. 2015; Elli et al. 2016). However, some studies have suggested that other components of the diet may be responsible for triggering such adverse effects (Biesiekierski et al. 2013). The pathophysiological mechanisms by which gluten may cause adverse symptoms in individuals with potential non-celiac gluten sensitivity (NCGS) remain poorly understood.

To gain a better understanding of the potential pathways affected by gluten intake which may become dysregulated in individuals with NCGS, the effects of gluten intake on plasma proteomic biomarkers in individual without CD were explored (see chapter 4). Gluten intake was associated with increased levels of $\alpha_2$-macroglobulin, a protein involved in the removal of both pro- and anti-inflammatory cytokines from the systemic circulation through receptor-mediated pathways (Garber, Gonias, and Webb 2000; Borth 1994; Rehman, Ahsan, and Khan 2013; Mocchegiani et al. 2006; Kurdowska et al. 1997). Exposure to gluten has been shown to lead to increased secretion of such pro- and anti-inflammatory markers in both healthy controls and those with potential NCGS (Lammers et al. 2011; Hollon et al. 2015), suggesting that pathways
involved in the clearance of such markers represent possible responses to gluten intake which may become dysregulated in individuals with sensitivities to gluten.

While specific risk variants in the human leukocyte antigen (HLA) gene region are necessary for the development of CD, the potential genetic basis of gluten sensitivity in the absence of CD has not been determined. Furthermore, the influence of genetic variation on the physiological effects of gluten intake in individuals without CD remains unknown. One approach to identifying genetic variants that interact with external (eg. dietary) factors to modify an outcome variable is to examine genetic variants with known associations with the outcome (Gauderman et al. 2013; Kooperberg and Leblanc 2008). Plasma $\alpha_2$-macroglobulin represents a marker associated with gluten intake in those without CD; however, genetic associations with plasma $\alpha_2$-macroglobulin concentrations have not been extensively studied. While common variations in $A2M$, the gene that encodes $\alpha_2$-macroglobulin, are not associated with serum $\alpha_2$-macroglobulin levels (Rugsarash et al. 2006), other loci may be involved. The objective of this study was to identify novel genetic variants that are associated with circulating levels of $\alpha_2$-macroglobulin and determine whether any such variants modify the association between gluten intake and $\alpha_2$-macroglobulin in individuals without CD.

5.3 Methods

5.3.1 GWAS Cohort

Subjects were participants from the Toronto Nutrigenomics and Health (TNH) study. Recruitment methods and the nature of the TNH study participants have been described
previously (Brenner et al. 2010). Briefly, the TNH study is a cross-sectional examination of young adults between the ages of 20 and 29 years who were recruited from the University of Toronto Campus between the fall of 2004 and the winter of 2010. Subjects completed a food frequency questionnaire (FFQ), a general health and lifestyle questionnaire, and a physical activity questionnaire in addition to providing a fasting blood sample. Pregnant or breastfeeding women and individuals who could not provide a blood sample were excluded from the study. The University of Toronto Ethics Review Board approved the study protocol. The TNH GWAS Cohort is comprised of a subset of individuals of self-reported Caucasian ancestry (n=532). Caucasians were selected for the GWAS cohort because they represent the largest ethnocultural group in the TNH population. Individuals with positive or equivocal CD serology (n=10), and those with missing values for $\alpha_2$-macroglobulin (n=31) were excluded from the present study.

5.3.2 $\alpha_2$-Macroglobulin Measurements

Circulating $\alpha_2$-macroglobulin concentrations were measured from plasma samples using an HPLC-MS/MS multiple-reaction monitoring technique at the University of Victoria-Genome British Columbia Proteomics Centre as described previously (Kuzyk et al. 2009; Garcia-Bailo et al. 2012). See Chapter 4.3.3 for more details on the proteomics measurements.

5.3.3 Gluten Assessment

Please refer to Chapter 4, section 4.3.4.
5.3.4 Genotyping

Genome-wide genotyping for each individual in the TNH GWAS Cohort was conducted using the Affymetrix Genome-Wide Human SNP Array 6.0. The Affymetrix Chip contains probe sets for >900,000 common and rare single nucleotide polymorphisms (SNPs). Genotypes were called from the Affymetrix Chip using the Birdseed v2 algorithm (Korn et al. 2008) for 906,601 SNPs.

5.3.4.1 Quality Control

All quality control analysis was done using PLINK software v1.90. Subjects and genome-wide SNPs were evaluated and excluded based on a number of standard quality control assessments in GWAS studies. Subjects with a composite genome-wide call rate <85% (n=3), mean autosomal heterozygosity >5SDs from the population average (n=5), or disagreement between reported gender and imputed gender based on X chromosome heterozygosity (n=3) were excluded. Furthermore, identical-by-descent (IBD) analysis was done to test for relatedness among all possible pairs of samples, as the inclusion of related individuals in GWAS can lead to increased type I and type II errors (Devlin and Roeder 1999; Turner et al. 2011). A threshold of Pihat > 0.2 (an estimate of the proportion of shared genomic information) was used as the cutoff for cryptic relatedness (Seddon et al. 2013; Yu et al. 2011), and the individual with the lowest composite call rate was excluded when pairs of related individuals were identified. A total of 3 pairs of related individuals were identified (Pihat > 0.2), and 3 individuals were subsequently excluded. SNPs with a call rate <95% (n=15,781), not in Hardy Weinberg-equilibrium (HWE) p<10^{-8} (n=30,711), associated with sex p<10^{-8} (n=81), and those that were monomorphic in the study
population (n=48,134) were excluded. SNPs with allele frequencies that significantly differed (p<10^{-8}) from those observed in European individuals in 1000 Genomes Project were excluded (n=157). We also excluded genotyped SNPs with a minor allele frequency <5% (n=170,228) for all trait association analyses presented. After these quality control measures, 641,509 SNPs genotyped were available for analysis.

5.3.4.2 Population Structure

Principle components analysis (PCA) was conducted to assess for population structure using PLINK v1.9 software. SNPs in regions of long-range linkage disequilibrium (LD) (n=20,278), as well as SNPs with LD r^2>0.2 (n=623,400) were removed prior to principal component extraction. The top 3 principal components were extracted from our population of European ancestry for the remaining 154,225 SNPs (Bishop et al. 2009). Iterative removal of outliers (individuals ± 6SDs from the mean of any PC) was conducted (Price et al. 2006), and a total of 85 individuals were excluded. Remaining subjects were then seeded with individuals from 3 populations from the 1000 Genomes Project (Caucasians from Utah, Dai Chinese, and Gambian individuals) to confirm Caucasian ancestry. The remaining TNH subjects (n=389) showed good conformance to the PCA patterns generated by Caucasian and European individuals from the 1000 Genomes Project (figures S1 and S2), and were included in all subsequent analyses.

5.3.4.3 Imputation

Genome-wide imputation for autosomal chromosomes was conducted using the Michigan
Imputation Server (https://imputationserver.sph.umich.edu/index.html). Phasing was done using SHAPEIT software (Delaneau, Marchini, and Zagury 2011) and imputation was done using Minimac3 (Howie et al. 2012). European individuals from the 1000 Genomes Project (Phase 3 v5) were selected as the reference panel. A total of 14,196,621 variants were initially imputed across 22 autosomal chromosomes. After excluding imputed variants that were non-SNPs, had imputation R² ≤ 0.3 (a measure of imputation quality) and minor allele frequency <0.05, a total of 5,399,846 imputed SNPs were included in the final analysis.

5.3.5 Statistical Analysis

Genome-wide scans were conducted using SNPTEST software v2.5.2 (Marchini et al. 2007). Descriptive statistics and individual SNP analyses were conducted using SAS Statistical Analysis Software v.9.2 (SAS Institute Inc, Cary, NC). Variables that were not normally distributed were loge- or square root-transformed prior to analysis. Genome-wide significance was set at p<5x10⁻⁸, a threshold which corresponds to the estimated number of truly independent genomic loci in European individuals (Dudbridge and Gusnanto 2008). The p-values from models using transformed variables are reported, but untransformed means and measures of spread are reported to facilitate interpretability.

The association between autosomal genome-wide genetic variants and plasma α₂-macroglobulin concentrations were explored using linear regressions with an additive mode of inheritance. Imputed variants were treated as continuous allele dosages with values ranging from 0 to 2. With allelic dosages, 0 corresponds to an imputed variant with 100% chance of being homozygous for the major allele, and 2 corresponds to a 100% chance of the minor allele (de Bakker et al. 2008).
Initial genome-wide scans were run without adjustment for any covariates. Fully-adjusted scans were then run accounting for age, sex, BMI, and the top 3 principal components from the population structure analysis. Genomic inflation factor $\lambda$ was calculated to assess the potential of any residual effects of population substructure. The genomic inflation factor $\lambda$ was calculated by dividing the median observed test statistic by the expected median (Devlin and Roeder 1999), with deviations from a value of 1 being indicative of potential residual substructure (Clayton et al. 2005).

Variants that reached genome-wide significance ($p<5\times10^{-8}$) in any models were carried over to subsequent analyses to determine whether such variants modify the association between gluten intake and $\alpha_2$-maroglobulin. General linear models (GLMs) adjusted for age, sex, BMI, physical activity, hormonal contraceptive use among women and fiber intake were then used to determine whether stratifying by genotype of such SNP(s) modified the association between gluten intake and plasma $\alpha_2$-maroglobulin. Interactions between genome-wide significant SNP(s) and gluten intake on circulation concentrations of $\alpha_2$-maroglobulin were examined.

### 5.4 Results

Subject characteristics for the TNH GWAS cohort are shown in Table 5.1. The Manhattan plot for the results of the fully-adjusted GWAS on circulating $\alpha_2$-maroglobulin is shown in Figure 5.1. The quantile-quantile (q-q) plot for the fully-adjusted GWAS is shown in Figure 5.2. There was no evidence of residual population structure ($\lambda = 1.006$). Variation on chromosome 21 was most significantly associated with circulating concentrations of $\alpha_2$-maroglobulin. The top hits obtained using both unadjusted and adjusted models are summarized in Table 5.2. These
variants are all within the inducible T-cell costimulatory ligand (*ICOSLG*) or its immediately surrounding regions. A regional-association plot is presented in Figure 5.3. rs7278004, a variant in the 3’ untranslated region (UTR) of *ICOSLG*, was most significantly associated with α2-maroglobulin. Associations at rs7278004 reached genome-wide significance in unadjusted models (p=3.13x10^{-8}), but were slightly attenuated after adjusting for age, sex, BMI and the top 3 principal components of population structure analysis (p=1.67x10^{-7}). Carriers of the minor allele at rs7278004 (C) had higher levels of circulating α2-maroglobulin than carriers of the major allele (T). No other variants reached genome-wide significance in unadjusted or adjusted models.

Associations between gluten intake and α2-maroglobulin stratified by rs7278004 genotype are shown in Table 5.3. We observed a significant positive association (p=0.02) between gluten intake and plasma α2-maroglobulin in minor allele homozygotes (CC genotype) at rs7278004. Among carriers of the major allele (CT and TT genotypes), α2-maroglobulin concentrations increased with gluten intake, but the association was not statistically significant (p>0.05). However, there was no significant interaction between rs7278004 genotype and gluten intake on circulating levels of plasma α2-maroglobulin (p=0.12).
Table 5.1: Subject characteristics for TNH GWAS Study Population.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (n)</td>
<td>389</td>
</tr>
<tr>
<td>Age (y)</td>
<td>23.06 ± 0.13(^1)</td>
</tr>
<tr>
<td>Sex [n(%)]</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>105 (27)</td>
</tr>
<tr>
<td>Female</td>
<td>284 (73)</td>
</tr>
<tr>
<td>Body mass index (kg/m(^2))</td>
<td>23.18 ± 0.17</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>75.13 ± 0.42</td>
</tr>
<tr>
<td>Energy intake (kcal)</td>
<td>2004 ± 32</td>
</tr>
<tr>
<td>Gluten intake (g/1000 kcal)</td>
<td>5.41 ± 0.12</td>
</tr>
<tr>
<td>(\alpha_2)-macroglobulin (pmol/μL)</td>
<td>5.86 ± 0.09</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>115 ± 0.55</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>69 ± 0.39</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>4.72 ± 0.02</td>
</tr>
<tr>
<td>Cholesterol</td>
<td></td>
</tr>
<tr>
<td>Total (mmol/L)</td>
<td>4.21 ± 0.04</td>
</tr>
<tr>
<td>HDL (mmol/L)</td>
<td>1.61 ± 0.02</td>
</tr>
<tr>
<td>LDL (mmol/L)</td>
<td>2.17 ± 0.03</td>
</tr>
<tr>
<td>Total:HDL cholesterol</td>
<td>2.72 ± 0.03</td>
</tr>
</tbody>
</table>

\(^1\)Mean ± SE (all such values).
Table 5.2: Summary of variants associated with plasma α2-macroglobulin with the highest genome-wide significance.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos*</th>
<th>Gene‡</th>
<th>Predicted Function‡</th>
<th>Alleles</th>
<th>MAF</th>
<th>β ± SE(^{¥})</th>
<th>p(^{1})</th>
<th>(p^{2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7278004</td>
<td>21</td>
<td>45636881</td>
<td>ICOSLG</td>
<td>3’ UTR</td>
<td>C/T</td>
<td>43%</td>
<td>0.11 ± 0.02</td>
<td>3.13x10^-8</td>
<td>1.67x10^-7</td>
</tr>
<tr>
<td>rs58813804</td>
<td>21</td>
<td>45636792</td>
<td>ICOSLG</td>
<td>downstream</td>
<td>C/G</td>
<td>44%</td>
<td>0.11 ± 0.02</td>
<td>7.65x10^-8</td>
<td>2.65x10^-7</td>
</tr>
<tr>
<td>rs56161361</td>
<td>21</td>
<td>45636793</td>
<td>ICOSLG</td>
<td>downstream</td>
<td>A/G</td>
<td>43%</td>
<td>0.12 ± 0.02</td>
<td>8.51x10^-8</td>
<td>2.72x10^-7</td>
</tr>
<tr>
<td>rs7279792</td>
<td>21</td>
<td>45635719</td>
<td>-</td>
<td>-</td>
<td>A/G</td>
<td>42%</td>
<td>0.11 ± 0.02</td>
<td>6.93x10^-8</td>
<td>3.68x10^-7</td>
</tr>
<tr>
<td>rs8130872</td>
<td>21</td>
<td>45635062</td>
<td>-</td>
<td>-</td>
<td>T/C</td>
<td>42%</td>
<td>0.11 ± 0.02</td>
<td>7.49x10^-8</td>
<td>3.99x10^-7</td>
</tr>
<tr>
<td>rs8134498</td>
<td>21</td>
<td>45639368</td>
<td>ICOSLG</td>
<td>intronic</td>
<td>C/T</td>
<td>45%</td>
<td>0.11 ± 0.02</td>
<td>1.85x10^-7</td>
<td>5.68x10^-7</td>
</tr>
<tr>
<td>rs7279626</td>
<td>21</td>
<td>45635616</td>
<td>-</td>
<td>-</td>
<td>A/G</td>
<td>42%</td>
<td>0.11 ± 0.02</td>
<td>1.35x10^-7</td>
<td>7.50x10^-7</td>
</tr>
<tr>
<td>rs2838527</td>
<td>21</td>
<td>45633226</td>
<td>-</td>
<td>-</td>
<td>T/C</td>
<td>46%</td>
<td>0.11 ± 0.02</td>
<td>1.32x10^-7</td>
<td>8.92x10^-7</td>
</tr>
<tr>
<td>rs9980139</td>
<td>21</td>
<td>45636731</td>
<td>ICOSLG</td>
<td>downstream</td>
<td>A/G</td>
<td>43%</td>
<td>0.11 ± 0.02</td>
<td>1.77x10^-7</td>
<td>9.15x10^-7</td>
</tr>
<tr>
<td>rs9977351</td>
<td>21</td>
<td>45637406</td>
<td>ICOSLG</td>
<td>3’ UTR</td>
<td>A/G</td>
<td>41%</td>
<td>0.11 ± 0.02</td>
<td>1.66x10^-7</td>
<td>1.12x10^-6</td>
</tr>
<tr>
<td>rs7280525</td>
<td>21</td>
<td>45637392</td>
<td>ICOSLG</td>
<td>3’ UTR</td>
<td>A/G</td>
<td>41%</td>
<td>0.11 ± 0.02</td>
<td>1.66x10^-7</td>
<td>1.12x10^-6</td>
</tr>
<tr>
<td>rs4818890</td>
<td>21</td>
<td>45646256</td>
<td>ICOSLG</td>
<td>downstream</td>
<td>C/A</td>
<td>41%</td>
<td>0.10 ± 0.02</td>
<td>3.01x10^-7</td>
<td>1.42x10^-6</td>
</tr>
<tr>
<td>rs2298565</td>
<td>21</td>
<td>45643396</td>
<td>ICOSLG</td>
<td>intronic</td>
<td>G/C</td>
<td>41%</td>
<td>0.10 ± 0.02</td>
<td>4.73x10^-7</td>
<td>2.20x10^-6</td>
</tr>
<tr>
<td>rs2329714</td>
<td>21</td>
<td>45642155</td>
<td>ICOSLG</td>
<td>intronic</td>
<td>A/T</td>
<td>42%</td>
<td>0.10 ± 0.02</td>
<td>3.56x10^-7</td>
<td>2.30x10^-6</td>
</tr>
</tbody>
</table>

Loci with most significant genome-wide association. Chr, chromosome; MAF, minor allele frequency; UTR, untranslated region.  
\(^{*}\)Genome assembly = GRCh37.  
\(^{¥}\)Refers to effect of the minor allele (transformed scale).  
\(^{1}\)P-value calculated using unadjusted linear regression with an additive mode of inheritance (dosage analysis for imputed SNPs) on transformed variables.  
\(^{2}\)P-value calculated using linear regression with an additive mode of inheritance (dosage analysis for imputed SNPs) on transformed variables adjusted for age, sex, BMI and top three principal components.
Table 5.3: Association between gluten and $\alpha_2$-macrogobulin stratified by rs7278004 genotype.

<table>
<thead>
<tr>
<th>rs7278004</th>
<th>Gluten Intake (g/1000 kcal)</th>
<th>$\alpha_2$-macrogobulin (pmol/μL)</th>
<th>$\beta \pm SE$</th>
<th>$p^1$</th>
<th>$p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC (n=72)</td>
<td>5.46 ± 0.28*</td>
<td>6.54 ± 0.23</td>
<td>0.04 ± 0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>CT (n=195)</td>
<td>5.27 ± 0.18</td>
<td>6.00 ± 0.14</td>
<td>0.008 ± 0.01</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>TT (n=122)</td>
<td>5.60 ± 0.21</td>
<td>5.23 ± 0.21</td>
<td>0.008 ± 0.01</td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>

*Mean ± SE (all such values).

1Test for association between gluten intake and plasma $\alpha_2$-macrogobulin concentrations among SNP genotypes using general linear models adjusted energy intake, age, sex, BMI, physical activity, hormonal contraceptive use among women and fiber intake.

2Test for the interaction between gluten intake and respective SNP genotype on plasma $\alpha_2$-macrogobulin concentration. The interactions and slopes were obtained using a general linear model adjusted for energy intake, age, sex, BMI, physical activity, hormonal contraceptive use among women and fiber intake.
Figure 5.1: Manhattan plot for $\alpha_2$-macrogobulin GWAS. X-axis shows chromosomal position. Y-axis shows the $-\log_{10}$ p-values for the association between each SNP and $\alpha_2$-macrogobulin. Linear regression with additive mode of inheritance adjusted for age, sex, BMI and the top three principal components from population structure analysis used to generate p-values.
Figure 5.2: Quantile-quantile plot for $\alpha_2$-macrogobulin GWAS. X-axis shows expected $-\log_{10} p$-values. Y-axis shows observed $-\log_{10} p$-values generated using linear regression with an additive mode of inheritance adjusted for age, sex, BMI and the top three principal components from population structure analysis. $\lambda_{gc}$, genomic inflation factor.
Figure 5.3: Regional association plot on chromosome 21. X-axis shows chromosomal position. Y-axis (left) shows the -log_{10} p-values for the association between each SNP and α2-macroglobulin. Y-axis (right) shows the recombination rate calculated from 1000 Genomes Project European individuals. SNP colours indicate linkage disequilibrium $r^2$ values with rs7278004. Gene annotations generated using the UCSC genome browser.
5.5 Discussion

Despite the rising popularity of gluten-free foods and the gluten-free diet among the general public, there is currently limited understanding of any physiological effects of gluten intake in individuals without CD (Sapone et al. 2012). Gluten intake has been shown to be positively associated with plasma $\alpha_2$-maroglobulin, a protein involved in various immune-related pathways (see Chapter 4), in those without CD, but it is unclear whether genetic variation modifies this association. The objective of the present study was to identify novel genetic variants associated with circulating levels of $\alpha_2$-maroglobulin which may modify the association between gluten intake and $\alpha_2$-maroglobulin. We identified such variants by employing a two-step procedure which has been previously utilized to identify gene-environment interactions (Gauderman et al. 2013; Kooperberg and Leblanc 2008). First, we conducted a GWAS to identify genetic variants significantly associated with circulating $\alpha_2$-maroglobulin. We then carried over variants with the highest genome-wide significance and determined whether they modified the association between gluten intake and $\alpha_2$-maroglobulin using a candidate gene approach. We identified variation in $ICOSLG$ as being most significantly associated with plasma $\alpha_2$-maroglobulin ($p=1.76 \times 10^{-7}$); however, there was no significant interaction between variation in $ICOSLG$ and gluten intake on circulating levels of $\alpha_2$-maroglobulin ($p>0.05$).

$ICOSLG$ maps to the q arm of chromosome 21 and encodes for the inducible T-cell costimulatory ligand (ICOSL) (NCBI Resource Coordinators 2016). ICOSL is primarily expressed on activated B cells and antigen-presenting cells (APCs) of the immune system (Aicher et al. 2000). When presented on APCs, ICOSL can interact with inducible T-cell costimulator (ICOS) molecules on T-cells, which ultimately leads to T-cell activation,
subsequent proliferation and cytokine release (Hutloff et al. 1999; Dong et al. 2001). Since this T-cell response is an important feature of a large number of immune-related functions, the role of ICOS/ICOSL interactions in the development of various immune responses has been extensively investigated. Animal studies have shown that ICOS signaling is important in both adaptive and innate immune responses (Tanaka et al. 2010; Maeda et al. 2011), as well as the development of autoimmune and allergic reactions (Rottman et al. 2001; Dong et al. 2001; Maeda et al. 2011). Interactions between ICOS and ICOSL are also thought to play an important role in local tissue responses to inflammation (Wang et al. 2012). These functions of ICOS/ICOSL suggest that variation in ICOSLG may be important in the dysregulation of immune function which occurs in a number of immune-related disorders. Furthermore, the observed association between variation in ICOSLG and circulating α2-maroglobulin has strong biological plausibility given the function of α2-maroglobulin in various immune responses, including cytokine release and clearance through receptor-mediated pathways (Mocchegiani et al. 2006; Borth 1994; Rehman, Ahsan, and Khan 2013).

Variation in ICOSLG has been associated with risk of developing gastrointestinal immune diseases, including both inflammatory bowel disease (IBD) and CD. Although the exact mechanisms remain unclear, there is biological plausibility for the causal nature of such associations given the dysregulation of intestinal immune homeostasis observed in the aforementioned disorders and the expression of ICOSLG on intestinal epithelial cells (Nakazawa et al. 2004; Zhang and Li 2014). Various GWAS and meta-GWAS have identified variation near ICOSLG as being associated with risk of developing Crohn’s disease and ulcerative colitis (Barrett et al. 2008; Jostins et al. 2012; Franke et al. 2010). The strongest identified association between IBD and ICOSLG maps to an area downstream of the coding region of the gene, and presence of the risk allele has been shown to be associated with up to an 18% increase in the risk
of developing IBD (Franke et al. 2010). A GWAS in subjects with CD also identified variation in \textit{ICOSLG} as being associated with risk of developing this autoimmune condition (Dubois et al. 2010). Presence of the G allele at rs4819388, an intronic variant in the 3’ UTR of \textit{ICOSLG}, was associated with a 12% increase in the risk of developing CD. Since defined variants in the HLA-DQ region are known to be necessary for the development of CD, rs4819388 only confers additional risk for CD in the presence of HLA risk alleles (Dubois et al. 2010).

The observed association between \(\alpha_2\)-maroglobulin and rs7278004 in the present study suggests that variation in \textit{ICOSLG} may also be implicated in the biological response to gluten intake in those without CD. While both rs7278004 and CD-associated rs4819388 are found in the 3’ UTR of \textit{ICOSLG} (Sherry et al. 2001), both variants are not synonymous with each other (LD \(r^2=0.26\)). Furthermore, the association between CD-associated rs4819388 and \(\alpha_2\)-maroglobulin did not approach genome-wide significance in the present study (\(p=0.04\)), suggesting that associations between \(\alpha_2\)-maroglobulin and this region of \textit{ICOSLG} are driven by rs7278004 and not rs4819388. Although the 3’ UTR of genes is not ultimately translated into the final protein product, untranslated regions are known to serve a number of important regulatory functions, and variants in such regions can have a large influence on gene expression (Mignone et al. 2002; Pesole et al. 2002). This suggests that variants identified in the present study may influence \textit{ICOSLG} expression, and could affect signaling pathways in the immune system and possibly alter immune responses to dietary factors like gluten.

While we observed a main effect for the association between rs7278004 in \textit{ICOSLG} and \(\alpha_2\)-maroglobulin, rs7278004 genotype did not modify the association between gluten intake and \(\alpha_2\)-maroglobulin. There was no significant interaction between rs7278004 and gluten intake on circulating levels of \(\alpha_2\)-maroglobulin (\(p=0.12\)); however, the positive association between gluten
and α2-maroglobulin was strongest in minor allele homozygotes (CC genotype) of rs7278004.

The present study may have been limited in its ability to detect significant gene X diet interactions. Indeed, it has been estimated that sample sizes required to detect interactions are up to four times greater than the sample sizes required to detect main effects of similar magnitude (Smith and Day 1984). It, therefore, remains possible that variation in ICOSLG may modify physiological effects of gluten intake in those without CD, and possibly play a role in the development of adverse responses to gluten, including those observed in NCGS.

The present study has some limitations in addition to the limited power to detect significant gene X diet interactions. The sample size of the GWAS cohort was relatively small, and this may have also limited our power to detect main effects. Furthermore, the study population consisted of a fairly homogenous group of young Caucasian individuals and, therefore, may not be representative of other populations. Additionally, the present study was not sufficiently powered to detect associations at loci with low minor allele frequencies, and all analyses focused on SNPs with a minimum minor allele frequency of 5%. There is growing interest in rare genetic variants (Gorlov et al. 2008), and it is possible that some of the variation in circulating levels of α2-maroglobulin could be attributed to such SNPs. Furthermore, there are likely a number of genetic variants with small effects on circulating α2-maroglobulin levels that, collectively, explain more of the variation in α2-maroglobulin concentration than the variants identified using common genome-wide scanning techniques (Borel 2012).

In summary, we conducted a GWAS of circulating α2-maroglobulin to identify novel genetic variants associated with levels of this plasma protein which may modify the association between gluten intake and α2-maroglobulin in individuals without CD. Genetic variation in the 3’UTR of ICOSLG was most significantly associated with circulating levels of α2-maroglobulin; however,
such variation did not modify the association between gluten intake and α2-maroglobulin. Future studies in those without CD who report adverse responses to gluten are warranted in order to determine whether variation in ICOSLG or other genes are associated with increased risk of developing such conditions.
CHAPTER 6:

GENERAL DISCUSSION
Chapter 6: General Discussion

6.1 Overview and Discussion

The overall aim of this thesis was to investigate both CD and possible physiological effects of gluten intake in those without CD in a Canadian population. Regarding CD, the aim was to determine the prevalence of seropositivity in a population of Canadian adults, and to determine whether the prevalence of positive CD serology, predisposing HLA risk genotypes and average gluten intake differ between major ethnocultural groups. Possible adverse effects of undiagnosed CD were also explored by assessing markers of cardiometabolic health and nutritional status in individuals with likely undiagnosed CD. Associations of gluten intake with plasma proteomic biomarkers and genetic variants in individuals without CD were investigated using omics techniques to elucidate pathways that could potentially become dysregulated in individuals who report adverse responses to gluten.

**Objective 1:** To determine the prevalence of positive CD serology in a population of Canadian adults, and determine whether the prevalence of CD seropositivity, HLA-DQ2/8 high risk alleles and average gluten intake differ between major ethnocultural groups.

**Results:** The prevalence of CD seropositivity in Canadian adults was 1 in 114 (0.88%; 95% CI, 0.57-1.30%). The majority of CD cases in Canada are likely undiagnosed (87%). The prevalence of positive CD serology was highest among Caucasians (1.47%; 95% CI, 0.93-2.22%), while no cases were identified among those of East Asian or South Asian descent. East Asians had lower
levels of both gluten intake and high risk HLA genotypes than Caucasians and South Asians (p<0.05).

**Objective 2:** To determine whether biomarkers of cardiometabolic health and nutritional status differ between Canadian adults with positive and negative CD serology.

**Results:** Undiagnosed CD was associated with lower levels of HDL-cholesterol (p=0.008), lower levels of apolipoprotein-AI (p=0.02), a higher ratio of total cholesterol to HDL cholesterol (p=0.006) and a higher apolipoprotein-B/AI ratio (p=0.03). Those with likely undiagnosed CD also had lower circulating levels of retinol (p=0.006) and a trend for lower vitamin D (p=0.08), but had similar levels of various carotenoids, tocopherols and ascorbic acid compared to individuals without CD.

**Objective 3:** To examine the association between gluten intake and plasma proteomic biomarkers in a population of young adults without CD.

**Results:** Intake of gluten was associated with increased concentrations of plasma α₂-macroglobulin, a proteomic marker of cytokine release, in individuals without CD in fully adjusted models accounting for energy intake, age, sex, BMI, fiber intake, physical activity and hormonal contraceptive use among women (p=0.004). Gluten intake was not significantly associated with any of the other 53 plasma proteins analyzed.
**Objective 4:** To identify genetic variants that are associated with plasma proteomic biomarkers of gluten intake using a genome-wide approach and determine whether any such variants modify the association between gluten intake and plasma protein concentration.

**Results:** Plasma $\alpha_2$-macroglobulin was most significantly associated with genetic variation (rs7278004) in *ICOSLG* ($p=1.67 \times 10^{-7}$), a T-cell stimulatory gene involved in cytokine release and various gastrointestinal immune-related disorders. There was no significant interaction ($p=0.12$) between variation in *ICOSLG* and gluten intake on plasma $\alpha_2$-macroglobulin; however, the positive association between gluten intake and $\alpha_2$-macroglobulin was strongest in minor allele homozygotes at rs7278004.

There has been increased awareness for gluten-related disorders over the past decade as evidenced by the increasing popularity of gluten-free foods (Sapone et al. 2012). Screening-based studies have shown that approximately 1% of many populations around the world are affected by CD, a lifelong autoimmune disorder of the intestinal mucosa triggered by gluten intake (Gujral, Freeman, and Thomson 2012; Kagnoff 2007). Non-celiac gluten sensitivity (NCGS) is thought to exist, but the pathophysiological mechanisms and prevalence of such a condition remain unclear (Fasano et al. 2015). The present thesis confirms that the prevalence of CD in North America is just below 1%, with the vast majority of cases being undiagnosed and of Caucasian ancestry. The present thesis also demonstrates that those with likely undiagnosed CD have less favourable lipid profiles and lower circulating levels of certain fat-soluble vitamins. In individuals without CD, gluten intake was found to be positively associated with plasma $\alpha_2$-macroglobulin, a marker of cytokine release. Plasma concentrations of $\alpha_2$-macroglobulin were associated with variation in *ICOSLG*, a gene involved in T-cell signaling and various
gastrointestinal immune-related disorders. While the interaction between variation in ICOSLG and gluten intake on α2-macroglobulin was not significant, the positive association between gluten and α2-macroglobulin was driven by individuals with a specific variant in ICOSLG.

The first two experimental chapters in the present thesis (Chapters 2 and 3) investigated CD, predisposing HLA-DQ2/DQ8 genotypes and gluten intake in an ethnoculturally diverse population of adults in Canada. The prevalence of CD has been shown to exhibit geographical variation (Gujral, Freeman, and Thomson 2012). In the United States, approximately 0.75% of individuals are thought to be affected by CD (Fasano et al. 2003). In Europe, although there can be considerable variation between countries, most screening studies demonstrate that ~1% of individuals likely have CD (Kang et al. 2013). In China and Japan, CD is thought to be extremely rare, with only a limited number of case reports appearing in the literature (Cummins and Roberts-Thomson 2009; Kang et al. 2013); however, accurate estimates on the true prevalence of CD from antibody screening-based studies are not available for East Asia. CD is now recognized as being relatively common in India, with prevalence estimates ranging from 0.1% in Southern India to >1% in Northern India (Ramakrishna et al. 2016). Our findings demonstrate that the prevalence of CD seropositivity among Canadian adults of diverse ethnocultural backgrounds is consistent with estimates from the U.S and many European countries. Furthermore, similar to screening studies in the U.S., we have shown that Caucasians comprise the majority of CD cases (Mardini, Westgate, and Grigorian 2015), and that CD is likely underdiagnosed despite increased awareness for this condition (Guandalini and Assiri 2014; Rubio-Tapia et al. 2012).

While the prevalence of certain autoimmune conditions has been shown to vary with latitude, with estimates being higher in northern climates (Ponsonby, McMichael, and van der Mei 2002),
our results regarding the prevalence of CD in a Canadian population are consistent with worldwide estimates and, therefore, do not support the notion that such a phenomenon occurs with CD. The proposed mechanism of latitudinal variation in the prevalence of certain autoimmune disease is thought to be partially attributed to the protective immunosuppressant effects of vitamin D, which is produced upon cutaneous exposure to ultraviolet B radiation (Ponsonby, McMichael, and van der Mei 2002; Arnson, Amital, and Shoenfeld 2007). Indeed, vitamin D deficiency has been shown to be associated with an increased risk of developing autoimmune diseases such as multiple sclerosis, rheumatoid arthritis and type I diabetes (Holick 2004). While CD can be associated with decreased absorption of fat-soluble vitamins like vitamin D due to flattening of the mucosal surface (Lerner et al. 2012), vitamin D deficiency has traditionally not been thought of as a risk factor for CD. Nevertheless, recent evidence has indicated that children with CD are more likely to be born in the summer (Capriati et al. 2015; Tanpowpong et al. 2013; Ivarsson et al. 2003; Lebwohl, Green, et al. 2013; Lewy, Meirson, and Laron 2009), suggesting a possible interaction between lower vitamin D levels during the winter months and timing of first gluten exposure may be important in the development of CD.

The ‘hygiene hypothesis’ offers an alternative explanation for observed differences in the geographical prevalence of autoimmune disorders. According to the ‘hygiene hypothesis’, countries that are increasingly industrialized have a lower infectious burden than developing countries, and this consequently leads to an increase in the prevalence of both autoimmune and allergic disorders since exposure to certain infectious agents may be protective against such immune-related disorders (Strachan 1989; Okada et al. 2010; Bach 2002). Such a hypothesis would explain the increasing prevalence of various autoimmune diseases (Rook 2012), including CD (Lebwohl, Ludvigsson, and Green 2015), over the past century. However, many aspects of the ‘hygiene hypothesis’ have proven to be inherently difficult to assess, and the extent to which
exposure to various infectious agents protects against the development of specific allergic and autoimmune reactions remains unknown. Furthermore, contrary to the predictions of the ‘hygiene hypothesis’, exposure to certain viral infections early in life has actually been shown to be positively associated with the development of CD (Stene et al. 2006). Additionally, the highest recorded prevalence of CD (>5%) has been observed in Saharawi refugee children living in Africa (Catassi et al. 1999). This represents a 5- to 10-fold higher prevalence of CD compared to developed countries, and suggests that the ‘hygiene hypothesis’ alone likely does not explain temporal or geographical differences in the prevalence of CD.

On the population level, the prevalence of predisposing risk variants in the HLA-DQ gene locus as well as average gluten consumption are established predictors of CD prevalence (Gujral, Freeman, and Thomson 2012). Indeed, Western countries with a high prevalence of HLA-DQ2 and DQ8 risk alleles and high per capita wheat consumption (eg. U.S. and Europe) have a distinctly higher prevalence of CD than countries like China and Japan, where predisposing HLA-DQ2 and DQ8 risk variants are rare and rice, rather than wheat, is a dietary staple (Cummins and Roberts-Thomson 2009; Gujral, Freeman, and Thomson 2012). Our results suggest that differences in the prevalence of CD in certain immigrant populations of varying ethnicities living in Canada may also be partially explained by differences in the prevalence of HLA-DQ2/DQ8 risk variants and average gluten intake. Nevertheless, it is becoming increasingly clear that there are a number of additional factors that influence the population prevalence of CD. Two such factors include non-HLA genetic variation as well as alterations in the composition of the gut microbiota (Dieli-Crimi, Cenit, and Nunez 2015; Marasco et al. 2016), with the latter suggesting an important role of global dietary patterns, and not necessarily just gluten intake.
The prevalence of CD has been shown to be increasing over time (Rubio-Tapia, Kyle, et al. 2009; Lohi et al. 2007), although the environmental factors responsible for this trend remain unclear. While increasing awareness for CD has likely resulted in a higher rate of diagnosis, this alone does not explain the increasing prevalence of CD as serologic screening-based studies have shown that undiagnosed cases of CD have also been on the rise over the past 50 years (Murray et al. 2003; Rubio-Tapia, Kyle, et al. 2009). Some studies have suggested that undiagnosed CD may be associated with a number of adverse health effects, including an increased rate of overall mortality (Rubio-Tapia, Kyle, et al. 2009; Metzger et al. 2006); however, not all studies have supported this association (Canavan et al. 2011; Lohi et al. 2009; Godfrey et al. 2010).

Furthermore, follow-up studies of individuals with undiagnosed CD detected in screening studies have found that the majority of such individuals refuse further testing for CD and do not go on to adopt a gluten-free diet (Shamir et al. 2007). This has led to considerable debate in the scientific community as to whether routine CD screening should be pursued in the general population (Aggarwal, Lebwohl, and Green 2012). Our findings indicate that those with undiagnosed CD may have less favourable lipid profiles and lower levels of certain circulating fat-soluble vitamins, but not widespread nutrient deficiencies indicative of severe villous atrophy. This suggests that most individuals with undiagnosed CD identified through screening studies may fall into the mild end of the CD spectrum. Whether such individuals are at risk for developing more severe forms of CD and any associated complications later in life remains unclear.

The last two objectives of the present thesis (Chapters 4 and 5) aimed to investigate the associations of gluten intake with plasma proteomic biomarkers and genetic variants in those without CD using a multi-omics approach. While there has been a recent increase in the popularity of gluten-free foods among those in the general population (Sapone et al. 2012), there is limited understanding of the physiological effects of gluten in individuals without CD. Non-
celiac gluten sensitivity (NCGS) is thought to affect an additional subset of the population, but may aspects of the pathogenesis of this condition remain unknown (Fasano et al. 2015). Our findings demonstrate that gluten intake may be associated with increased circulating concentrations of α2-macroglobulin in individuals without CD. This protein is a marker of cytokine release, and it has been shown to bind to a number of pro-inflammatory (TNF-α, IL-6 and IL-8) and anti-inflammatory cytokines (IL-10 and IL-10) in order to facilitate their receptor-mediated clearance from the systemic circulation (Rehman, Ahsan, and Khan 2013; Garber, Gonias, and Webb 2000). It is possible that such pathways could become dysregulated in individuals with NCGS. An upregulation of such pathways in individuals with NCGS could potentially be indicative of enhanced secretion of pro-inflammatory cytokines. Alternatively, gluten has been shown to lead to increased secretion of IL-10, an anti-inflammatory cytokine and important regulator of innate immunity, in individuals without CD, but not in those with NCGS (Hollon et al. 2015). Since α2-macroglobulin is known to bind IL-10 and facilitate its removal from the circulation (Garber, Gonias, and Webb 2000), a downregulation of α2-macroglobulin in those with NCGS could suggest an important role for the dysregulation of IL-10 secretion in those without CD who report adverse symptoms to gluten.

Variation in ICOSLG, a gene involved in T-cell proliferation and cytokine release, was associated with circulating levels of α2-macroglobulin in individuals without CD. Although the association between gluten intake and α2-macroglobulin was not modified by variation in ICOSLG, the positive association between gluten and α2-macroglobulin was strongest in individuals possessing a certain ICOSLG genotype. Variation in ICOSLG has previously been associated with the development of gastrointestinal immune-related disorders, including IBD and CD (Barrett et al. 2008; Franke et al. 2010; Jostins et al. 2012; Dubois et al. 2010). Our results suggest that such variation may also play a role in the physiological effects of gluten intake in
those without CD, and potentially be involved in the development of adverse symptoms to gluten intake experienced by those with possible NCGS. Clearly, further research into the physiological effects of gluten intake in both healthy individuals without CD and those with potential NCGS is necessary to better understand the mechanisms which ultimately lead to the development of symptoms in affected individuals.

6.2 Limitations

The work presented in this thesis has some potential limitations.

6.2.1 Celiac Disease Assessment

The assessment of CD using a combination of genetic and serologic screening precludes a conclusive diagnosis in all cases identified in the present thesis. While the assessment of HLA-DQ2/DQ8 genotypes as well as tTG autoantibodies is a highly sensitive and specific method for CD screening, additional collection of duodenal biopsy samples and the demonstration of gluten-specific damage to the mucosa remains the gold standard in diagnosing CD (Barakauskas, Lam, and Estey 2014). Indeed, there is the potential for false positive findings when relying solely on serology for diagnosis. However, our method of CD assessment relied on positive tTG serology on two separate ELISA kits. Furthermore, individuals were initially screened using genetic testing, and only individuals with the HLA-DQ2/8 genotypes necessary for the development of CD underwent tTG serologic screening. Restricting serologic testing to such individuals has been shown to decrease the number of false positive findings and increase the specificity of serologic
tests (Husby et al. 2012; Barakauskas, Lam, and Estey 2014). Alternatively, relying solely on serologic screening can result in false negative test results as a small subset of individuals with CD are known to present with negative serology despite being on a gluten-containing diet (Ierardi et al. 2015). Individuals with CD could also present with negative serology if gluten is eliminated from their diet before serologic testing. Despite these limitations, the assessment of tTG serology is considered a highly valuable tool in assessing the prevalence of CD on the population level, and our findings are consistent with other screening-based studies conducted in North America which utilize similar screening approaches (Rubio-Tapia, Kyle, et al. 2009; Rubio-Tapia et al. 2012; Fasano et al. 2003; Mardini, Westgate, and Grigorian 2015; Kim et al. 2016).

The lack of biopsy samples and data on CD-specific symptoms also limits the conclusions which can be drawn regarding the severity of CD in identified cases of positive serology. Indeed, CD is considered a spectrum based on varying degrees of symptoms and histological findings, with individuals falling into the broad categories of ‘typical’, ‘atypical’, ‘silent’, ‘potential’ or ‘refractory’ CD (see Chapter 1.3.5). While estimates of the prevalence of CD include all aforementioned sub-categories, knowledge of the distribution of cases with regards to this spectrum has many important implications. Different categories of CD may be associated with variable long-term risks of developing CD-associated complications (see Chapter 1.3.6). This is important when considering the usefulness and feasibility of population screening for CD in populations not considered to be at risk. While biopsy samples and data on CD-specific symptoms were not available in the present study populations, other lines of data allowed for inferences about the severity of CD in likely undiagnosed cases. In Chapter 3, data on serum biomarkers of cardiometabolic health and nutritional status were compared between individuals with positive and negative CD serology. This enabled a better understanding of the degree of
mucosal damage that may be present in individuals with likely undiagnosed CD identified in screening-based studies.

6.2.2 Biochemical Measurements

Biochemical measurements in the present thesis, including markers of cardiometabolic health (Chapter 3), nutritional status (Chapter 3) and a panel of plasma proteins (Chapter 4/5), were all measured at a single time point. This represents a potential limitation since random within-subject variability in day-to-day biomarker assessment can attenuate associations between such biomarkers and health outcomes. Nevertheless, a single measurement of most of the utilized nutrition-related biomarkers has been estimated to produce correlation coefficients within 80-88% of true correlations between such biomarkers and health outcomes (Block et al. 2006). Furthermore, the majority of the aforementioned markers of nutritional status were measured using HPLC-based methods which have been extensively validated and represents the gold standard for assessing such markers (Jansen and Ruskovska 2015; Wootton 2005).

The simultaneous measurement of 54 plasma proteins using an MRM assay (see Chapter 4.3.3) has some unique limitations. Incomplete digestion of plasma proteins could lead to a lower concentration of peptide fragments which are ultimately measured as a proxy for protein concentrations. However, all peptides measured had CVs <15%, indicating that digestion rates likely approached 100% and that variability in digestion rates between samples were minimal. Additionally, post-translational modifications in peptide sequences could negatively affect the accuracy of the MRM assay. However, peptide sequences with no known post-translational modification sites were selected to minimize this possibility (Kuzyk et al. 2009). Finally, while
the plasma proteins assessed are traditionally measured using immunoassay methods (eg. ELISA), the values obtained using the MRM assay were reproducible and comparable to previously published values obtained using immunoassay methods (Hortin, Sviridov, and Anderson 2008; Farrah et al. 2011; Garcia-Bailo et al. 2012).

6.2.3 Dietary Assessment

The assessment of dietary intakes represents a limitation of all nutritional epidemiology studies. Prominent methods used to assess habitual food consumption patterns include 24hr recalls, food records and food frequency questionnaires (FFQs) (Shim, Oh, and Kim 2014). With FFQs, subjects indicate the frequency at which they typically consume foods from a pre-selected list within a pre-defined time (ie. week, month, or year). While repeated 24hr recalls and food records over multiple days are required to achieve an accurate measure of true dietary intake over a period of time, FFQs are typically administered at a single time point. Although FFQs are associated with lower participant burden than other forms of dietary assessment, they are limited by the ability of participants to accurately recall the foods they have consumed over a relatively long period of time. FFQs are also limited by the finite number of pre-specified foods that have been listed, which may not be exhaustive for all subjects. This can be particularly problematic when assessing dietary intakes in individuals of differing ethnocultural backgrounds, as foods which are popular in certain cultures may be absent from FFQ food lists. However, compared to other FFQs, the Toronto-modified Willett questionnaire used in the present thesis contains a relatively large number of food items (n=196). Furthermore, the questionnaire probes for dietary intakes over the past month, so it may be subject to a lesser degree of recall bias than FFQs
which assess intakes over longer timeframes. The base Willett FFQ has been extensively validated in North American populations (Subar et al. 2001; Holmes et al. 2007), and has been shown to provide reliable estimates of intake for many nutrients (Cahill, Corey, and El-Sohemy 2009; Brenner et al. 2011), including carbohydrates (Eny et al. 2008). Finally, while the utilized FFQ was not designed or validated for gluten, a previously established method for estimating gluten intake from questionnaire-based studies was utilized (Hopman et al. 2007; Hopman et al. 2012; van Overbeek et al. 1997), and estimates obtained for total gluten intake were similar to estimates previously reported in other Western populations (van Overbeek et al. 1997).

6.2.4 Nature of Study Populations

The nature of the study populations utilized in the present thesis pose some potential limitations on the conclusions that can be drawn from the results. While the inclusion of major ethnocultural groups likely strengthens the external generalizability of the findings presented, subjects from both populations were recruited exclusively from sites in Toronto, and therefore may not be representative of all Canadian adults. According to Canadian census reports, approximately 8% of the Canadian population is comprised of East Asians, while 5% of Canadians are South Asians (Statistics Canada 2011). This is compared to the present study populations, where 22% and 8% of individuals were East Asian and South Asian, respectively. Furthermore, both study populations assessed were limited to adults over the age of 18. While there is considerable interest in gluten-related disorders in adults, the findings presented in this thesis may not be applicable to pediatric populations. Indeed, childhood CD is characterized by a different clinical presentation than adult-onset CD, and often involves more severe gastrointestinal symptoms.
Finally, there are some limitations associated with the cross-sectional nature of the TNH Study and cross-sectional analysis of the THD Study. Although cross-sectional studies are well suited to define the prevalence of conditions like CD at specific points in time, this study design precludes the investigation of changes in prevalence over time, the follow-up of individuals with undiagnosed disease, as well as the identification of exposures that are causally linked to the development of such conditions. Likewise, although a number of potential confounders were accounted for when assessing the association between gluten intake and various proteomic biomarkers in those without CD, causal inferences in any of the observed associations cannot be made due to the possibility of unaccounted for residual confounding.

6.3 Future Research

6.3.1 Celiac Disease

Future research into CD in a Canadian context should aim to build upon the findings of the present thesis. To improve estimates of the prevalence of CD across Canada, results of screening studies conducted in all Provinces and Territories should be considered. The Canadian Health Measures Survey (CHMS) collects important health information from households across Canada on a biannual basis (Giroux 2007), and represents an ideal platform for assessing the prevalence of CD across Canada. As part of the CHMS examination, blood samples are collected for the assessment of chronic and infectious disease, as well as for markers of nutrition and environmental exposures. Such samples could undergo screening for predisposing HLA-DQ2/DQ8 variants as well as tTG antibodies. Such screening studies for CD have already been
conducted in multiple iterations the National Health and Nutrition Examination Survey (NHANES), a nationally representative survey in the U.S. which is analogous to the CHMS in Canada (Mardini, Westgate, and Grigorian 2015; Kim et al. 2016). Since the CHMS is collected biannually, routine screening for CD in this study population would enable to assessment of changes in the prevalence of CD over time. With data on biomarkers of nutritional status and environmental exposures also available in CHMS, routine CD screening could ultimately facilitate an improved understanding of factors that are associated with the observed rise in the prevalence of CD in North America. Such screening studies in CHMS or other populations could also collect biopsy samples form a random subset of individuals with positive tTG serology to assess the severity of mucosal damage in these individuals. This would improve our understanding of the distribution of cases within the spectrum of CD, and enable further research into the factors that may be associated with the development of more severe, symptomatic forms of CD compared to relatively asymptomatic presentations. Follow-up of individuals identified as having previously undiagnosed CD in longitudinal studies would also be invaluable in assessing whether those with untreated CD are at risk for adverse health complications later in life, and whether adherence to a gluten-free diet attenuates this risk.

Future research should also aim to investigate CD in pediatric populations in Canada. Birth cohort studies in which infants are recruited at birth and followed for a number of years represent an ideal study design for assessing the development of CD in young children. Given the large cost associated with longitudinal studies of infants and the relatively low prevalence of CD (~1%), such studies could be restricted to infants with the HLA-DQ2/8 genotypes which are necessary for the development of CD. Such studies have already been conducted in the U.S. and Europe (Norris et al. 2005; Lionetti et al. 2014; Akobeng et al. 2006), and some have demonstrated possible influences of season of birth and viral infections on risk of developing CD.
(Capriati et al. 2015; Tanpowpong et al. 2013; Ivarsson et al. 2003; Lebwohl, Green, et al. 2013; Lewy, Meirson, and Laron 2009). Longitudinal birth cohorts in Canada represent a unique opportunity to replicate and build on these findings, as true effects could be expected to be even more pronounced given the relatively large seasonal fluctuations in Canada. Furthermore, previous birth cohort studies have yielded inconsistent findings regarding the influence of breastfeeding habits and the first introduction of gluten on risk of developing CD (Norris et al. 2005; Lionetti et al. 2014; Vriezinga et al. 2014), and birth cohort studies in Canada could help clarify such inconsistencies. Findings from such studies could have broad public health implications for the prevention of CD in at-risk children.

6.3.2 Non-Celiac Gluten Sensitivity

Future research on NCGS should initially aim to determine whether patients with apparent gluten-related symptoms in which CD has been conclusively ruled out are, in fact, experiencing symptoms in response to gluten and not another component of the diet. Further research should then aim to improve upon the diagnostic criteria for this condition. The identification of sensitive and specific biomarkers for diagnosing NCGS is important, as this would allow practitioners to make definitive diagnoses using objective measures. It would also facilitate research to better understand the pathogenesis and true prevalence of such a condition. Randomized controlled trials in which gluten is administered to patients with potential NCGS in a double blinded fashion and paired with the assessment of a panel of biomarkers of immune and intestinal barrier function are well suited to demonstrate the presence of gluten-restricted effects and identify potential diagnostic biomarkers of NCGS. While studies conducted to date have demonstrated
some evidence for gluten-specific effects and identified certain markers which differ between those with potential NCGS and controls (Brottveit et al. 2013; Sapone et al. 2011; Sapone et al. 2010; Hollon et al. 2015; Uhde et al. 2016), none of these markers are sensitive or specific enough to be utilized in the diagnosis of NCGS. Building on the findings of the present thesis, future RCTs in individuals with potential NCGS should investigate whether gluten intake is associated with changes in α2-macroglobulin or other immune-related proteomics markers in such individuals. Since highly abundant plasma proteins are ubiquitously expressed, it is unlikely that these proteins will serve as diagnostic biomarkers of NCGS. Rather, they may be useful in elucidating the pathophysiological mechanisms behind such a condition. An improved understanding of these mechanisms may ultimately assist with the identification of diagnostic biomarkers for NCGS.

The identification of diagnostic biomarkers for NCGS could facilitate further research into the possible genetic basis of such a condition. This could be achieved through both candidate gene and genome-wide studies. Candidate genes could be selected on the basis of known biological associations with pathophysiological pathways implicated in the development of NCGS. While such pathways currently remain poorly understood, it is possible that genes involved in intestinal barrier function as well as innate immunity could emerge as promising candidates (Lebwohl, Ludvigsson, and Green 2015; Uhde et al. 2016). Clear diagnostic criteria could also enable the recruitment of NCGS-restricted populations with sample sizes large enough to conduct a case-control GWAS. Variants that reach genome-wide significance in such a study could ultimately prove to be predisposing risk factors for the development of gluten sensitivity in the absence of CD. Further research would be necessary to determine whether such genetic variation interacts with environmental factors to influence risk of developing gluten sensitivity as is the case with variation in HLA and CD.
6.4 Implications

Findings from the present thesis have a number of implications. Results of the first screening-based study for CD conducted in a Canadian population (Chapter 2) demonstrate that approximately 1% of individuals are likely affected by this condition, with the vast majority of cases likely being undiagnosed. This will result in increased awareness for CD among both physicians and the general population. For physicians, these results highlight the under-diagnosis of CD and suggest the need for increased screening among those considered to be at risk for CD. This includes individuals presenting with gastrointestinal and extra-gastrointestinal symptoms consistent with CD, those with a family history of CD, type I diabetes or other autoimmune disorders, Down or Turner syndrome and, if known, those possessing high-risk HLA-DQ2/8 genotypes (Fasano et al. 2003; Ludvigsson et al. 2015). For the general population, increased awareness for undiagnosed CD can encourage those experiencing unexplained symptoms of varying severities to seek further testing for CD. This is likely to work synergistically with the increase in popularity of many consumer genetic tests which provide individuals with information regarding their genetic risk for CD (Marietta and McGuire 2009). Indeed, increased awareness for CD may lead to a better understanding of the results of genetic testing for CD, and could facilitate a higher rate of diagnosis among those experiencing symptoms consistent with CD with knowledge of their HLA risk profile. This is important since increased diagnosis of CD would be expected to lead to a significant reduction in health care costs over time (Green et al. 2008).
Results from the present thesis also have important implications for public health policy. Currently, there is debate within the scientific community as to whether mass screening for CD in the general population is warranted (Evans, McAllister, and Sanders 2009; Fasano 2009; Aggarwal, Lebwohl, and Green 2012; Ludvigsson et al. 2015). According to guidelines set by the World Health Organization, a disease must fulfill each of the following criteria in order for mass screening to be considered appropriate (Wilson and Jungner 1968): (1) early detection of the disorder must be difficult; (2) the condition must be common; (3) screening tests must be highly sensitive and specific; (4) screening tests must be culturally acceptable; (5) effective treatment must be available; (6) untreated disease can lead to adverse health complications, and (7) testing and treatment must be cost-effective. It is clear that CD satisfies criteria 1-5, as evidenced by the high prevalence of undiagnosed CD, the overall prevalence of ~1% in North America, the accuracy of serologic tests (see Chapter 1.3.3), the widespread acceptance of serologic testing, and effectiveness of a gluten-free diet, respectively. However, the extent of adverse health effects associated with untreated CD, particularly among those who remain undiagnosed, and the resulting cost-effectiveness of screening/treatment remains unclear given the wide and variable clinical spectrum of CD (Ludvigsson et al. 2015). Results of the present thesis suggest that individuals with untreated CD may have less favourable lipid profiles and lower circulating concentrations of certain fat-soluble vitamins. However, such individuals may not have severe and widespread nutrient deficiencies, suggesting that the majority of individuals with undiagnosed CD identified through population screening likely fall under the mild end of the CD spectrum. This is an important consideration that could improve the accuracy of models designed to predict the cost-effectiveness of mass screening for CD (Shamir, Hernell, and Leshno 2006; Dorn and Matchar 2008), and ultimately inform public policy on the usefulness of population screening.
Regarding gluten sensitivity beyond the autoimmune mechanisms of CD, results of this thesis demonstrate possible physiological effects of gluten intake in those without CD. This has implications for future research investigating the possible pathophysiological mechanisms by which gluten may lead to adverse symptoms in individuals with potential NCGS. Recent research has indicated that a subset of individuals without CD may indeed experience adverse symptoms in response to gluten intake (Biesiekierski et al. 2011; Di Sabatino et al. 2015; Zanini et al. 2015; Elli et al. 2016), but the mechanisms underlying the development of these symptoms remain poorly understood. Plasma proteomics markers of immune function represent candidate markers of pathways associated with gluten intake that may become dysregulated in individuals with NCGS. Further investigation of such markers and a number of other potential candidates in individuals with NCGS may ultimately lead to a clearer understanding of the pathogenesis of such a condition. Furthermore, the identification of genetic variants that modify the physiological effects of gluten intake in those without CD can facilitate research into the potential genetic basis of NCGS. Our results suggest that genes involved in immune system cell signaling, such as ICOSLG, represent ideal candidates for gene regions that warrant further investigation in individuals with possible NCGS.

6.5 Summary of Thesis

In conclusion, the present thesis investigated CD and the effects of gluten intake in those without CD in a population of Canadian adults. The results indicate that the prevalence of CD autoimmunity in Canada is 0.88%, with the vast majority of likely CD cases being Caucasian and undiagnosed. Undiagnosed CD was associated with decreased levels of HDL-cholesterol,
apo-A2, and certain fat-soluble vitamins, but not widespread nutrient deficiencies. In individuals without CD, gluten intake was shown to be associated with a plasma $\alpha_2$-macroglobulin, a proteomic biomarker with various immune-related functions. The association between gluten and $\alpha_2$-macroglobulin was strongest in individuals with a specific ICOSLG gene variant. Overall, the research presented in this thesis has defined the prevalence of CD autoimmunity in Canada and the degree of adverse health effects in those with likely undiagnosed CD, and also shed light on possible physiological effects of gluten intake in those without CD.
References


Dickey, W. 2002. 'Low serum vitamin B12 is common in coeliac disease and is not due to autoimmune gastritis', *Eur J Gastroenterol Hepatol*, 14: 425-7.


Symptoms: Results from a Multicenter Randomized Double-Blind Placebo-Controlled Gluten Challenge', *Nutrients*, 8.


Food and Drug Administration. 2013. 'Food labeling; gluten-free labeling of foods', *Federal Register*, 78: 47154-79.


analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci', *Nat Genet*, 42: 1118-25.


Gilbert, C. 2013. 'What is vitamin A and why do we need it?', Community Eye Health, 26: 65.


Han, Y., W. Chen, P. Li, and J. Ye. 2015. 'Association Between Coeliac Disease and Risk of Any Malignancy and Gastrointestinal Malignancy: A Meta-Analysis', Medicine (Baltimore), 94: e1612.


disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease', *Hum Immunol*, 64: 469-77.


Tucker, E., K. Rostami, S. Prabhakaran, and D. Al Dulaimi. 2012. 'Patients with celiac disease are increasingly overweight or obese on presentation', J Gastrointestin Liver Dis, 21: 11-5.


symptom recurrence in only a minority of patients who meet clinical criteria for non-coeliac gluten sensitivity', *Aliment Pharmacol Ther*, 42: 968-76.


## Appendices

**Appendix A1**: Gluten content of gluten-rich food items listed in Toronto-modified Willett FFQ\(^1\).

<table>
<thead>
<tr>
<th>Gluten-Rich Food Category</th>
<th>Food Item (Serving Size)</th>
<th>Gluten (g) per Serving</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bread</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pizza (150g)</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>Bagel (98g) or Roll (28g)(^2)</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>English Muffins (52g), Bagels (98g) or Rolls (28g)</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>Dark Bread (25g), including Pita (45g)(^2)</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Pancakes, Waffles or French Toast (100g)</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>White Bread (25g), including Pita (45g)</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Pretzels (28g)</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Heavy Bread (25g)</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Crackers (15g)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Pasta</strong></td>
<td>Macaroni &amp; Cheese (140g)</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>Whole Grain Pasta (140g)</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Pasta (140g)</td>
<td>5.8</td>
</tr>
<tr>
<td><strong>Cereal</strong></td>
<td>Cold Breakfast Cereal (30g)</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Barley (157g)</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Baked Goods</strong></td>
<td>Pie, Home-Baked (130g)</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Pie, Ready-Made (130g)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Doughnuts (45g)</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Brownies (40g)</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Cake, Home-Baked (50g)</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Wheat Germ (7g)</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Sweet Roll or Coffee Cake, Ready Made (30g)</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Cake, Ready-Made (50g)</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Cookies, Home-Baked (20g)</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Sweet Roll or Coffee Cake, Home Baked (30g)</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Cookies, Fat Free (20g)</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Cookies, Ready-Made (20g)</td>
<td>0.4</td>
</tr>
</tbody>
</table>

\(^1\)USDA National Nutrient Database for Standard Reference and Harvard FFQ Database used to determine serving size and vegetable protein content for each food item. Intermediate serving sizes were selected when multiple options were present. Vegetable protein content was multiplied by 0.8 to obtain the approximate amount of gluten. For food categories that contained multiple food items (eg. ‘English muffins, bagels or rolls’), the approximate gluten content of all sub-items was calculated and divided by the total number of sub-items.

\(^2\)Made with whole wheat or whole grain flour
Appendix A2: Population structure analyses for Toronto Nutrigenomics and Health GWAS Cohort.

Figure A2.1: Principal components analysis for TNH GWAS Cohort seeded with individuals of various ethnicities from the 1000 Genomes Project.
Figure A2.2: Principal components analysis for TNH GWAS Cohort seeded with European individuals from the 1000 Genomes Project. CEU, Utah Residents with Northern and Western Ancestry (Caucasians); FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian Population in Spain; TNH, Toronto Nutrigenomics and Health GWAS Cohort; TSI, Toscani in Italy