

# Association of Clinical Phenotype and Genetic Burden in Coeliac Disease

Patricia Dominguez Castro



Trinity  
College  
Dublin

The University of Dublin

A thesis submitted to the University of Dublin for the  
degree of Doctor of Philosophy

Supervisor: Prof Ross McManus

Department of Clinical Medicine

University of Dublin

Trinity College

2017

## **Declaration**

I hereby declare that this thesis has not been previously submitted for a degree at this or any other university and that the work described in this thesis is entirely my own work. I hereby give my permission to the library or Trinity College Dublin to lend or copy my thesis upon request for the purpose of study.

---

Patricia Dominguez Castro

## Summary

*Coeliac disease (CD) is a chronic small intestinal immune-mediated enteropathy precipitated by exposure to a gluten containing diet in genetically predisposed individuals. The clinical presentation of CD is highly heterogeneous, ranging from severe intestinal symptoms such as diarrhoea and weight loss to subclinical or silent forms of the disease. Many studies have reported a change in the clinical phenotypic presentation of CD in recent years, with an increase in the prevalence of a non-classical and subclinical phenotype at diagnosis as well as the median age of diagnosis. The co-existence of CD with other immune-mediated conditions (IMC) has been well reported in the literature. It is recommended that patients with type 1 diabetes mellitus (T1DM) and autoimmune thyroid disease (ATD) are screened for CD. There is a paucity of data regarding the clinical presentation of CD in the Republic of Ireland (ROI).*

*Genome wide association studies (GWAS) have helped elucidate the polygene architecture of CD, with HLA and non-HLA genes known to contribute to the development of the condition. Little is currently known as to the effects of genotype on clinical phenotype in CD. The use of polygene risk scores (PRS) in CD, combining HLA and non-HLA predisposing loci, have been shown to improve the identification of potential CD patients. However, whilst correlation of PRS with overall disease risk has been well explored, the association of the genotype with CD clinical phenotype has not been extensively investigated. Therefore, studies combining clinical phenotypes with genetic information could potentially explain differences in disease presentation and development.*

*This thesis aimed to provide an updated description of the clinical phenotype of CD in Ireland at present, and to report any potential changes in the presentation of the condition over time. Furthermore, we wanted to explore the ability of PRS to predict differences in disease presentation and prognosis.*

*Our results have corroborated findings reported internationally, that the clinical profile of the disease has changed radically over the past 55 years. Among the changes we found were an increased median age at diagnosis in adulthood, a higher prevalence of non-classical CD at diagnosis, an equal prevalence of patients diagnosed in childhood (<18 years) and patients diagnosed over 65 years and a large percentage of patients presenting either overweight or obese at diagnosis. Also less associated ATD and dermatitis herpetiformis (DH) are observed over time. It seems plausible that these changes are due to improved detection of non-classical disease, and/or potentially*

*shifts in environmental exposures which are engendering changes in the clinical manifestation of CD.*

*Our results also support previous findings that CD patients have an increased occurrence of immune mediated conditions IMCs. Sharing of genetic predisposition loci partially explain the co-existence of CD with other IMCs, however, it is intriguing to see that many of the CD associated IMCs are, like CD itself, increasing their incidence over time. This may indicate that common shared environmental factors have morphed giving rise to greater number of cases of these different conditions. In our study, this hypothesis is supported by the fact that the observed decrease in co-existent ATD is not explained by the age of the patients at data collection or the age of CD diagnosis, but by calendar year of diagnosis.*

*We also found that HLA & non-HLA PRS can predict the age of CD diagnosis in the case of patients with a family history of CD. Moreover, similarly to the case of T1DM, where lower risk genotypes are more prevalent in recent years, our PRS indicate a decreasing genetic burden in CD over the period studied. Notably, HLA-DQ2.5 heterozygotes have latterly become more common in the patient population, which may partly explain the increase in frequency of the disease. Also certain HLA and non-HLA SNPs were identified as significant contributors to this decrease in PRS due to variation in their frequency. The MHC marker rs241420 is of special interest because it shows the largest association with CD, and has been linked to expression of the antiviral gene SKIV2L. Viral infections may be a triggering event in CD, and the association with rs241420 based on existing data, suggests that patients in the later time periods of our study may express less SKIV2L. This observation is interesting given that viral infection is an environmental factor that could therefore be associated with changes in presentation or prevalence.*

*In conclusion, this body of work has reported a changing clinical picture that is linked to a decrease in genetic burden over the duration of our study. This implies that patients are being drawn from a changing genetic pool under the influence of mutating environmental factors.*



## **Acknowledgements**

I would like to thank everyone who has supported and helped me to complete this thesis, especially my supervisor Professor Ross McManus for his unconditional support throughout this tempestuous journey. He has always been there for me when my patience was exhausted. Also especially Professor Nassir Mahmud for his invaluable help even during extremely difficult circumstances for him. To all consultants and clinical and administrative staff at the various participating hospitals that have helped in data access and collection. To all academic staff that helped with data interpretation and analysis. I would also like to thank Drs David Kevans, Shigeki Nakagome and Lina Zgaga for their interesting and helpful discussion and suggestions. To my husband whose eccentric fascination with formatting documents to perfection saved the day, and made possible the timely submission of this thesis. To my son Liam for keeping everything in perspective. Finally, to Sandy for her silent canine understanding.

***“ We choose to go to the Moon. We choose to go to the Moon in this decade and do the other things, not because they are easy, but because they are hard; because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one we intend to win”***

John F. Kennedy, Rice University Houston, Texas, September 12 1962

***“After landing his invasion forces on the shores of Veracruz, Mexico the sixteenth-century Spanish conquistador Hernan Cortes would immediately burn his own boats. He was sending his army a message: ‘We can’t turn back. Either we succeed here or we die here.’ Excuses were not an option”***

Lou Holtz, football coach Notre Dame Fighting Irish, 1986-1996

***“ Failure is not an option ”***

Apollo XIII, The Movie

## **Statement of work**

This work was undertaken by the author with the Department of Clinical Medicine, Genomics Research Group, Trinity Translational Medicine Institute, Trinity College Dublin and St. James Hospital

## **Publications arising from this thesis**

1. Dominguez Castro P, Harkin G, Hussey M et al. Changes in Presentation of Celiac in Ireland from the 1960s to 2015. *Clin Gastroenterol Hepatol*. 2017 Jun;15 (6): 864-871.

## **List of abbreviations**

**AA**= Alopecia areata

**AGA**= Antigliadin antibodies

**AS**= Ankylosing spondylitis

**ATD**= Autoimmune thyroid disease

**BLD**= Blood

**BMD**= Bone mineral density

**BMI**= Body mass index

**CD**= Coeliac disease

**DGPA**= Anti-deamidated gliadin peptide antibodies

**DH**= Dermatitis herpetiformis

**DR**= Distant relative

**EMA**= Endomysium antibodies

**eQTL**= Expression quantitative trait loci

**ESPGHAN**= European society for paediatrics gastroenterology and hepatology

**ESR**= Erythrocyte sedimentation rate

**FASLG**= Fas ligand

**FDR**= First degree relative

**GFD**= Gluten free diet

**GI**= Gastrointestinal

**GUH**= Galway university hospital

**GWAS**= Genome wide association studies

**GWS**= Genome wide significance

**IBD**= Inflammatory bowel disease

**IBS**= Inflammatory bowel syndrome

**IDA**= Iron deficiency anaemia

**IEL**= Intraepithelial lymphocyte

**IL-15**= Interleukin 15

**IMC**= Immune-mediated condition

**LA**= Lazarotide acetate

**MHC**= Major histocompatibility complex

**MS**= Multiple sclerosis

**NCGS**= non-coeliac gluten sensitivity

**tTGA**= anti-Tissue Transglutaminase Antibodies

## Table of Contents

|                                                                                                                               |           |
|-------------------------------------------------------------------------------------------------------------------------------|-----------|
| Declaration.....                                                                                                              | i         |
| Summary.....                                                                                                                  | ii        |
| Acknowledgements.....                                                                                                         | iv        |
| Statement of work .....                                                                                                       | vi        |
| Publications arising from this thesis .....                                                                                   | vii       |
| List of abbreviations .....                                                                                                   | viii      |
| Table of Contents .....                                                                                                       | x         |
| Index of Tables.....                                                                                                          | xiv       |
| Index of Figures .....                                                                                                        | xvi       |
| <b>Chapter 1 General Introduction .....</b>                                                                                   | <b>1</b>  |
| 1.1 Overview.....                                                                                                             | 2         |
| 1.2 CD prevalence, incidence and clinical presentation.....                                                                   | 5         |
| 1.3 CD pathogenesis the immune system, dietary gluten and other environmental factors..                                       | 9         |
| 1.4 Immunopathogenesis .....                                                                                                  | 14        |
| 1.5 CD and co-existent autoimmune conditions; the protective role of the GFD. ....                                            | 17        |
| 1.6 CD serology and biopsy: past, present and future.....                                                                     | 22        |
| 1.7 Novel serological biomarkers to monitor CD.....                                                                           | 27        |
| 1.8 Alternatives to the gluten free diet.....                                                                                 | 30        |
| 1.9 CD Genetics.....                                                                                                          | 34        |
| <b>Chapter 2 Investigation of Coeliac Disease Clinical Phenotype in Ireland and its<br/>Potential Changes Over Time .....</b> | <b>37</b> |
| 2.1 Introduction.....                                                                                                         | 38        |
| 2.2 Aims.....                                                                                                                 | 41        |
| 2.3 Methods .....                                                                                                             | 42        |
| 2.3.1 Study design and study population.....                                                                                  | 42        |
| 2.3.2 Data collection and definition of study variables .....                                                                 | 43        |
| 2.3.3 Assays utilised to determine coeliac serology .....                                                                     | 43        |
| 2.3.4 Statistical analysis .....                                                                                              | 44        |
| 2.3.4.1 Statistical analysis of changes in clinical presentation over time .....                                              | 44        |
| 2.3.4.2 Statistical analysis of changes in clinical presentation over time per geographical<br>area .....                     | 45        |
| 2.3.4.3 Statistical analysis of BMI changes over time.....                                                                    | 45        |
| 2.3.4.4 Regression analysis of clinical changes over time per geographical area. ....                                         | 45        |
| 2.4 Results .....                                                                                                             | 47        |

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| 2.4.1 Characteristics of the sample.....                                                | 47 |
| 2.4.2 Characteristics of the sample over time .....                                     | 56 |
| 2.4.3 Comparison of characteristics of the sample segregated by geographical area. .... | 59 |
| 2.4.4 Changes in BMI and weight over time .....                                         | 66 |
| 2.4.5 Correlation of serology with biopsy at diagnosis .....                            | 75 |
| 2.5 Discussion .....                                                                    | 76 |
| 2.5.1 Changes in CD clinical presentation over time .....                               | 76 |
| 2.5.2 Gender differences in CD diagnosis and clinical presentation.....                 | 78 |
| 2.5.3 CD and associated T1DM and ATD over time.....                                     | 78 |
| 2.5.4 CD differences per geographical location of diagnosis.....                        | 79 |
| 2.5.5 CD clinical presentation in the elderly .....                                     | 81 |
| 2.5.6 Body mass index (BMI) changes in CD patients .....                                | 84 |
| 2.5.7 Serum biomarkers at CD diagnosis.....                                             | 88 |
| 2.6 Conclusion .....                                                                    | 90 |

### **Chapter 3 Investigation of Coeliac Disease Co-existence With Other Immune**

#### **Mediated Conditions and Pattern of Autoimmune Thyroid Disease Over Time..... 91**

|                                                                                                                                                    |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.1 Introduction.....                                                                                                                              | 92  |
| 3.2 Aims.....                                                                                                                                      | 95  |
| 3.3 Methods .....                                                                                                                                  | 96  |
| 3.3.1 Study design and study population.....                                                                                                       | 96  |
| 3.3.2 Data collection and definition of study variables .....                                                                                      | 96  |
| 3.3.3 Statistical analysis .....                                                                                                                   | 97  |
| 3.2 Results .....                                                                                                                                  | 99  |
| 3.2.1 General population prevalence studies of immune mediated conditions (IMCs) and prevalence of IMCs among CD patients and their relatives..... | 99  |
| 3.2.2. Immune mediated conditions (IMCs) prevalence ratios in CD patients and their parents.....                                                   | 106 |
| 3.2.3. Regression analysis to predict the development of any second autoimmunity or ATD in CD patients. ....                                       | 111 |
| 3.2.4 Genetics, environmental factors, and epidemiology of co-existent IMCs in CD patients.....                                                    | 116 |
| 3.3 Discussion .....                                                                                                                               | 119 |
| 3.3.1 CD and Type 1 Diabetes Mellitus (T1DM).....                                                                                                  | 120 |
| 3.3.2 CD and autoimmune thyroid disease (ATD). ....                                                                                                | 121 |
| 3.3.3 CD and inflammatory bowel disease (IBD). ....                                                                                                | 124 |
| 3.3.4 CD comorbidities with other immune-mediated conditions (IMCs).....                                                                           | 126 |



|                                                                                                                                                             |            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.4 Conclusion .....                                                                                                                                        | 130        |
| <b>Chapter 4 Genetic Burden and Coeliac Disease Phenotype.....</b>                                                                                          | <b>131</b> |
| 4.1 Introduction.....                                                                                                                                       | 132        |
| 4.2 Aims.....                                                                                                                                               | 135        |
| 4.3 Methods .....                                                                                                                                           | 136        |
| 4.3.1 Study design and study population.....                                                                                                                | 136        |
| 4.3.2 Clinical phenotypes.....                                                                                                                              | 136        |
| 4.3.3 Genotype/phenotype analysis method .....                                                                                                              | 137        |
| 4.3.4 PRS calculation.....                                                                                                                                  | 137        |
| 4.3.4 Analysis of changing alleles frequencies over time .....                                                                                              | 138        |
| 4.4 Results .....                                                                                                                                           | 140        |
| 4.4.1 Correlation of HLA & Non-HLA, HLA only and Non-HLA PRS with clinical phenotype. ....                                                                  | 141        |
| 4.4.2 Correlation between age of diagnosis with CD and HLA & non-HLA, HLA only and non-HLA only PRS.....                                                    | 148        |
| 4.4.3 Correlation between age of diagnosis with CD and HLA & non-HLA, HLA only and non-HLA only PRS in CD patients with a family history of CD (n=154)..... | 150        |
| 4.4.4 Analysis of HLA & non-HLA, HLA only and non-HLA only PRS changes over time. ....                                                                      | 152        |
| 4.4.5 Investigation of whether allele frequencies contributing to PRS change over time. ....                                                                | 159        |
| 4.5 Discussion .....                                                                                                                                        | 173        |
| 4.5.1 Genetic burden and clinical phenotype.....                                                                                                            | 173        |
| 4.5.2 Genetic burden over time.....                                                                                                                         | 175        |
| <b>Chapter 5 General Discussion .....</b>                                                                                                                   | <b>178</b> |
| <i>Future Directions</i> .....                                                                                                                              | 184        |
| <b>Chapter 6 Bibliography .....</b>                                                                                                                         | <b>185</b> |
| Appendices .....                                                                                                                                            | 215        |
| Appendix I. Template used to extract data from patients' charts. ....                                                                                       | 216        |
| Appendix II. Questionnaire used during patient visit to CD clinic. ....                                                                                     | 218        |
| Appendix III. Supplementary Table 1. The Oslo classification of coeliac disease.....                                                                        | 221        |
| Appendix IV. Supplementary Figure 1. Geographical division of CD patients sample in the present study (n=749). ....                                         | 222        |
| Appendix V. Supplementary Figure 2 .....                                                                                                                    | 223        |
| Appendix VI. Supplementary Figure 3.....                                                                                                                    | 224        |
| Appendix VII. Supplementary Table 2.....                                                                                                                    | 225        |

|                                                                                                                                         |     |
|-----------------------------------------------------------------------------------------------------------------------------------------|-----|
| Appendix VIII. Supplementary Table 2 .....                                                                                              | 226 |
| Appendix IX. Supplementary Figure 4.....                                                                                                | 228 |
| Appendix X. Supplementary Figure 5. ....                                                                                                | 229 |
| Appendix XI. Supplementary Table 3. Characteristics of the 436 patients with genotypic<br>information. ....                             | 230 |
| Appendix XII. Analysis of HLA & Non-HLA (Table A) and HLA only (Table B) and PRS<br>changes over time using a binary time variable..... | 231 |
| Appendix XIII. Supplementary Table 3. HLA SNPs which differ $\geq 0.04$ in frequency as shown<br>in Figure 4.12.....                    | 233 |
| Appendix XIII. Supplementary Table 3 (cont.) .....                                                                                      | 234 |
| Appendix XIV. Supplementary Table 4. Non-HLA SNPs which differ $\geq 0.04$ in frequency as<br>shown in Figure 4.12.....                 | 235 |
| Appendix XIV. Supplementary Table 4 (cont.).....                                                                                        | 236 |
| Appendix XV. Selected list of abstracts accepted for presentation at national and<br>international meetings. ....                       | 237 |
| Appendix XVI Paper Published in Clinical Gastroenterology and Hepatology .....                                                          | 238 |

## Index of Tables

|                                                                                                                                                                                                                                                           |            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Table 1.1 Studies assessing the risk to develop co-existent IMCs in CD patients according to gluten exposure.....</i>                                                                                                                                  | <i>19</i>  |
| <i>Table 1.2 Novel serological biomarkers for the diagnosis of CD. ....</i>                                                                                                                                                                               | <i>26</i>  |
| <i>Table 1.3 Novel serological biomarkers to monitor CD.....</i>                                                                                                                                                                                          | <i>29</i>  |
| <i>Table 2.1 Characteristics of coeliac patients (n=749).....</i>                                                                                                                                                                                         | <i>47</i>  |
| <i>Table 2.2 Prevalence of coeliac disease extra-intestinal symptoms and co-existent immune-mediated conditions (n=749).....</i>                                                                                                                          | <i>48</i>  |
| <i>Table 2.3 Characteristics of the sample segregated by gender (n=749).....</i>                                                                                                                                                                          | <i>49</i>  |
| <i>Table 2.4 Characteristics of the CD patient sample diagnosed in adulthood segregated by degree of damage to the small bowel mucosa classified according to Marsh criteria (n=355).....</i>                                                             | <i>50</i>  |
| <i>Table 2.5 Characteristics of the sample segregated geographically by patients diagnosed in Galway University Hospital in the west of Ireland and those diagnosed in the other participating centres in Dublin in the east of Ireland (n=749). ....</i> | <i>51</i>  |
| <i>Table 2.6 Characteristics of the sample diagnosed in adulthood segregated by BMI category at diagnosis (n=402).....</i>                                                                                                                                | <i>52</i>  |
| <i>Table 2.7 Characteristics of the sample diagnosed in adulthood segregated by the age at diagnosis with coeliac disease; ≥18-35 years, &gt;35-64 years and ≥65 years.....</i>                                                                           | <i>55</i>  |
| <i>Table 2.8 Clinical characteristics at the time of coeliac disease diagnosis in adulthood and co-existence of immune-mediated conditions over 5 time periods (n=652). ....</i>                                                                          | <i>57</i>  |
| <i>Table 2.9 Comparison of clinical characteristics and associated immune-mediated conditions between CD patients diagnosed in adulthood in GUH (Table A) and in the east of Ireland (Table B) over 4 time periods. ....</i>                              | <i>60</i>  |
| <i>Table 2.10 Binary logistic regression using time period and geographical location of diagnosis with their unadjusted and adjusted ORs (n=602).....</i>                                                                                                 | <i>65</i>  |
| <i>Table 2.11 Coeliac patients BMI at diagnosis and at follow-up one year after diagnosis (kg/m<sup>2</sup>). ....</i>                                                                                                                                    | <i>67</i>  |
| <i>Table 2.12 BMI and weight changes in CD patients 1 year after diagnosis (n=166).....</i>                                                                                                                                                               | <i>70</i>  |
| <i>Table 3.1 Disease prevalence studies for comparison. ....</i>                                                                                                                                                                                          | <i>99</i>  |
| <i>Table 3.2 Hypothyroidism, Type 1 diabetes mellitus (T1DM) and rheumatoid arthritis (RA) prevalence studies for comparison used by Neuhasen et al.....</i>                                                                                              | <i>102</i> |
| <i>Table 3.3 CD patients' IMCs prevalence and prevalence of CD and IMCs among relatives categories. ....</i>                                                                                                                                              | <i>103</i> |
| <i>Table 3.4 Prevalence and type of other co-existent IMCs in CD patients and their relatives. ....</i>                                                                                                                                                   | <i>106</i> |
| <i>Table 3.5 Prevalence ratios of co-existent IMCs in CD cases using prevalence studies in Table 3.1. ....</i>                                                                                                                                            | <i>107</i> |
| <i>Table 3.6 Prevalence ratios and presentation of co-existent T1DM, Hypothyroidism and RA in CD cases. ....</i>                                                                                                                                          | <i>110</i> |
| <i>Table 3.7 Prevalence ratios (PRs) of co-existent immune mediated conditions IMCs in parents of CD patients using prevalence studies in Table 3.1. ....</i>                                                                                             | <i>111</i> |
| <i>Table 3.8 Regression analysis to predict the probability of CD patients diagnosed in adulthood of developing co-existent ATD (n=652).....</i>                                                                                                          | <i>113</i> |

|                                                                                                                                                                                                                                            |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Table 3.9 Epidemiology, shared genetics and potential environmental factors of co-existent IMCs and CD.....</i>                                                                                                                         | <i>117</i> |
| <i>Table 4.1 Mean PRS for all markers according to phenotypic characteristics of CD patients. ....</i>                                                                                                                                     | <i>141</i> |
| <i>Table 4.2 Mean PRS for HLA only markers according to phenotypic characteristics of CD patients....</i>                                                                                                                                  | <i>143</i> |
| <i>Table 4.3 Mean PRS for non-HLA only markers according to phenotypic characteristics of CD patients. ....</i>                                                                                                                            | <i>145</i> |
| <i>Table 4.4 Correlation of clinical phenotype with mean PRS for all markers (HLA &amp; non-HLA) segregated into patients with a family history of CD (n=154; Table A) and those without, a family history of CD (n=282;Table B). ....</i> | <i>147</i> |
| <i>Table 4.5 Analysis of HLA &amp; non-HLA PRS at different p-value levels over time. ....</i>                                                                                                                                             | <i>153</i> |
| <i>Table 4.6 Analysis of HLA only PRS at different p-value levels over time. ....</i>                                                                                                                                                      | <i>154</i> |
| <i>Table 4.7 Analysis of non-HLA only PRS at different p-value levels over time.....</i>                                                                                                                                                   | <i>155</i> |
| <i>Table 4.8 Analysis of non-HLA only PRS at different p-value levels over time using binary time variable <math>\leq 2005</math> and <math>\geq 2006</math>. ....</i>                                                                     | <i>159</i> |
| <i>Table 4.9 HLA markers showing significant changes in frequency between 1995-2006.....</i>                                                                                                                                               | <i>168</i> |
| <i>Table 4.10 Non-HLA markers showing significant changes in frequency between 1995-2006.....</i>                                                                                                                                          | <i>169</i> |

## Index of Figures

|                                                                                                                                                                                                                                                                                                                                                                                         |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Figure 1.1 Suggested causal model of the multifactorial aetiology of coeliac disease taken from Ivarsson 2005 (Ivarsson, 2005).</i>                                                                                                                                                                                                                                                  | 12  |
| <i>Figure 1.2 Hypothetical microbiome biomarkers of CD organized by strength of evidence taken from Sanz et al. 2015.</i>                                                                                                                                                                                                                                                               | 13  |
| <i>Figure 1.3 Proposed model for CD pathogenesis TGt = Tissue transglutaminase; IEL = intraepithelial lymphocytes; NKG2D = natural killer group 2D receptor; MICA = MHC class I chain-related protein A; CD71 = cluster of differentiation 71 or transferrin receptor protein 1; DCs = dendritic cells (Sanz, 2015).</i>                                                                | 16  |
| <i>Figure 1.4 Organs involved in CD co-existent with other IMCs and genetic overlap between CD and other IMCs (Lundin and Wijmenga, 2015).</i>                                                                                                                                                                                                                                          | 21  |
| <i>Figure 1.5 Suggested algorithm for CD testing.</i>                                                                                                                                                                                                                                                                                                                                   | 23  |
| <i>Figure 1.6 Characteristics of CD biopsies.</i>                                                                                                                                                                                                                                                                                                                                       | 25  |
| <i>Figure 1.7 CD pathogenesis and future non-dietary treatments (Lindfors et al., 2012).</i>                                                                                                                                                                                                                                                                                            | 33  |
| <i>Figure 2.1 Degree of damage to the small bowel mucosa classified according to Marsh criteria in patients diagnosed with CD in adulthood classified per BMI category &lt;25 kg/m<sup>2</sup> and ≥25 kg/m<sup>2</sup>.</i>                                                                                                                                                            | 54  |
| <i>Figure 2.2 Changes in clinical presentation of CD at the time of diagnosis in adulthood over 5 time periods (n=602).</i>                                                                                                                                                                                                                                                             | 56  |
| <i>Figure 2.3 Trend analysis of age at diagnosis with CD in adulthood over 5 time periods (n=652).</i>                                                                                                                                                                                                                                                                                  | 58  |
| <i>Figure 2.4 Trend analysis of BMI at diagnosis with CD in adulthood over 5 time periods (n=402).</i>                                                                                                                                                                                                                                                                                  | 58  |
| <i>Figure 2.5 Changes in clinical presentation of CD patients at the time of diagnosis in adulthood in Galway University Hospital (GUH) (Figure A) (p=0.41), and other participating centres on the east coast (Figure B) (p=0.004) over 4 time periods.</i>                                                                                                                            | 63  |
| <i>Figure 2.6 Changes in CD patients BMI at diagnosis from 1995 up to 2015 compared to the general population.</i>                                                                                                                                                                                                                                                                      | 68  |
| <i>Figure 2.7 Figure A, CD patients BMI at diagnosis (n=265) and at follow-up (2 years) in the northern Ireland paper by Dickey et al. (n=188); Figure B, CD patients BMI at diagnosis and at follow-up in the US paper by Kabbani et al. (mean 3.3 years) (n=679); Figure C, CD patients BMI at diagnosis (n=369) and at follow-up (2 years) in the paper by Cheng et al. (n=309).</i> | 71  |
| <i>Figure 2.8 Percentage of patients classified as having excellent or good adherence to the diet a year after diagnosis with CD classified per BMI category at diagnosis (n=27).</i>                                                                                                                                                                                                   | 73  |
| <i>Figure 2.9 CD patients diagnosed in adulthood BMI at diagnosis and DH co-existence over five time periods.</i>                                                                                                                                                                                                                                                                       | 74  |
| <i>Figure 2.10 Median tissue transglutaminase antibodies fold rise in patients with coeliac disease having different levels of small bowel damage classified according to Marsh criteria.</i>                                                                                                                                                                                           | 75  |
| <i>Figure 3.1 Prevalence and type of co-existent IMCs in CD patients (Figure A) and their relatives (Figure B).</i>                                                                                                                                                                                                                                                                     | 105 |

|                                                                                                                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Figure 3.2 Analysis of age at diagnosis with CD, age at diagnosis with co-existent ATD, median age at data collection and prevalence of co-existent ATD for patients diagnosed in adulthood over 5 time periods (n=652).</i> | 114 |
| <i>Figure 3.3 Analysis of parameters in Figure 3.2 with subcohort of patients diagnosed in adulthood and &gt;55 years at data collection over 5 time periods (n=354).</i>                                                       | 115 |
| <i>Figure 3.4 Prevalence of co-existent ATD for female CD patients diagnosed in adulthood and &gt;55 years at data collection divided into 5 periods according to the year of diagnosis (n=226).</i>                            | 116 |
| <i>Figure 4.1 Diagram showing total sample and its subsets classified according to availability of genotype information.</i>                                                                                                    | 136 |
| <i>Figure 4.2 Correlation between PRS for all markers (HLA &amp; non-HLA) and age of CD Diagnosis.</i>                                                                                                                          | 148 |
| <i>Figure 4.3 Correlation between HLA only PRS and age of CD Diagnosis.</i>                                                                                                                                                     | 149 |
| <i>Figure 4.4 Correlation between non-HLA only PRS and age of CD diagnosis.</i>                                                                                                                                                 | 149 |
| <i>Figure 4.5 Correlation between PRS for all markers (HLA &amp; non-HLA) and age of CD diagnosis in patients with family history of CD.</i>                                                                                    | 151 |
| <i>Figure 4.6 Correlation between HLA only PRS and age of CD diagnosis in patients with family history of CD.</i>                                                                                                               | 151 |
| <i>Figure 4.7 Correlation between non-HLA only PRS and age of CD diagnosis in patients with family history of CD.</i>                                                                                                           | 152 |
| <i>Figure 4.8 Mean PRS for All SNPs According to Time Period.</i>                                                                                                                                                               | 156 |
| <i>Figure 4.9 Mean HLA only PRS According to Time Period.</i>                                                                                                                                                                   | 157 |
| <i>Figure 4.10 Mean non-HLA PRS According to Time Period.</i>                                                                                                                                                                   | 157 |
| <i>Figure 4.11 Percentage of HLA-DQ2 homozygous patients over three time periods (Figure A) and over five time periods (Figure B).</i>                                                                                          | 161 |
| <i>Figure 4.12 Scatterplot representing changes in gene variants frequencies between the first time period <math>\leq 1995</math> and the last time period <math>\geq 2006</math>.</i>                                          | 163 |
| <i>Figure 4.13 HLA alleles simulation distribution plot based on 1000 simulation data.</i>                                                                                                                                      | 164 |
| <i>Figure 4.14 Non-HLA alleles simulation distribution plot based on 1000 simulation data.</i>                                                                                                                                  | 166 |
| <i>Figure 4.15 Changes in gene expression of SKIV2L associated eQTL locus rs411337 in various tissues.</i>                                                                                                                      | 172 |

## **Chapter 1 General Introduction**

## 1.1 Overview

Coeliac disease (CD) is a chronic small intestinal immune-mediated enteropathy triggered by dietary gluten in genetically predisposed individuals (Ludvigsson et al., 2013a). Gluten refers to a large group of alcohol soluble prolamins found in wheat (gliadins and glutenins) which constitute approximately half of the protein in the cereal grain. Other similar prolamins which are also harmful for CD patients are found in rye (secalins) and, barley (hordeins) (Sollid, 2002, Ludvigsson et al., 2013a). CD has an estimated worldwide prevalence of 1% in Western societies affecting both adults and children (Dube et al., 2005, Reilly and Green, 2012).

In order for the pathogenic reaction to develop, dietary gluten must cross the mucosal epithelial layer into the lamina propria of susceptible individuals. There it is deamidated by tissue transglutaminase, and subsequently presented by antigen presenting cells on major histocompatibility complex (MHC) class II molecules HLA-DQ2 and DQ8. This causes an immune cascade which ultimately results in the infiltration of the intestinal epithelium by lymphocytes and ensuing tissue destruction leading in many cases to total villous atrophy (Kupfer and Jabri, 2012). However, HLA-DQ2 and DQ8 molecules have a carrier frequency of 30-40% in western populations and gluten is a ubiquitous dietary staple, indicating that other factors must be involved in causing CD (Kagnoff, 2007). In the last decade, genome wide association studies (GWAs) have identified many additional genetic risk loci, confirming the belief that CD is a polygenic autoimmune condition (van Heel et al., 2007, Trynka et al., 2011). In addition, mounting evidence that CD is increasing in prevalence and incidence suggest that the environmental trigger of CD is not only gluten, and other potential environmental factors most likely play a role in the development of the condition (Lebwohl et al., 2014b).

Symptoms resulting from gluten consumption in CD patients may include manifestations such as diarrhoea, steatorrhea and weight loss which ultimately leads to malabsorption of nutrients. This type of CD presentation is referred to as classical CD (Ludvigsson et al., 2013a). However, CD can also present without malabsorption symptoms with patients instead suffering from constipation or abdominal pain, or with extra-intestinal symptoms such as osteoporosis or anaemia. This type of CD has been termed non-classical CD (Ludvigsson et al., 2013a, Leffler et al., 2015a). CD can also present subclinically with manifestations which are below the threshold for detection by clinicians. Therefore, a high level of awareness is needed by physicians to avoid misdiagnosis or underdiagnosis of the condition. Classical CD predominates in



paediatric patients under the age of 3, whereas non-classical presentation is more prevalent in older children and adults (Telega et al., 2008, Vivas et al., 2008, Volta et al., 2014, Kivela et al., 2015).

CD diagnosis is carried out by measuring levels of anti-tissue transglutaminase antibodies (Hadley et al.) which is the serological test of choice in most countries due to its high sensitivity and specificity (Leffler and Schuppan, 2010). However, until 2012, despite the high efficacy of serologic tests, the gold-standard for diagnosis was a duodenal biopsy with histological changes that correlate with CD (Marsh 2-3) (ESPGHAN, 1990). In 2012, new guidelines from the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) allowed the biopsy to be omitted in children with CD symptoms based on the presence of tTGA 10 or more times the normal upper test limit and positive EMA in individuals positive for HLA DQ2/DQ8 result (Husby et al., 2012). However, a biopsy is still required in adults (Ludvigsson et al., 2014a). Due to the continuous increase in CD prevalence and the limitations of the biopsy for monitoring disease progression after diagnosis, there is a need to develop better serology based diagnostic techniques (Kurppa et al., 2016).

The only available treatment for CD is strict life-long adherence to a gluten free diet (GFD) which is in practice, difficult. Given its important role in preventing complications such as bone disease in CD patients, and its potential role in preventing co-existent immune-mediated conditions (IMCs), long term management of the condition by a physician and/or dietician is necessary (Haines et al., 2008). The majority of patients respond well to the GFD with their symptoms resolving completely, followed in most cases by normalisation of abnormal serology and histology (Newnham et al., 2016). Whereas small bowel mucosa recovery happens fast in the paediatric population, it is slow in CD patients diagnosed in adulthood. Studies indicate it takes approximately five years for the majority of adult patients to achieve mucosal healing (Wahab et al., 2002, Vecsei et al., 2014, Bannister et al., 2014, Galli et al., 2014, Haere et al., 2016). However, from 7% to 30% of CD patients suffer from non-responsive CD (NRCD) with no resolution of symptoms and/or mucosal recovery. Nonetheless, in the majority of cases of NRCD, the cause is thought to be gluten exposure (either inadvertent or deliberate) or a different or additional diagnosis, such as refractory CD (Leffler et al., 2007a, Al-Bawardy et al., 2017). Due to the high percentage of patients suffering from NRCD due to inadvertent gluten exposure, novel non-dietary therapies are needed and are an area of active research; among others, gluten digestion by endopeptidases or tight junction enhancers, are being investigated (Maki, 2014, Castillo et al., 2015, Kurada et al., 2016).

Refractory CD is defined as persistent villous atrophy despite being over 6-12 months on a verified GFD, and in the absence of other causes for non-responsive treated CD or the presence of malignancy (Rubio-Tapia and Murray, 2010, Malamut et al., 2012b). However, most patients are diagnosed with refractory CD following the reoccurrence of malabsorptive symptoms after an initial response to the GFD for years (Rubio-Tapia and Murray, 2010). The prevalence of refractory CD is not known with precision, but it is rare and probably affects approximately 2% of adult CD patients (Rubio-Tapia and Murray, 2010, Malamut et al., 2012a).

## 1.2 CD prevalence, incidence and clinical presentation

CD was mainly diagnosed based on malabsorption symptoms until the late 1970's, and therefore its prevalence was thought to be low worldwide (0.0003%) (Green and Jabri, 2003, Lohi et al., 2007). CD was also initially thought to affect mainly the paediatric population, however the advent of highly sensitive and specific serological tests, led to the realisation that the condition occurs frequently in adults in the absence of malabsorption symptoms and even in an asymptomatic scenario (Lohi et al., 2007, Reilly et al., 2012). Serological testing allowed mass population screening, which has showed a prevalence of approximately 1% in Europe and the United States (Fasano et al., 2003, Mustalahti et al., 2010, Rubio-Tapia et al., 2012).

CD prevalence varies markedly between countries in Europe ranging from 2.4% in Finland to 0.3% in Germany in adults aged 30-64 years (Mustalahti et al., 2010). In the US, the highest prevalence is found among non-Hispanic whites (1.01%), with black and Hispanic populations having lower CD prevalences of 0.3% and 0.2% respectively (Rubio-Tapia et al., 2012, Choung et al., 2015). Although historically CD was considered to be a condition affecting mainly people of European ancestry, the highest prevalence (5.6%) is found in the Saharawi people, a Berber population indigenous to the Western Sahara (Teresi et al., 2010). The wide variability in prevalences within Europe remains largely unexplained and has lead to the hypothesis that it is caused by different environmental exposures (Mustalahti et al., 2010). Even the Saharawi population having one of the highest frequencies of the HLA-DR3 haplotype in the world (38.9%) does not fully explain the high prevalence of CD in this population (Catassi et al., 2001).

Similar HLA-DR3 frequencies to those in the Saharawis are observed in other populations such as the Sardinian population where the prevalence of CD is approximately 1% (Catassi et al., 2001). Catassi et al. suggest that the change in dietary habits in this population, where the intake of gluten was traditionally low before Spanish colonisation, and where subsequently, wheat, and specially bread, became the main staple food, could play a role in the high CD prevalence. Not only the amount of gluten would play a role in the high CD prevalence in the Saharawi population, but also the timing of its introduction may be important in this context, since it is sometimes introduced as early as in the first month of life.

Catassi's proposed explanation correlates with the so-called Swedish epidemic of CD from 1985 to 1994, which led to the hypothesis that infant feeding practices increased

the risk of developing CD in young children (Ivarsson, 2005). During this time period the incidence of CD in Sweden in children <2 years tripled from 65 to 198 cases per 100,000 person-years, whilst incidence for older children remained practically stable (Ivarsson et al., 2000, Lebowitz et al., 2014b). These changes coincided with low breastfeeding rates at the age of 6 months (50%), a doubling in the amount of flour contained in formula milk from 1981-1983 and a national recommendation to postpone gluten introduction from 4 months to 6 months of age (Ivarsson et al., 2000, Ivarsson, 2005).

A rapid decline in CD incidence among infants started in 1995 which coincided with higher breastfeeding rates and a change in the national recommendation to introduce small amounts of gluten from 4 months of age (Ivarsson et al., 2000, Ivarsson, 2005). However, the cumulative incidence remained high for the cohort of infants born during the epidemic compared to the post epidemic cohort (Ivarsson, 2005). The Swedish epidemic led to the hypothesis that infant feeding has an effect on the risk of developing CD (Lebowitz et al., 2014b). Although recent large randomised controlled trials have failed to find an association between age at gluten introduction and breastfeeding and CD, there seems to be a link between the amount of gluten intake in the first two years of life and CD (Vriezinga et al., 2014, Lionetti et al., 2014, Andren Aronsson et al., 2016a).

The prevalence of CD is extensively reported to be increasing worldwide. Whereas the most obvious explanation for this increase is higher awareness of the condition and improved diagnostic techniques, there also appears to be a parallel increase in true incidence which is not so easily explained (Catassi et al., 2014). A study by Lohi *et al.* (2007) in Finland checked CD prevalence change over time using a historic population sample from 1978-1980 and comparing it to a population sample in 2000-2001. They found that the total prevalence of CD in Finland had increased from 1.05% in 1978-1980 to 1.99% in 2000-2001. In order to reduce the effect of the introduction of better diagnostic techniques in the increase in prevalence, Lohi *et al.* added together diagnosed and unrecognised CD cases. They concluded that a true rise in CD total prevalence had happened in Finland probably due to the action of environmental factors. Moreover, this increase in CD occurred parallel to an increase in the prevalence of Type 1 diabetes (T1DM), with the prevalence of both conditions nearly doubling (Lohi et al., 2007).

These findings have been corroborated by posterior studies (Rubio-Tapia et al., 2009, Catassi et al., 2010). Rubio-Tapia *et al.* reported a 4.5 and 4 fold increase in the prevalence of CD in two recent US population cohorts compared to a historic cohort

from 1948-1954. Moreover, they also reported increased mortality rates in those with undiagnosed CD compared to seronegative people. Catassi *et al.* also reported a doubling of CD prevalence in the US in the period between 1974 and 1989 (from 0.2% to 0.46% respectively). This increase in prevalence was due to increasing numbers of people diagnosed with CD in adulthood. Untreated CD was also linked to a 2.6 fold increase in co-existent autoimmunity and a 1.6 fold increase in osteoporosis, as well as increased death due to malignancy (Catassi *et al.*, 2010).

Therefore, CD prevalence has increased over the past decades, and although better diagnostic techniques and increased awareness have potentially played a role, the above mentioned studies demonstrate a true increase in incidence that is not explained only by improved detection. The rapid increase in prevalence and incidence cannot either be explained by changes in genotype frequencies which do not change quickly in large populations in the absence of, for example, severe selection pressures or extensive migration. Thus environmental factors must be playing a role in this change in CD epidemiology. Moreover, it is intriguing that parallel to this increase in CD incidence, other autoimmune conditions such as T1DM and inflammatory bowel disease (IBD), as well as allergic conditions are becoming more common (Tuomilehto, 2013, Kaplan and Ng, 2016). Co-existence of CD with other autoimmune conditions is well reported, and is likely due to shared genetics, but also importantly, potentially shared environmental factors could be playing a role in the co-existence of conditions and their increase in incidence (Fasano, 2006, Neuhausen *et al.*, 2008, Lundin and Wijmenga, 2015).

Despite the well reported change in CD epidemiology, mass population screening for CD is not recommended because the effect of the GFD on quality of life and life expectancy on undetected patients is still unclear, and it may not be cost effective (Ludvigsson *et al.*, 2014a, Ludvigsson *et al.*, 2015, Wise, 2016). However, serological screening is recommended in a number of situations such as the presence of potential symptoms of the condition (i.e. diarrhoea, failure to thrive in children, unexplained weight loss and anaemia), a previous diagnosis of IBS, presence of associated autoimmune conditions, such as T1DM, dermatitis herpetiformis (DH) and autoimmune thyroid disease (ATD), or a family history of CD (Ludvigsson *et al.*, 2014a, Wise, 2016).

The increase in CD incidence over the past two decades has been accompanied by a change in its clinical presentation, from classical presentation to the majority of patients presenting non-classically or asymptomatic (Murray *et al.*, 2003, Rampertab *et al.*, 2006, Volta *et al.*, 2014). This shift in the condition's clinical phenotype has happened

together with an increase in the age at diagnosis both in the adult and paediatric population (Catassi et al., 2010, Volta et al., 2014, Kivela et al., 2015).

### **1.3 CD pathogenesis the immune system, dietary gluten and other environmental factors**

CD is the result of an interplay between an individual's genotype and environmental factors (Escudero-Hernandez et al., 2016, Al-Bawardy et al., 2017). The role of genetics in the development of the condition has been proven by the high prevalence of CD among first degree relatives of CD patients at 5% and 20% (Rubio-Tapia et al., 2008, Dogan et al., 2012). There is also a more than 80% concordance in monozygotic twins (Greco et al., 2002, Nistico et al., 2006). The disease has been mainly linked to the alleles DQA1\*05 and DQB1\*02 which encode the HLA-DQ2 heterodimer, and alleles DQA1\*03 and DQB1\*0302 which encode the HLA-DQ8 heterodimer. Generally, between 90-95% of CD patients express HLA-DQ2, while the great majority of the remaining patients are HLA-DQ8 positive (Murray et al., 2007, Wijmenga and Gutierrez-Achury, 2014). These HLA alleles are also frequently found in the general Caucasian populations (up to 40% of whom will carry one or more alleles) suggesting that, though they are necessary for the development of the disease, they are not sufficient on their own to cause it. Therefore, a negative test for HLA-DQ2 or HLA-DQ8 has a negative predictive value over 99% (Hadithi et al., 2007b). In recent years, a battery of non-HLA predisposing genes have been discovered through genome wide association studies (GWAS) (Garner et al., 2009, Dubois et al., 2010, Trynka et al., 2011, Ostensson et al., 2013, Garner et al., 2014, Coleman et al., 2015).

The discovery of gluten as the main environmental factor causing CD was made by the Dutch paediatrician Willem Dicke who realised that the protein contained in wheat made his patients sick. He arrived at this conclusion after his CD patients improved noticeably during the Dutch famine in 1944 when there was a shortage of wheat. These patients relapsed after the end of the famine when wheat was reintroduced (Dicke et al., 1953). Yet CD is not a mere combination of the right genotype and gluten exposure, as the latter does not explain the onset of the disease, which for many people happens in adulthood after many years of gluten exposure, and probably depends on other unknown contributing environmental factors. Therefore, although gluten is the essential environmental factor for CD to occur, there are very likely other environmental factors that need to be present before CD can develop (Ivarsson, 2005, Catassi et al., 2010, Lebowitz et al., 2014b).

Other factors associated with a higher risk of CD are; infections such as rotavirus among children and campylobacter in the case of adults, or the so called "hygiene hypothesis" which claims that reduced exposure to microbes in recent decades could

underlie the increase in autoimmune conditions and atopy (Stene et al., 2006, Riddle et al., 2013, Lebwohl et al., 2014b). Elective caesarean delivery has also been shown to increase the risk of developing CD compared to children who had an emergency caesarean delivery and therefore have potentially had contact with the birth canal, however this association is controversial with several recent studies indicating that this is not true (Decker et al., 2011, Marild et al., 2012, Kristensen and Henriksen, 2016, Koletzko et al., 2017). However, caesarean delivery is still associated with other conditions such as T1DM which is closely related to CD (Walker, 2017).

The hygiene hypothesis is supported by several studies, in 2014 a UK study by Whyte *et al.* reported a higher prevalence of CD among children in higher socioeconomic groups which was attributed to both the “hygiene hypothesis” and higher healthcare contact in this group (Whyte et al., 2014). Lebwohl *et al.* (2013) found less prevalence of *H.pylori* infection in CD patients compared to controls. They postulated that *H.pylori* may protect against CD through the recruitment of T-regulatory lymphocytes which would have systemic effects and be protective, a similar situation has been postulated in asthma (Oertli et al., 2012, Arnold et al., 2012, Lebwohl et al., 2013). A 2008 study showing significantly different seropositive CD prevalence between Finnish (1.4%) and Russian Karelia (0.6%) provides further support for the “hygiene hypothesis”. These are geographically close areas which have a similar prevalence of HLA-DQ2 and HLA-DQ8 but differ in their level of economic development (Kondrashova et al., 2008).

However, all the above findings seem to be in contrast with others that have found increased CD rates after *Campylobacter* infection (Riddle et al., 2013). An increased risk of CD has also been found in patients who have taken antibiotics, therefore it is not clear whether the increased risk of CD is caused by the bacterial infection or, on the contrary, is a consequence of the use of antibiotics to treat the bacterial infection (Marild et al., 2014, Lebwohl et al., 2014a). The most likely scenario is that CD pathogenesis involves several environmental factors working synergistically (Ivarsson, 2005, Lebwohl et al., 2014a). Therefore, the “viral infection hypothesis” and the “hygiene hypothesis” do not exclude each other, but it is possible that infection with *H.pylori* could prevent a second hit which could be an enteric viral or bacterial infection and therefore prevent the development of CD (Lebwohl et al., 2014a).

This idea is represented in Figure 1.1 taken from the 2005 paper on the Swedish CD epidemic by Ivarsson. It organises the different factors contributing to CD development hierarchically according to their level of action into structural, associated or directly causal. Therefore, the genotype and dietary gluten are described as necessary causal factors or essential factors for CD to occur. On the other hand, the timing and amount



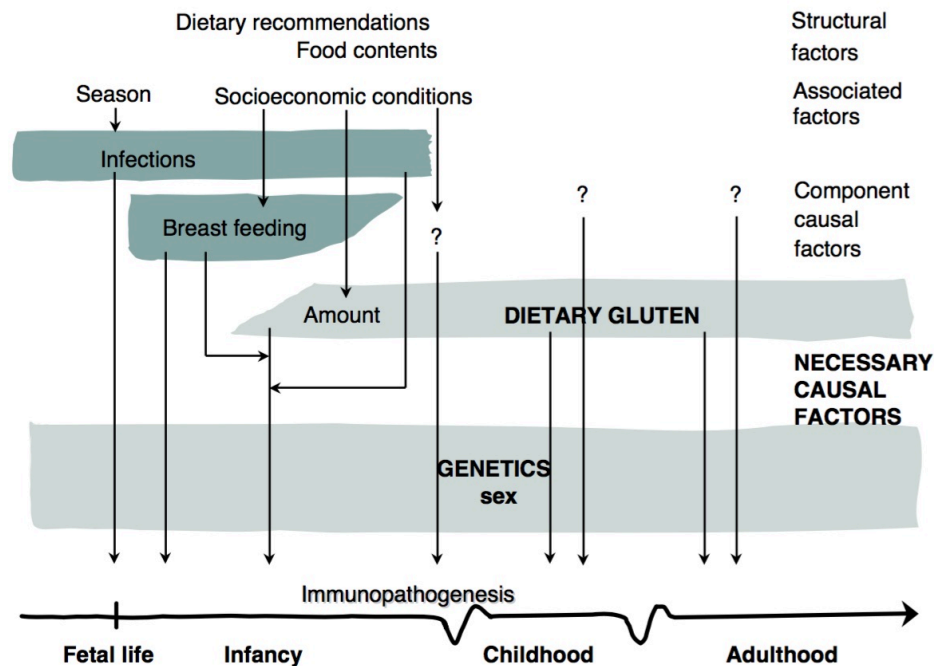
of gluten introduced in the diet during infancy, as well as any infectious episodes suffered would be component causal factors. Thus, when the necessary causal factors and one or more component causal factors are combined, disease may occur. This implies that disease occurrence can be prevented by removing the necessary causal factor 'dietary gluten', and/or one or several of the component causal factors (i.e. viral infection). Ivarsson considers dietary recommendations and food components as structural factors, and socioeconomic status and season of birth as associated factors. Structural and associated factors increase CD risk, but do not have a causal effect (Ivarsson, 2005).

The model by Ivarsson is supported by a recent publication by Kemppainen *et al.* (2017) based on the TEDDY study cohort which follows 8676 children with 1 of 10 HLA-DR-DQ genotypes associated with T1DM and/or CD. They report that rotavirus GI infections increased the risk of developing CD in genetically susceptible children, and interestingly, this risk changes with HLA genotype, time of gluten introduction, breastfeeding, season of birth and rotavirus vaccination (Kemppainen *et al.*, 2017). The risk of developing CD after rotavirus infection in the TEDDY cohort was more pronounced in the cohort of children without HLA-DQ2 alleles who were breastfed for less than four months. This is in contrast with the fact that breastfeeding duration was not found to protect against CD in recently published studies (Lionetti *et al.*, 2014, Aronsson *et al.*, 2015, Andren Aronsson *et al.*, 2016b).

Kemppainen *et al.* conclude that although breastfeeding duration is not directly protective, it provides passive immunity to infants. The lack of this passive immunity, together with a particular composition of the microbiota, which has been proven to be HLA dependent in the case infants genetically predisposed to CD as well as modulated by the type of milk feeding, could favour a proinflammatory environment in the intestine (Olivares *et al.*, 2015, Kemppainen *et al.*, 2017, Palma *et al.*, 2012). Not only does breast milk provide passive immunity but studies have indicated that protective bacteria in breast milk seed the infant gut in early life (Pannaraj *et al.*, 2017). Therefore, lack of breastfeeding coupled to an unfavourable HLA dependent microbiome would favour CD development after rotavirus infection in the presence of dietary gluten. The study by Kemppainen *et al.* reveals the previously described complex interactions between the necessary casual factors (genotype and dietary gluten) and the additional factors needed to develop CD. In conclusion, although CD is the final outcome, the path from patient to patient varies (Leonard and Fasano, 2016).

**Figure 1.1 Suggested causal model of the multifactorial aetiology of coeliac disease taken from Ivarsson 2005 (Ivarsson, 2005).**

A hypothetical model of the complex and multifactorial development of CD is depicted below. In this model factors are organised according to their level of action as; structural, associated or directly causal. Causal factors are divided between necessary causal factors and component causal factors.

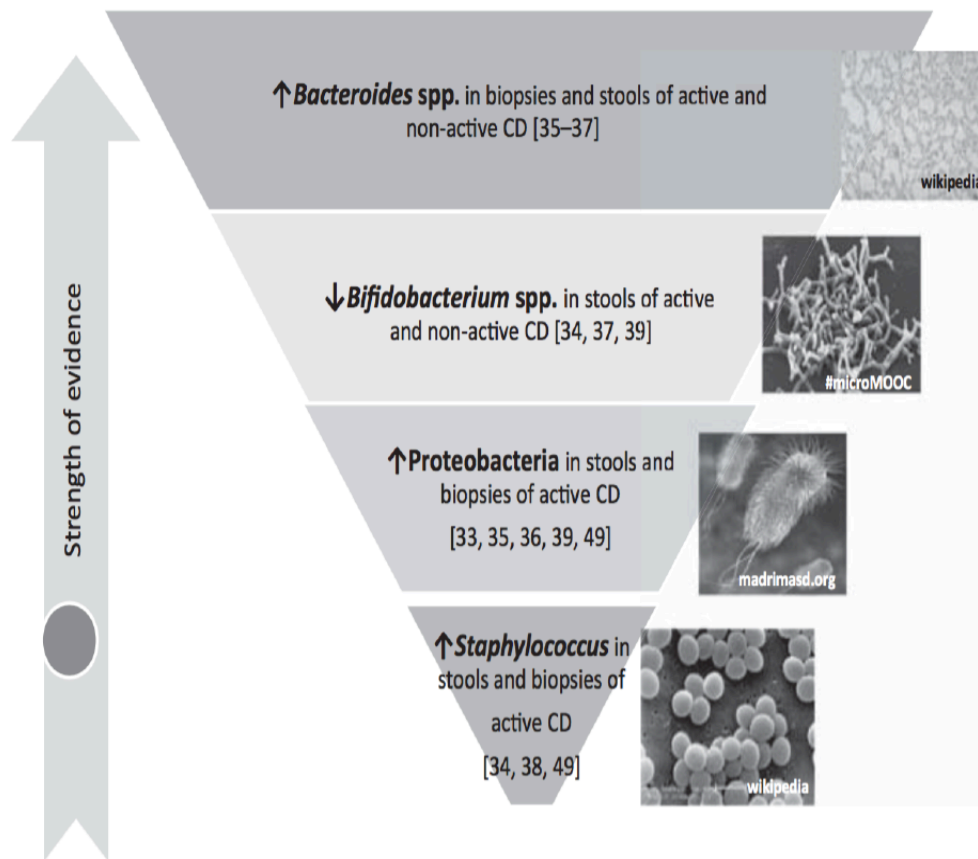


Despite the progress made in understanding the immunopathogenesis of CD, the early events happening after the exposure of the intestinal mucosa to gluten, leading to loss of tolerance and giving rise to autoimmunity, are not fully understood. Changes in the composition of the gut microbiome could be involved in the loss of tolerance to gluten and precede the onset of CD in genetically susceptible individuals (Sellitto et al., 2012, Verdu et al., 2015, Leonard and Fasano, 2016). This microbiome dysbiosis has also been found to persist when patients are following a GFD (Verdu et al., 2015). The exact process by which the microbiome would affect disease initiation and progression is not known, but it has been hypothesized that it could be through activation of the innate immune system, alteration of the permeability of the gut epithelial barrier or intensification of the immune response to gliadin (Verdu et al., 2015)

Gut microbiome dysbiosis is a characteristic of several other immune-mediated conditions (IMCs) apart from CD, however whether the alterations observed are causal or just secondary changes accompanying disease development is still a matter of debate (Levy et al., 2017). Figure 1.2 taken from Sanz *et al.* (2015) shows hypothetical CD microbiome biomarkers organized according to the strength of evidence. In any

case, if the microbiome is important in the initiation of CD, then the identification of specific microbiome profiles which lead to loss of gluten tolerance, could result in early preventive interventions (Sanz, 2015, Leonard and Fasano, 2016).

**Figure 1.2 Hypothetical microbiome biomarkers of CD organized by strength of evidence taken form Sanz et al. 2015.**



Studies have demonstrated microbiome dysbiosis both in active and non-active CD, with an abundance of *Bacteroides* spp. and a decrease of *Bifidobacterium* spp.

## 1.4 Immunopathogenesis

As previously mentioned, CD autoimmunity is the result of an abnormal innate and adaptive immune system response to dietary gluten (Sanz, 2015, Escudero-Hernandez et al., 2016). Gluten proteins have a high content of glutamine and proline residues which makes them resistant to gastrointestinal digestion. The undigested gluten peptides accumulate in the lumen prior to transfer into the sub-luminal space, initiating the cascade of events that leads to enterocyte destruction and eventually to CD. In wheat, gluten peptides are divided into gliadin and glutenins. Gliadin has been suggested to activate innate immunity without resulting in T-cell specific responses (Lauret and Rodrigo, 2013, Sanz, 2015, Escudero-Hernandez et al., 2016). One specific peptide,  $\alpha$ -gliadin 31-43, has been shown to stimulate the production of IL-15 by epithelial cells, macrophages and dendritic cells. Interestingly, increased IL-15 production leads to weakening of the tight junctions between intestinal epithelial cells thus increasing epithelial permeability and potentially allowing the translocation of gluten peptides to the lamina propria (Lauret and Rodrigo, 2013, Escudero-Hernandez et al., 2016) (Figure 1.3).

In a paper in 2000 Fasano *et al.* reported the discovery of the tight junction regulator, the protein zonulin (Fasano et al., 2000). Increased zonulin leads to looser tight junctions which resume their baseline state once the zonulin signal is gone (Fasano, 2012). Increased and continuous production of zonulin has been observed in CD patients with active disease (Lammers et al., 2008). The two triggers identified so far which release zonulin are small intestinal bacterial infection and gliadin (Fasano, 2011). Gliadin binds to the chemokine receptor CXCR3 recruiting the adapter protein MyD88 which leads ultimately to the release of zonulin (Lammers et al., 2008). A recent paper by Lerner and Matthias has also linked the increasing use of food additives with changes in intestinal tight junction permeability as a potential explanation for the increase in autoimmune conditions in recent years (Lerner and Matthias, 2015a). However, the precise mechanism for the development of nutrient-induced autoimmunity is not clear and causality has not been established. Lerner and Matthias highlight several potential mechanisms; molecular mimicry between food additives and tight junction self-antigen, luminal alteration of a nutrient resulting in neo-epitopes in contact with tight junctions, nutrient-epigenetic regulation of genes involved in tight junction regulation, and the hapten hypothesis which proposes that certain chemical products may combine with self components in the body to produce new antigenic molecules (Lerner and Matthias, 2015a).

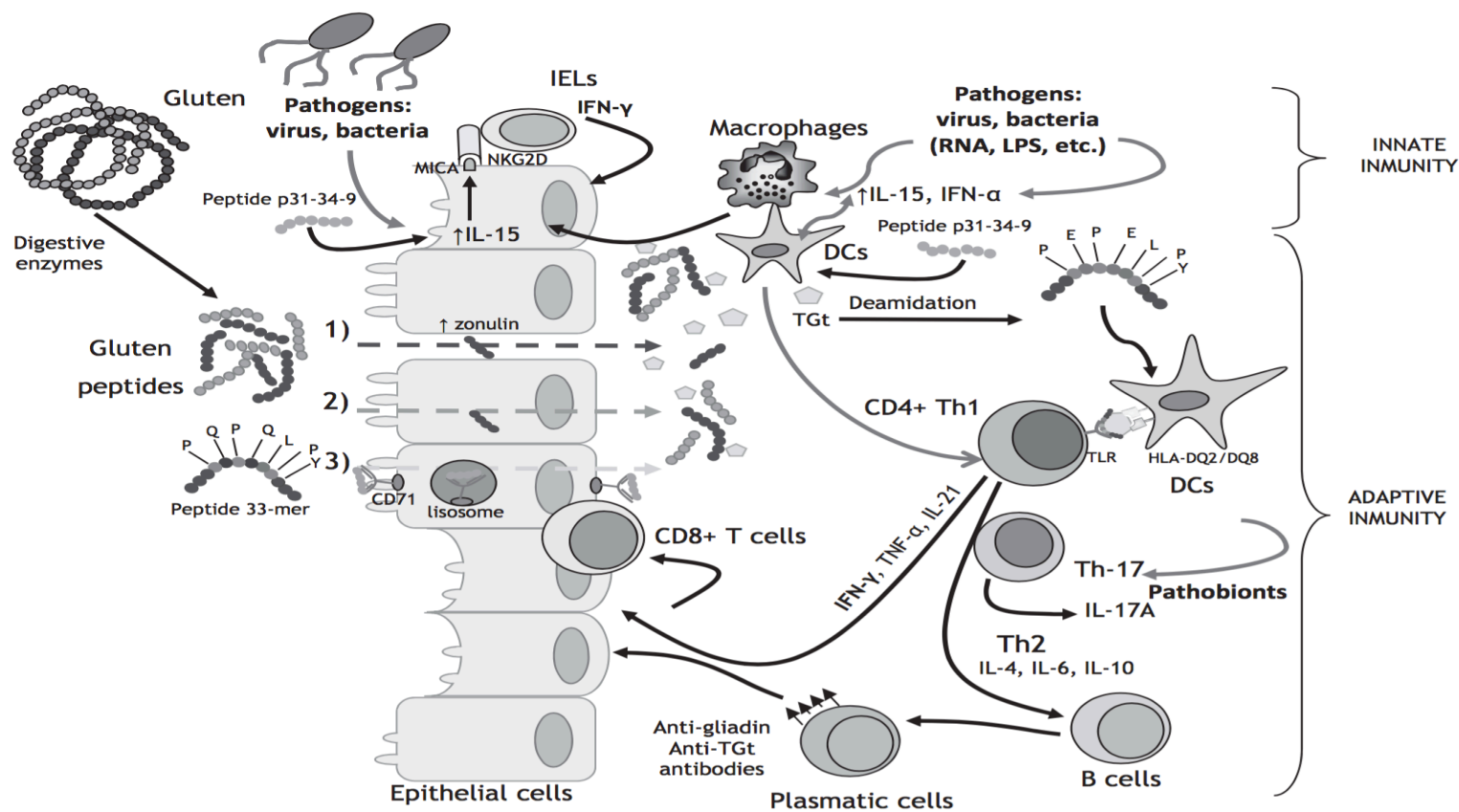
Increased IL-15 is thought to play a pivotal role in the destruction of intestinal epithelial cells (Escudero-Hernandez et al., 2016). Once in the lamina propria, gluten peptides are deamidated by the enzyme transglutaminase 2 (TG2). IL-15 activated dendritic cells endocytose deamidated gluten peptides, presenting them to CD4<sup>+</sup> T-cells via their HLA-DQ molecules. CD4<sup>+</sup> T-cells become gluten reactive, releasing Th1 cytokines; IFN $\gamma$ , TNF- $\alpha$ , IL-18 and IL-21 (Escudero-Hernandez et al., 2016). These inflammatory cytokines promote the expression of stress molecules by intestinal epithelial cells and their receptors/ligands in IELs causing epithelial cell death, therefore increasing intestinal permeability (Lauret and Rodrigo, 2013, Escudero-Hernandez et al., 2016) (Figure 1.3).

Whereas the role of the adaptive immune system in CD is well described, the role of the innate immune system is less well understood (Junker et al., 2012). In recent years it has been reported that the non-gluten alpha-amylase/trypsin inhibitors (ATI) found in wheat and related cereals are important players in innate immune system activation not only in CD, but perhaps also in other intestinal inflammatory disorders (Junker et al., 2012). Wheat contains the following proteins; albumins, globulins, gliadins and glutenins. Commercial wheat consists of 10%-20% albumins/globulins and 80%-90% gluten. Albumins/globulins are composed of 14 protein families, including ATIs, which are water soluble  $\approx$ 15-kD proteins which are active in defending cereals from pests (Tilg et al., 2013, Schuppan and Zavallos, 2015, Zavallos et al., 2017). In humans, ATIs can activate macrophages and dendritic cells through Toll-like receptor 4 (TLR4), causing the release of proinflammatory cytokines, and thus initiating an inflammatory response (Junker et al., 2012, Tilg et al., 2013).

Interestingly, a recent study found that gluten-containing foods had levels of TLR4 activating ATIs which were 100 fold higher than in gluten free foods. Moreover, ATIs were more abundant in modern varieties of wheat and were also highly resistant to proteolysis and retained their activity in processed or baked foods (Zavallos et al., 2017). This discovery may have implications for patients with irritable bowel syndrome (IBS) or even IBD patients, who do not have evidence of CD but improve on a GFD. Up to now there has been no scientific evidence to support the GFD in these patients, but it might be possible that their symptoms are driven by ATIs (Tilg et al., 2013, Herfarth et al., 2014, Zanwar et al., 2016).

Nevertheless, it is not clear whether the innate immune response precedes, is the result of, or occurs at the same time as, the adaptive immune response. However, it is clear that HLA-DQ molecules, TG2 enzyme, zonulin and IL15 are among the key players in the pathogenesis of CD (Fasano, 2012, Escudero-Hernandez et al., 2016).

**Figure 1.3 Proposed model for CD pathogenesis** TGt = Tissue transglutaminase; IEL = intraepithelial lymphocytes; NKG2D = natural killer group 2D receptor; MICA = MHC class I chain-related protein A; CD71 = cluster of differentiation 71 or transferrin receptor protein 1; DCs = dendritic cells (Sanz, 2015).



## **1.5 CD and co-existent autoimmune conditions; the protective role of the GFD.**

The co-existence of CD and other immune-mediated conditions (IMCs), especially T1DM and ATD, is well reported in the literature (Fasano, 2006, Neuhausen et al., 2008, Lundin and Wijmenga, 2015). Shared genetics partially explains the co-existence of CD with other IMCs (Figure 1.4). However, whether untreated CD precipitates the development of other IMCs in genetically susceptible individuals is still controversial. A few studies have tried to answer this question, however, these have resulted in contradictory findings, probably due to their cross-sectional retrospective nature, small sample sizes and different methodology (Table 1.1).

In 1999 Ventura *et al.* reported that the age at diagnosis with CD was the only predictor of co-existent IMC in a cohort of 909 CD patients. Using age at diagnosis as a measure of gluten exposure, they concluded that the co-existence of CD with other IMCs is linked to the duration of gluten exposure (Ventura et al., 1999). Previous to this study, another found a higher prevalence of anti-insulin antibodies in untreated compared to treated CD children (27% and 20% respectively) (Di Mario et al., 1992). Similarly, Toscano *et al.* (2000) found a higher prevalence of anti thyroid peroxidase antibodies in untreated versus treated CD adolescents (53% vs. 20%) (Toscano et al., 2000).

However, a posterior study by Sategna-Guidetti *et al.* (2001) still reported an association between increased age at diagnosis and co-existent IMC, but no association with gluten exposure measured as diet compliance, follow-up with serology and biopsy, and age at diagnosis with the co-existent IMC (Sategna Guidetti et al., 2001). Ventura's findings were subsequently confirmed by a study in Finland, which again reported a higher prevalence of IMCs in CD patients compared to controls, however gluten exposure was not associated with this increased risk (Viljamaa et al., 2005). Cosnes *et al.* (2008) reported that those diagnosed with CD earlier in life and with a family history of CD were more likely to develop other IMCs. They also found the cumulative risk of developing a co-existent IMC is higher in patients who were non-compliant with the GFD (Cosnes et al., 2008). In 2012, an Italian study confirmed that paediatric CD patients were more prone to develop other IMCs (Elli et al., 2012). However, they did not find an association between gluten exposure and the development of IMCs, and even found a protective effect of long term gluten exposure on the development of co-existent IMCs, which they speculated could be due to sequestration of reactive T lymphocytes in affected tissues (Elli et al., 2012).

All the above studies were retrospective in nature which makes assessing compliance with the GFD and gluten exposure difficult. They have measured gluten exposure retrospectively as the time elapsed until adoption of a GFD, or until developing a second IMC if this happens before CD diagnosis. This approach can lead to an incorrect estimation of gluten exposure as CD or other IMCs may have been present and undiagnosed for years (Viljamaa et al., 2005). Two studies have followed newly diagnosed CD patients prospectively to assess the relationship of IMCs to the GFD. The first looked specifically for the development of ATD in a sample of 27 CD patients followed for a year after diagnosis. No relationship was found between gluten exposure and ATD development (Metso et al., 2012). The second looked at the occurrence of Th1, Th2 and Th17 driven IMCs in a sample of 1255 newly diagnosed CD patients followed for a 5 year period (Table 1.1). Co-existent IMCs increased during the follow-up period despite the GFD (Imperatore et al., 2016).

Despite their partially prospective nature, the above studies had a retrospective component as establishing the prevalence of IMCs preceding CD diagnosis had to be done retrospectively. Moreover, CD could have been undiagnosed in these patients for years. It is also possible that longer follow-up periods are necessary to observe a benefit of the GFD. Finally, the study by Imperatore *et al.* assesses compliance with the diet using tTGA serology as a surrogate marker, and others have suggested that this method has significant limitations and is not an accurate marker of GFD compliance (Vahedi et al., 2003).

Ideally, in order to figure out whether the development of co-existent IMCs in CD patients can be prevented, large prospective observational studies would be desirable both in genetically predisposed children and adults, followed for long periods of time collecting clinical, dietary and immunological information. The Environmental Determinants of Diabetes in the Young (TEDDY) study is an international prospective study following children predisposed to T1DM and CD over 15 years with the aim of identifying environmental factors related to the development of both conditions. A nested case-control study from the TEDDY cohort following infants up to two years of age resulted in the finding that the amount of gluten consumed until two years of age increases the risk of CD in genetically predisposed children (Andren Aronsson et al., 2016b). Whether this is also true in genetically predisposed adults and other IMCs is not known.



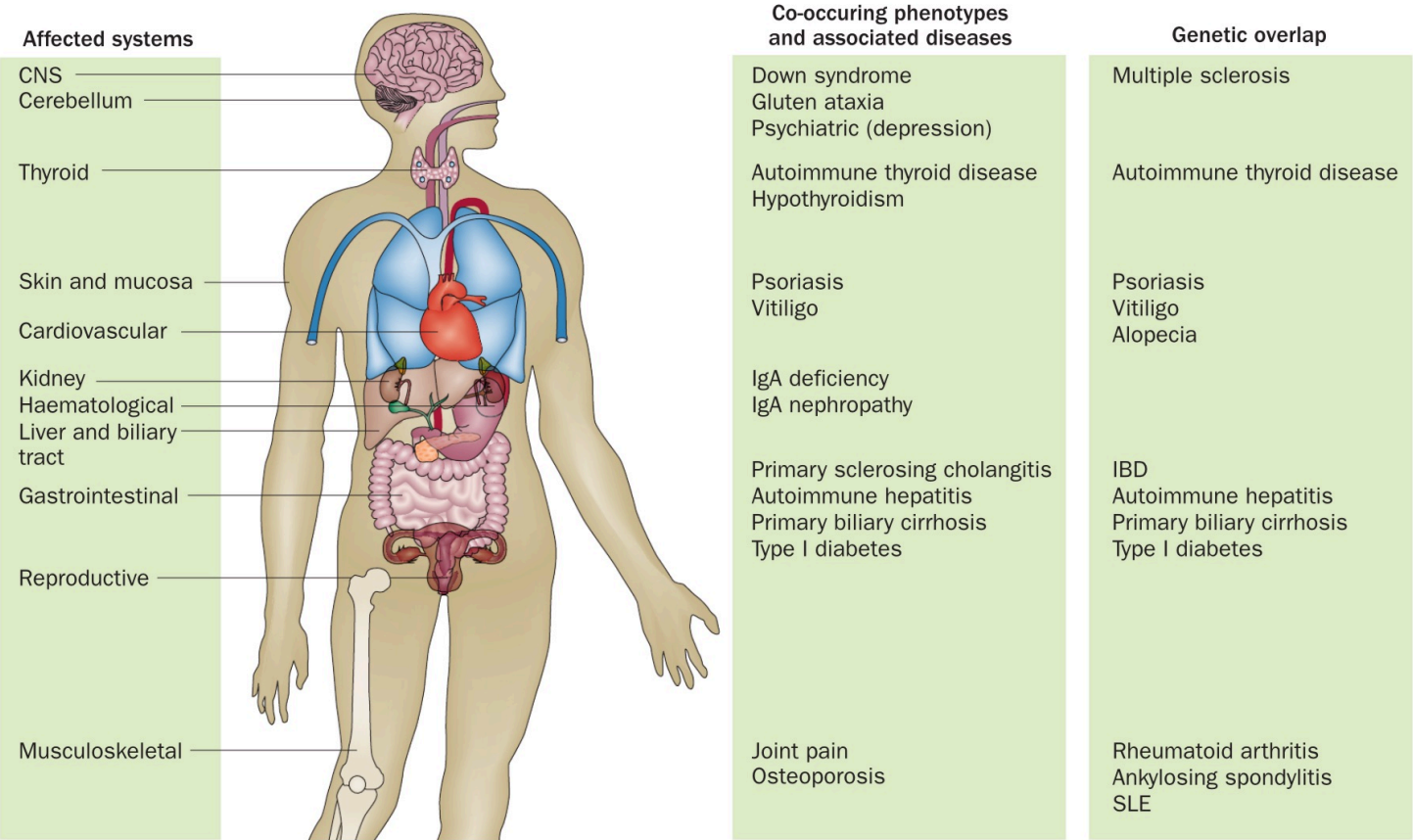
**Table 1.1 Studies assessing the risk to develop co-existent IMCs in CD patients according to gluten exposure.**

| Study                                 | Sample (n) | Sample stratification                                   | Age of Sample (years)                                                                                        | Parameters for CD diagnosis                                                                              | Tool to assess presence of comorbidities                                                                                                                                                                                                                                                                    |
|---------------------------------------|------------|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ventura <i>et al.</i> (1999)          | 2340       | 909 (CD)<br>1268 (Controls)<br>163 (Chron's disease)    | CD patients mean age (16.1+/- 3.1)<br>Controls mean age (20.5+/-4.5)<br>Crohn's disease mean age (28.8+/-10) | ESPGHAN criteria<br>All patients had several or subtotal villous atrophy and crypt hyperplasia on biopsy | Structured questionnaire plus review of patients' charts                                                                                                                                                                                                                                                    |
| Sategna-Guidetti <i>et al.</i> (2001) | 1027       | 422 (CD)<br>605 (controls)                              | CD patients (16-84)<br>Controls (16-84)                                                                      | Duodenal biopsy                                                                                          | Structured questionnaire plus review of patients' charts                                                                                                                                                                                                                                                    |
| Viljamaa <i>et al.</i> (2005)         | 1002       | 703 (coeliac disease)<br>299 (controls)                 | CD patients mean age 46.5 (2-88)                                                                             | Biopsy-proven CD patients                                                                                | Structured questionnaire plus review of patients' charts                                                                                                                                                                                                                                                    |
| Cosnes <i>et al.</i> (2008)           | 924        | 378 (CD paediatric patients)<br>546 (CD adult patients) | Not specified                                                                                                | Duodenal biopsy showing increased IEL's, crypt hypertrophy and villous atrophy                           | Structured questionnaire plus review of patients' charts                                                                                                                                                                                                                                                    |
| Ouaka <i>et al.</i> (2008)            | 64         | 64 (CD patients)                                        | 29 (16-63)                                                                                                   | Histology according to Marsh criteria                                                                    | Review of patients records                                                                                                                                                                                                                                                                                  |
| Meloni <i>et al.</i> (2009)           | 8364       | 324 (CD patients)<br>8040 (controls)                    | CD (6.6, +/- 4.5)<br>Controls (6-15)                                                                         | IgA antiendomysial antibody positivity<br>Duodenal biopsy Marsh criteria<br>Reponse to GFD               | Thyroid peroxidase antibody (AbTPO) & thyroid globulin antibody (AbTG) at the onset of CD and once per year after the institution of the GFD. When abnormal values were detected serum-free triiodo- thyronine (FT3), free thyroxine (FT4), thyroid-stimulating hormone (TSH), and thyroid ultrasonography. |
| Metso <i>et al.</i> (2012)            | 54         | 27 (CD patients)<br>27 (Controls)                       | CD patients median age 44 years<br>Controls median age 51 years                                              | Biopsy proven CD                                                                                         | Free thyroxine (FT4), plasma levels of thyrotropin (TSH) and thyroid volume measured at baseline and a year later at follow-up                                                                                                                                                                              |
| Elli <i>et al.</i> (2012)             | 4494       | 1015 (CD patients)<br>3479 (controls)                   | CD patients median age 35 (13-85)                                                                            | Positive tTGA & Marsh 3 biopsy                                                                           | Retrospective review of patients records (yearly follow-up)                                                                                                                                                                                                                                                 |
| Imperatore <i>et al.</i> (2016)       | 1255       | 1255 (CD patients)                                      | CD patients mean age 28.1 (+/- 15.7)                                                                         | Positive tTGA & EMA & Marsh biopsy $\geq 1$<br>HLA genotyping fro Marsh 0 & 1                            | Information on IMCs preceding CD gathered at first consultation, patients posteriorly underwent clinical examination and serological tests once a year                                                                                                                                                      |

**Table 1.1 (cont.)**

| CD Status                                                                       | Tool to assess compliance with GFD                                                                  | Type of study                                                                                                              | Results                                                                                                                                                                                                                                             |
|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| All CD patients on GFD at recruitment                                           | EMA antibodies                                                                                      | Cross-sectional retrospective                                                                                              | Prevalence of co-existent IMCs in coeliac disease is related to the duration of gluten exposure measured as age at diagnosis with CD.                                                                                                               |
| All CD patients on GFD for at least one year on recruitment                     | Review with expert or follow-up biopsy and/or EMA antibodies                                        | Cross-sectional retrospective                                                                                              | Prevalence of co-existent IMCs in coeliac disease is NOT related to the duration of gluten exposure (diet compliance, follow-up and age at diagnosis with co-existent IMC), but is related to age at diagnosis with CD.                             |
| All CD patients on GFD                                                          | Retrospectively assessed from patients' charts and questionnaires, follow-up serology and/or biopsy | Cross-sectional retrospective                                                                                              | Prevalence of co-existent IMCs in coeliac disease is NOT related to the duration of gluten exposure (diet compliance, follow-up and age at diagnosis with co-existent IMC), but is related to age at diagnosis with CD.                             |
| All coeliac patients on GFD at recruitment                                      | Assessed retrospectively by physician                                                               | Cross-sectional retrospective                                                                                              | Coeliac patients most at risk of autoimmune disease are those diagnosed early in life and having a family history of autoimmunity. Cumulative risk of developing co-existent IMCs higher in patients non-compliant with the GFD. GFD is protective. |
| All coeliac patients on GFD                                                     | Assessed by physician and dietician at diagnosis and follow-up                                      | Cross-sectional retrospective                                                                                              | Prevalence of autoimmune conditions does not correlate with gluten intake.                                                                                                                                                                          |
| All patients on GFD                                                             | Assessed by serology, AGA, tTG and EMA every 12 months                                              | Retrospective (mean follow-up period 8 years)                                                                              | ATD is strongly associated with CD. Gluten exposure does not correlate with the development of autoimmune thyroiditis.                                                                                                                              |
| Newly diagnosed CD patients at baseline on GFD a year later at follow-up        | Assessed by experienced dietician serology and follow-up biopsy                                     | Prospective (follow-up a year after CD diagnosis). Retrospective to check for presence of T1DM and ATD before CD diagnosis | ATD is strongly associated with CD. Gluten exposure does not correlate with the development of autoimmune thyroiditis.                                                                                                                              |
| All patients on GFD                                                             | interview with expert physician and serology test                                                   | Retrospective (follow-up 0 to 46 years)                                                                                    | Paediatric CD patients more likely to develop IMCs. Gluten exposure not associated to development of IMCs. Long-term gluten exposure protective against co-existent IMCs.                                                                           |
| Newly diagnosed CD patients at enrolment on GFD for the 5 year follow-up period | tTGA levels used as surrogate marker of compliance with GFD                                         | Prospective (follow-up 5 years after diagnosis). Retrospective to check for presence of IMCs before CD diagnosis           | The prevalence of co-existent IMCs was high at the time of diagnosis & increased during the follow-up period despite the GFD.                                                                                                                       |

Figure 1.4 Organs involved in CD co-existent with other IMCs and genetic overlap between CD and other IMCs (Lundin and Wijmenga, 2015).



## **1.6 CD serology and biopsy: past, present and future.**

Serology tests for CD have notably progressed in the past 20 years. Initially, antibodies against cereal derived peptides were used, as in the case of anti-gliadin antibodies (AGA) (Oxentenko and Murray, 2015). However, whereas these antibodies work well for patients where there is high suspicion of CD, in general they have low sensitivity and specificity (Rostom et al., 2005, Rubio-Tapia et al., 2013). Sensitivity and specificity for both IgA and IgG AGA is estimated to be approximately between 80% and 90% and lower specificities are seen in adults compared to children (Leffler and Schuppan, 2010). Despite these drawbacks, recent studies indicate that AGA levels correlate well with the presence of gluten ataxia and non-coeliac gluten sensitivity (NCGS); however, this is a controversial area and further research on their potential use in diagnosing NCGS is needed (Mansueto et al., 2014, Hadjivassiliou et al., 2016, Kurppa et al., 2016).

Therefore, AGA were soon replaced in the diagnosis of CD by autoantibodies first against reticulin and then endomysium antibodies (EMA) in 1989. EMA sensitivity and specificity was found to be largely above 90%, representing an improvement over AGA (Leffler and Schuppan, 2010). Despite EMA being a better diagnostic tool, the immunofluorescence assay they are based on required that each individual sample had to be checked under a microscope, making them difficult to standardise (Leffler and Schuppan, 2010).

The introduction of tTGA in the period 1998-2000 solved the problem of standardisation of EMA because it allowed for the development of an ELISA-based test. Due to the high sensitivity and specificity of IgA tTGA (approximately 98%) it is the recommended serological marker for the diagnosis of CD (Dieterich et al., 1998). tTGA levels should be ordered upon suspicion of CD together with endomysial antibodies (EMA) and total IgA levels to detect potential IgA deficiency which is common in CD patients and would lead to a false negative (Cataldo et al., 1998, Leffler and Schuppan, 2010, Chow et al., 2012).

Active CD patients also have circulating antibodies against deamidated gliadin peptides (DGPA) (Schyum and Rumessen, 2013). Unlike AGA, which measure antibodies against unmodified gliadin peptides, DGPA perform as well as tTGA. Moreover, some DGPA assays measure IgA and IgG antibodies at the same time increasing their sensitivity to diagnose CD (Schyum and Rumessen, 2013). DGPA antibodies have also been shown to be superior at detecting CD in its early stages generally and in infants in

particular, although further studies are needed to confirm these findings (Liu et al., 2007, Kurppa et al., 2011, Kurppa et al., 2016). Assays combining DGPA antibodies and tTGAs have also shown better accuracy (Schyum and Rumessen, 2013). In the case of IgA deficient patients, it is preferable to use IgG DGPA than IgG tTGA, as the latter has very low sensitivity (Dahlbom et al., 2005, Absah et al., 2017). Therefore, IgA tTGA testing together with total IgA would be the initial serological test of preference, and this is complemented with IgG DGPA in IgA deficient individuals (Figure 1.5) (Leffler and Schuppan, 2010).

**Figure 1.5 Suggested algorithm for CD testing.**

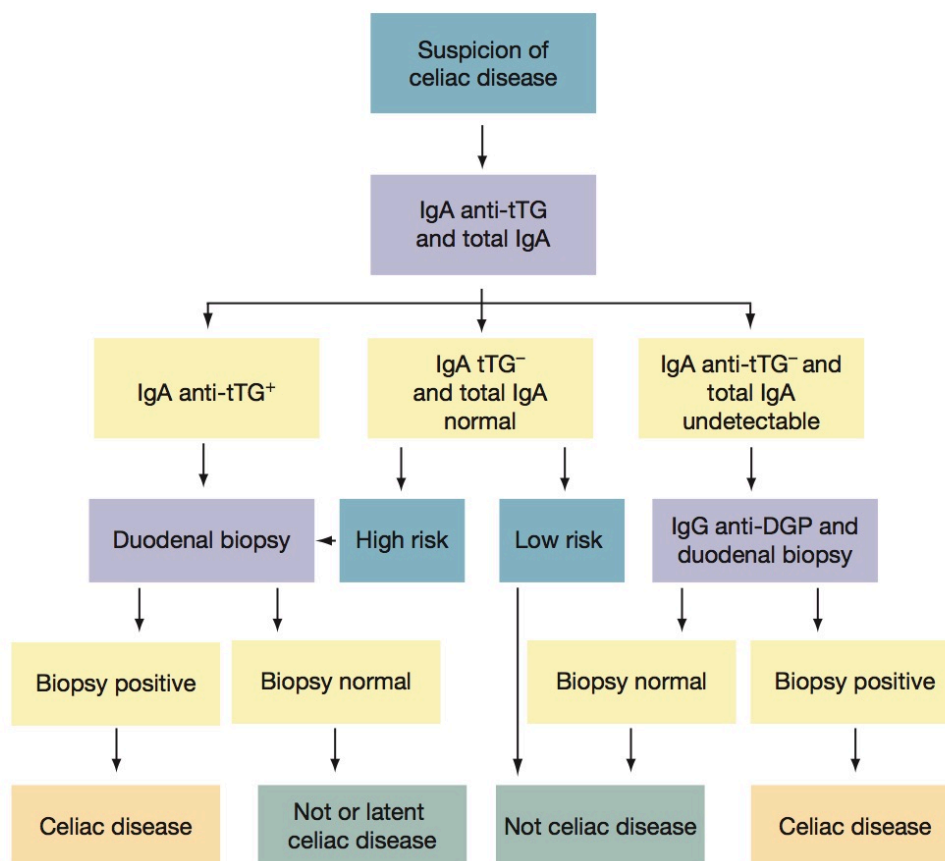


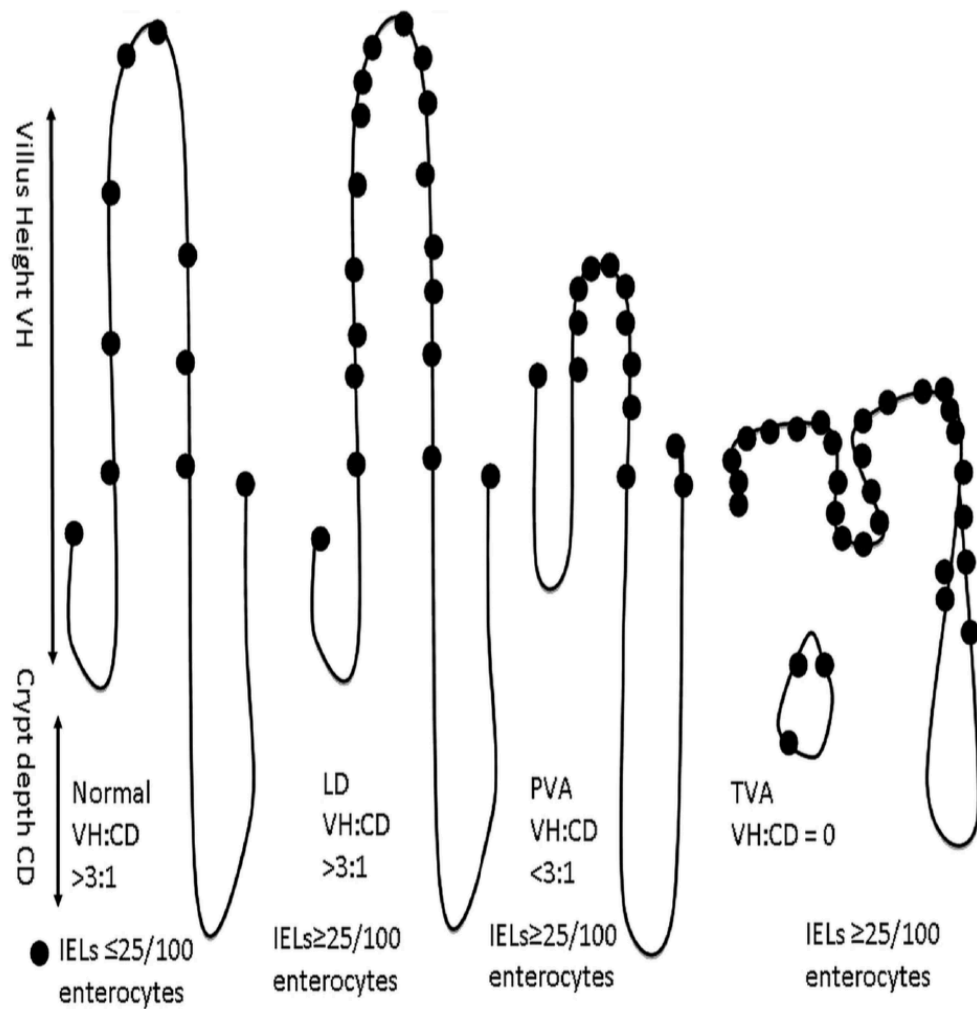
Figure 1.5 above taken from Leffler *et al.* (2010) shows how serological testing marks the inflection point in the investigation of CD. Therefore, if a person is tTGA negative and has normal total IgA, but they are high risk because they have a strong family history or are HLA-DQ2/DQ8 positive, for example, they still should be considered for a biopsy as the presence of CD is still possible. On the contrary, if the same serological result is available for a patient with low risk (i.e. no family history and not HLA-DQ2/DQ8 positive), it can be assumed they do not have CD. Latent CD is diagnosed when there is positive tTGA but normal histology in a high risk scenario, this is to say in the presence of strong family history or HLA-DQ2/DQ8 positivity (Figure 1.5), in this case there is no treatment recommended, but follow-up is necessary.

Biopsies are not only a central component of CD diagnosis but also in monitoring recovery after diagnosis. Upon diagnosis, an upper intestinal biopsy is carried out taking at least one sample from the bulb and at least four samples from the second or third part of the duodenum (Moreno et al., 2017). A CD patient biopsy has one or more of the following characteristics; increased IELs, crypt elongation and partial to total villous atrophy (Figure 1.6) (Ludvigsson et al., 2014b). The original Marsh classification used to describe the CD biopsy spectrum was modified by Oberhuber and later simplified by Corazza and Villanaci to make it less subjective (Marsh, 1992, Oberhuber et al., 1999, Corazza and Villanacci, 2005). Rostami and Villanaci introduced the concept of microscopic enteritis, also called lymphocytic duodenosis or lymphocytic enteropathy, which allows diagnosis of patients with non-classical presentation at an earlier stage as well as the confident initiation of the GFD in symptomatic patients with very mild histological damage (Rostami and Villanacci, 2009).

In 2012, new guidelines from the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) allowed the biopsy to be omitted in children with CD symptoms if they were HLA-DQ2/DQ8 positive and their tTGA levels were ten or more times the normal upper limit (Husby et al., 2012). These new paediatric guidelines were established to avoid the use of general anaesthesia and its potential complications (Kurppa et al., 2016). Despite increasing debate as to whether a biopsy is needed in adults due to the good correlation of high tTGA levels with small mucosal damage, it still remains the gold-standard for CD diagnosis in this group (Hill and Holmes, 2008, Hopper et al., 2008, Taavela et al., 2013, Singh et al., 2014). However, research into less expensive and invasive diagnostic options, as well as to improve the current biopsy methods is on-going (Salmi et al., 2010, Catassi and Fasano, 2010, Wakim-Fleming et al., 2013, Ludvigsson et al., 2014a, Kurppa et al., 2016, Haines et al., 2008). The role of serology testing is pivotal in the progress towards less invasive methods. In the area of novel serological biomarkers at diagnosis several options have been considered some of which are summarised in Table 1.2. Furthermore, new serological biomarkers are needed to allow monitoring of small bowel mucosa recovery and adherence to the GFD because current methods are not predictors of mucosal recovery or GFD adherence (Hopper et al., 2008, Sharkey et al., 2013, Kurppa et al., 2016).

**Figure 1.6 Characteristics of CD biopsies.**

**CD**, crypt depth; **LD**, lymphocytic duodenosis; **PVA**, partial villous atrophy; **TVA**, total villous atrophy; **VH**, villous height; dots representing IELs. Figure taken from Ludwigsson et al. (2014).



**Table 1.2 Novel serological biomarkers for the diagnosis of CD.**

| <b>Biomarker</b>                                                                                     | <b>Scientific Evidence</b>                                                      | <b>Drawbacks</b>                                       | <b>References</b>                                                                                                                        |
|------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| <b>TG2 open form</b>                                                                                 | Better accuracy at diagnosis & potentially identifying not compliance with GFD. | Additional studies with larger cohorts needed          | Pinkas <i>et al.</i> , 2007<br>Lindfors <i>et al.</i> , 2011<br>Pallav <i>et al.</i> , 2012                                              |
| <b>ELISA tests with CD specific TG2 epitopes</b>                                                     | Differentiate between CD autoimmunity and other autoimmune conditions           | Additional studies needed                              | Simon-Vecsei <i>et al.</i> , 2012<br>Kurppa <i>et al.</i> , 2016                                                                         |
| <b>Assays combining IgA &amp; IgG DGP antibodies or combinig DGP antibodies &amp; TG2 antibodies</b> | Better accuracy at diagnosis.                                                   | Performance in clinical routine to be studied          | Schyum and Rumessen, 2013<br>Kurppa <i>et al.</i> , 2016                                                                                 |
| <b>TG3 antibodies for DH &amp; TG6 antibodies for gluten ataxia</b>                                  | High specificity for both conditions.                                           | Not yet widely introduced in routine clinical practice | Sardy <i>et al.</i> , 2002<br>Reunala <i>et al.</i> , 2015<br>Hadjivassiliou <i>et al.</i> , 2008<br>Hadjivassiliou <i>et al.</i> , 2013 |
| <b>Intestinal Fatty Acid Binding Protein (I-FABP)</b>                                                | I-FABP correlates with the severity of the histology at diagnosis.              | Not specific for CD                                    | Adriaanse <i>et al.</i> , 2013<br>Adriaanse & Leffler, 2015<br>Adriaanse <i>et al.</i> , 2016                                            |

Some serological biomarkers aimed at improving CD diagnosis are listed in Table 1.2. Other potential serological biomarkers for CD diagnosis are; simvastatin, cholesterol lowering drug which is present at high levels in the blood of untreated CD patients; gliadin reactive CD4+ T-cells; specific microbiome composition, microRNAs profiles and novel peptides.



## **1.7 Novel serological biomarkers to monitor CD.**

As mentioned, CD antibody tests are poor predictors of histological recovery or as a means of monitoring dietary compliance (Vecsei et al., 2014, Kurppa et al., 2016, Moreno et al., 2017). Due to the lack of effective biomarkers, the process to examine compliance with the diet and mucosal recovery currently consists of, monitoring the resolution of the clinical symptoms present at diagnosis or the appearance of new symptoms, the use of validated surveys and dietary interviews, follow-up biopsies and serology biomarkers (Ludvigsson et al., 2014a).

Whereas the previous four step process is the most comprehensive approach to assess compliance and mucosal recovery in CD patients, it is not exempt from flaws and drawbacks. For example, although it has been shown that intestinal bowel symptoms are common in CD patients not adhering to the GFD, monitoring the resolution of symptoms is of no use in patients who presented with no or minimal symptoms at diagnosis. Moreover, symptoms may be present despite absence of gluten ingestion (Lahdeaho et al., 2011, Sainsbury et al., 2013, Sharkey et al., 2013). Some validated surveys such as the celiac dietary adherence test (CDAT) have been shown to be superior to tTGAs in monitoring adherence (Leffler et al., 2009). However, potential disadvantages to the use of surveys and diet interviews are their subjective nature and the propensity of patients to have biased recall (Moreno et al., 2017). A recent paper found no correlation between gluten immunogenic peptides (GIPs) in stools and dietary questionnaires to measure compliance in CD (Comino et al., 2016). GIP testing is a promising tool to prevent symptoms in CD patients due to inadvertent gluten exposure (Ludvigsson et al., 2018).

A follow-up biopsy is currently the best available method to assess both compliance with the GFD and mucosal recovery. Morphological changes, such as the ratio of the villous height to crypt depth, have been shown to correlate with the presence of GI symptoms and quality of life of CD patients on a GFD (Taavela et al., 2013, Ludvigsson et al., 2014a). However, despite its merits, biopsy is an expensive, invasive and impractical approach for frequent monitoring. Moreover, there is no consensus on the need for, and timing of follow-up biopsies (Ludvigsson et al., 2014a). Finally, as previously mentioned, CD antibody tests are poor predictors of compliance or mucosal recovery, probably because rather than measure intestinal damage, they detect immunological response, the measured components of which may have long half-lives. After commencing the GFD it can take 6-24 months for serological biomarkers to

normalise (Kaukinen et al., 2002, Tursi et al., 2003, Kaukinen et al., 2007, Sharkey et al., 2013, Rashid and Lee, 2016, Moreno et al., 2017).

Therefore, there is a need for standardised, reliable and non-invasive biomarkers, that indicate compliance with the GFD and mucosal recovery and can be used both in the clinical and research setting. Several monitoring biomarkers have been suggested and are summarised in Table 1.3 (Moreno et al., 2017).

**Table 1.3 Novel serological biomarkers to monitor CD.**

| <i>Method</i>                                                             | <i>Scientific Evidence</i>                                                                                                                                                                                                                                                                          | <i>Drawbacks</i>                                                                                                                              | <i>References</i>                                                                                                               |
|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| <b>Permeability test (lactulose/mannitol)</b>                             | Intestinal permeability proven to correlate gluten ingestion in CD patients.                                                                                                                                                                                                                        | Not specific for CD                                                                                                                           | Catassi <i>et al.</i> , 1993<br>Duerksen <i>et al.</i> , 2005<br>Vecsei <i>et al.</i> , 2009                                    |
| <b>Fecal calprotectin (FC)</b>                                            | Proven to detect bowel inflammation.<br>Some studies found FC to be increased in CD patients with active disease compared to controls and CD patients following a GFD.                                                                                                                              | Controversial findings in CD.<br>Not specific for CD.                                                                                         | Tibble <i>et al.</i> , 2001<br>Montalto <i>et al.</i> , 2007<br>Ertekin <i>et al.</i> , 2010<br>Balamtekin <i>et al.</i> , 2012 |
| <b>Regenerating Gene Ia (REG Ia)</b>                                      | REG Ia is a protein increased in the target tissue of some autoimmune diseases (i.e T1DM and IBD).<br>REG Ia found to be increased in sera of CD patients and not in sera of controls, T1DM & IBD patients.                                                                                         | Further studies needed.<br>Not known if REG Ia elevated in sera of patients with other autoimmune conditions.                                 | Taavela <i>et al.</i> , 2013<br>Ludwigsson <i>et al.</i> , 2014<br>Moreno <i>et al.</i> , 2017                                  |
| <b>Serum Alkylresorcinols (AR)</b>                                        | Phenolic lipids found almost exclusively in whole-grain wheat and rye products.<br>Found to be present at higher concentrations in untreated CD patients compared to controls and treated CD patients.                                                                                              | Only traces found in refined products.<br>Not known if they would detect occasional gluten intake.                                            | Kyro <i>et al.</i> , 2014<br>Lind <i>et al.</i> , 2016<br>Choung <i>et al.</i> , 2017                                           |
| <b>Citrulline</b>                                                         | Non-essential aminoacid almost exclusively present in the enterocytes of the small intestine.<br>Some studies have found and inverse correlation between serum citrulline and intestinal mucosal damage.                                                                                            | Controversial findings in CD.<br>Not specific for CD.<br>Negative correlation only existing in the presence of significant intestinal damage. | Miceli <i>et al.</i> , 2008<br>Ioannou <i>et al.</i> , 2011<br>Adriaanse & Leffler, 2015                                        |
| <b>Intestinal Fatty Acid Binding Protein (I-FABP)</b>                     | I-FABP is an intestinal enterocyte cytosolic protein involved in the metabolism of fatty acids.<br>Correlation between serum I-FABP and histological damage on patients on the GFD.<br>After CD diagnosis serum I-FABP decrease fast in children.<br>Early response marker during gluten challenge. | Serum levels decrease fast in children (6 months) so better than tTGA, but not in adult CD (3-4 years).<br>Not specific for CD.               | Adriaanse <i>et al.</i> , 2013<br>Adriaanse & Leffler, 2015<br>Adriaanse <i>et al.</i> , 2016                                   |
| <b>Detection of gluten immunogenic peptides (GIPs) in feces and urine</b> | CD triggered by GIPs which are resistant to gastrointestinal digestion, these peptides can be detected in human feces using ELISA.<br>A proportional fraction of the GIPs absorbed in the GI tract makes it to the circulation and is excreted in urine.                                            | Further research needed                                                                                                                       | Moreno <i>et al.</i> , 2015<br>Comino <i>et al.</i> , 2016                                                                      |

## 1.8 Alternatives to the gluten free diet

The only available treatment for CD is a lifelong strict adherence to a GFD (Ludvigsson et al., 2014a). A balanced GFD should be constituted by a combination of naturally gluten-free foods and certified artificially processed gluten-free foods (Freeman, 2017). Therefore, patients are guided to consume naturally gluten free cereals such as rice, buckwheat, corn and millet; and non-processed foods such as meat, fish, eggs, vegetables and fruits. Finally, artificially processed gluten-free foods, as established by Codex Alimentarius containing  $\leq 20$  ppm (mg/kg) gluten, complete the recommended CD diet (Codex, 2008).

Following a GFD strictly is very challenging, especially in Western countries where wheat, rye and barley are staples. Compliance levels vary a lot according to factors such as the patients age and age at CD diagnosis (Hall et al., 2009, Wild et al., 2010). Moreover, gluten is quite ubiquitous as it is commonly used as a food additive and cross-contamination of inherently gluten free foods may occur. In fact, foods labelled as gluten free may exceed the limit established by Codex Alimentarius (Codex, 2008, Thompson et al., 2010, Diaz-Amigo and Popping, 2012, Lee et al., 2014, Farage et al., 2017). Large variability in the methods utilised to detect gluten levels in gluten free foods are also a problem yielding inconsistent results, therefore new methodologies to quantify gluten in food is an active area of research (Diaz-Amigo and Popping, 2012, Farage et al., 2017).

Hence, it is practically impossible to achieve a diet with absolutely no gluten intake. Safe gluten intake levels for CD patients are not clear, but should lie between 10 and 100mg daily, indicating that some patients need only a small amount such as 10mg to develop small mucosa damage (Hischenhuber et al., 2006, Akobeng and Thomas, 2008). Moreover, other disadvantages to the GFD are its cost and the limited nutritional value of many commercial purified gluten free products (Penagini et al., 2013, Barone et al., 2016). A study by Wild *et al.* (2010) reported high intake of sugar accompanied by low intake of fibre, calcium and iron, among other minerals, in a cohort of patients on a GFD (Wild et al., 2010). Therefore, novel non-dietary therapies are desirable and are an area of active research (Maki, 2014, Castillo et al., 2015, Kurada et al., 2016). Better understanding of CD pathogenesis has led to different non-dietary strategies targeting different points in the molecular cascade that leads to its development (Lindfors et al., 2012, Kurada et al., 2016). These different therapies are summarised in Figure 1.7 taken from the 2012 paper by Linfors *et al.*

A series of compounds are currently in phase II clinical trial including glutenase enzymes and tight junction modulators (Vanga and Kelly, 2014). The logical basis for protease enzyme supplementation therapy is that incomplete digestion of dietary gluten leads to the appearance of immunogenic gluten peptides in the lumen of the small intestine. Therefore, enzyme supplementation could detoxify these immunogenic peptides avoiding their translocation into the lamina propria and the activation of the innate immune system (Lindfors et al., 2012, Vanga and Kelly, 2014).

Prolyl endopeptidases (PEP) are a family of enzymes with potential for the treatment of CD. Although these enzymes are expressed in the human small intestine mucosa, they cannot digest highly immunogenic 33-mer peptide (Shan et al., 2002, Garcia-Horsman et al., 2007). PEPs are also produced by several fungi and bacteria. The most promising type of PEP is the one produced by *Aspergillus niger* (AN-PEP) which has been shown to degrade gluten in an environment similar to that of the stomach and small intestine (Mitea et al., 2008). However, a clinical trial with a small group of CD patients on gluten containing diets and taking an AN-PEP supplement showed good tolerance to the enzyme, but no conclusive findings on the efficacy of AN-PEP, with studies including a larger number of patients and a longer period for the gluten challenge being needed (Tack et al., 2013). A study by Syage *et al.* found that lactiglutenase supplementation did show benefit in those CD patients who were seropositive, likely indicating that benefit from glutenase approach will really only occur or be seen in patients who are exposed to gluten (Syage *et al.*, 2017). On the other hand, commercial glutenase supplements currently available have been found to be ineffective in degrading immunogenic gluten peptides, and could represent a hazard to CD patients. Therefore, further studies and clinical trials are also required in this area (Janssen et al., 2015, Krishnareddy et al., 2017).

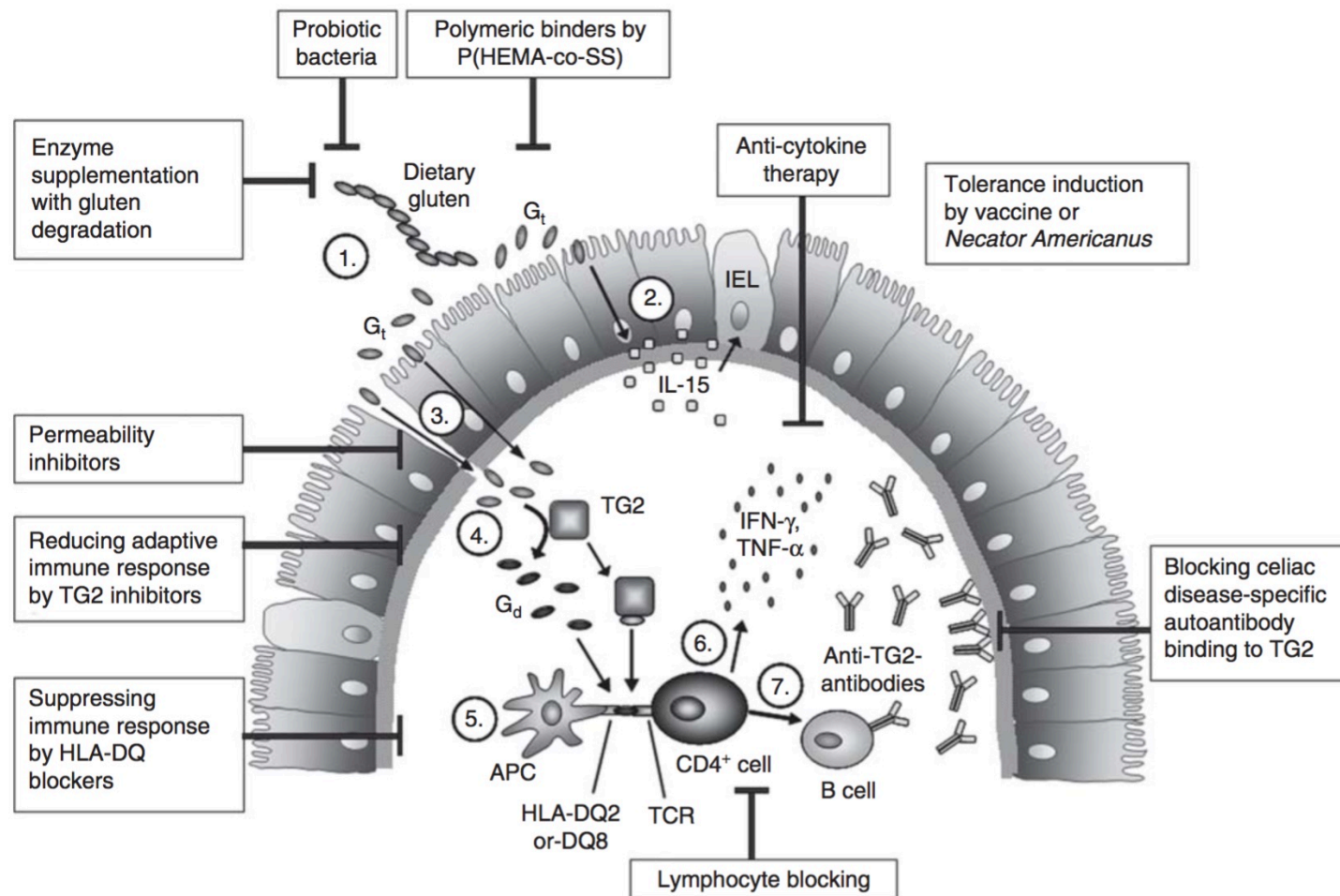
Intestinal permeability is a characteristic of active CD as discussed above, and seems to relate to the overexpression of the protein zonulin in CD patients in response to gliadin (Fasano et al., 2000, Lammers et al., 2008, Fasano, 2012). Therefore, zonulin inhibition is potentially an interesting non-dietary therapy for CD (Lindfors et al., 2012). Larazotide acetate (LA) is a synthetic octapeptide which shares aminoacid sequences with zonulin, and works by blocking zonulin receptors, thereby decreasing tight junction permeability. Phase I and II clinical trials have shown that LA is safe to be used and has a potential positive effect in preserving tight junctions and reducing paracellular permeability (Paterson et al., 2007, Gopalakrishnan et al., 2012).

Moreover, CD patients receiving LA reported less GI symptoms compared to controls, and it has also been proven to plateau the increase of tTGAs during gluten challenge

(Kelly et al., 2013, Leffler et al., 2015b). Despite these promising results certain areas remain unclear in the use of LA, such as an inverse dose effect, and the fact that although it stops the increase in tTGAs on gluten challenge, no difference in lactulose to mannitol ratios were observed between patients receiving the drug and placebos (Kelly et al., 2013, Khaleghi et al., 2016). This means that although the drug apparently decreases the immunological response, intestinal permeability is still present in CD patients treated with it. Thus, it could be that LA prevents the passage of gliadin peptides to the lamina propria via the paracellular pathway, but not the transcellular pathway (Khaleghi et al., 2016).

Other potential non-dietary therapies include the inhibition of TG2, blocking antigen presentation by HLA molecules, building up of tolerance to gluten or modulation of the immune system response and therapeutic vaccines (Vanga and Kelly, 2014, Goel *et al.*, 2017). Both glutenases and tight junctions regulators have the potential to be approved to treat NRCD or as a complement to the GFD (Kurada et al., 2016). However, before moving into phase III clinical trials, it is fundamental to develop accurate, sensitive, non-invasive biomarkers that indicate histological recovery and cessation of the autoimmune process. Biopsies are problematic in large scale studies and current serological biomarkers too insensitive to detect potential responses from non-dietary treatments (Kurppa et al., 2014).

Figure 1.7 CD pathogenesis and future non-dietary treatments (Lindfors et al., 2012)



## 1.9 CD Genetics

CD has a strong genetic component. In the 1970's it was discovered that certain HLA molecules were associated with the development of CD. The disease has been mainly linked to the alleles DQA1\*05 and DQB1\*02 which encode the HLA-DQ2 heterodimer, and alleles DQA1\*03 and DQB1\*0302 which encode the HLA-DQ8 heterodimer. Generally, between 90-95% of CD patients express HLA-DQ2, while the great majority of the remaining patients are HLA-DQ8 positive (Sollid, 2002). These HLA alleles are also frequently found in the general population (up to 40% of whom are carriers) suggesting that, though they are necessary for the development of the disease, they are not sufficient on their own to cause it. A so-called HLA DQ2 dose effect has been described in paediatric CD, with homozygosity for DQ2 conferring a higher risk to develop the condition as well as having an earlier onset (Liu et al., 2014).

The advent of array-based genotyping of single nucleotide polymorphisms (SNPs), allowed genotyping studies to progress from linkage based and candidate gene association studies to genome wide association studies (GWAS). A series of GWAS and derivative studies, such as the ones by Trynka *et al.*, Dubois *et al.*, and Van Heel *et al.* have resulted in the identification of more than 40 non-HLA CD associated loci (van Heel et al., 2007, Dubois et al., 2010, Trynka et al., 2011). Recently published papers on the Irish and US populations, have identified three further loci harbouring *ZNF335*, *NFIA*, *FRMD4B* and *DUSP10* genes among others, taking the total number of CD susceptibility loci to in excess of 50 (Ostensson et al., 2013, Garner et al., 2014, Coleman et al., 2015).

Many of these loci contain genes related to the immune response such as; T-cell development, T-cell and B-cell co-stimulation, detection of viral RNA by the innate immune system and cytokines and chemokines receptors (Dubois et al., 2010, Abadie et al., 2011, Wijmenga and Gutierrez-Achury, 2014). According to recent estimates, the HLA and non-HLA loci associated with CD together explain only roughly 50% of the genetic variation in the condition, although whether this estimate is accurate is a matter of debate currently (Ricano-Ponce et al., 2015). It may be, that the remaining missing heritability is explained by rare variants and/or common variants with modest effects, epistatic interactions and/or epigenetic phenomena for the identification of which current studies are underpowered (Ricano-Ponce et al., 2015, Fitzpatrick et al., 2015, Rose and Bell, 2012, Juran et al., 2012, Capuano et al., 2011, Fernandez-Jimenez et al., 2014). Another 2,550 SNPs with modest effect may contribute to the development of CD (Stahl et al., 2012, Ricano-Ponce et al., 2015). Alternatively, there is



considerable debate as to whether missing heritability may have been overestimated (Lee et al., 2013, Golan et al., 2014, Chen et al., 2015).

Sharing of susceptibility genes, including HLA, with several other immune-mediated conditions may in part explain the predisposition of CD patients to develop these functionally related conditions (Gutierrez-Achury et al., 2011). Interestingly, GWAS have shown that some of the non-HLA CD associated risk variants are indeed also shared with other immune-mediated conditions (Wijmenga and Gutierrez-Achury, 2014). A desire to characterise susceptibility variants across multiple diseases resulted in the development of the Immunochip, a customized genotyping array containing more than 250,000 genetic variants (SNPs) in its most recent iteration, in genes linked to immune mediated disease. The use of the Immunochip has contributed to the increased detection of non-HLA CD associated variants and more uniform characterisation of these variants across related diseases (Cortes and Brown, 2011, Wijmenga and Gutierrez-Achury, 2014).

There is little evidence as to the effects of genotype on phenotype in CD. The studies available have mainly focused in the HLA predisposing alleles or in a small subset of non-HLA genes with inconclusive results (Greco et al., 1998, Mustalahti et al., 2002, Zubillaga et al., 2002, Gudjonsdottir et al., 2009). All of these studies were published before the majority of the new non-HLA loci were discovered. It is suspected that CD has a polygenic architecture with HLA loci and other common and possibly rare variants with modest effects contributing to its phenotype (Wijmenga and Gutierrez-Achury, 2014). Epistatic effects are widely assumed to exist in polygenic diseases. Therefore, trying to establish differences in phenotypes based on a single SNP or allele seems to be of little benefit, and combining information from all the known risk variants seems to be a better choice (Romanos et al., 2014).

GWAS represent a rich source of information to carry out translational research that will link genetic science with clinical and public health practice (Janssens, 2008). The use of polygene risk scores (PRS) in CD, combining HLA and non-HLA predisposing loci, have been shown to improve the identification of potential CD patients (Romanos et al., 2014, Coleman et al., 2015). Moreover, genetic risk scores could be used as a predictive tool if shown to describe in advance the disease prognosis and its related complications. This concept has already been applied successfully in some studies related to other immune-mediated conditions. A recent study of inflammatory bowel disease found a significant association between increased genetic burden and early age of diagnosis in Crohn's disease (Ananthakrishnan et al., 2014).

In Ananthakrishnan *et al.* paper, genotyping was performed using the Illumina Immunochip, a weighed PRS was later calculated using the 163 distinct IBD risk loci by multiplying the log of the odds ratio (OR) for each risk locus by the risk allele dose at that locus (wild type, homozygous and heterozygous). Finally, a cumulative PRS for each patient was calculated by adding the contribution of each of the risk loci ( $\sum \log OR \times \text{allele dose}$ ). This paper also stratified the GRS into quartiles to examine a dose dependent response (Ananthakrishnan *et al.*, 2014). A different paper on rheumatoid arthritis (RA), using similar methods to the IBD paper, found a significant association between genetic risk score, seropositivity and erosive changes in RA (Chibnik *et al.*, 2011). Therefore, potentially, genotyping of HLA and non-HLA predisposing loci in CD patients and generating a composite genotypic risk score could be a practical way of identifying CD patients that will need close follow-up, like for example those who are prone to develop a co-existent IMC or will develop the disease earlier in life.

In summary, in this thesis, we wanted to establish the current situation regarding clinical presentation in Ireland and whether this has changed since our earliest patients were diagnosed in the 1980's. This includes age of onset, whether the patients have classical or other presentation, and the presence of comorbidities. We also wished to relate genotype to phenotype to see if any of these features could be related to particular genotypes. Thus we used the immunochip genotyping array to determine the genotypes of as many patients as possible whose detailed phenotypes could be collected. This required an extensive retrospective assessment of patient charts in several Irish hospitals and the development of a questionnaire for prospectively collected patients. From this approach we have established that the phenotype of CD patients has shifted over the last 50 years and that this is linked to changes in genotype, as will be outlined in the following chapters.

## **Chapter 2 Investigation of Coeliac Disease Clinical Phenotype in Ireland and its Potential Changes Over Time**

## 2.1 Introduction

Coeliac disease (CD) is a chronic small intestinal immune-mediated enteropathy precipitated by exposure to a gluten containing diet in genetically predisposed individuals (Ludvigsson et al., 2013a). Once considered to be a disorder characterized by overt malabsorptive symptoms and confined mainly to childhood, serological screening studies have proven that CD occurs in both paediatric and adult populations at a similar prevalence of approximately 1% (Reilly and Green, 2012). Some studies have also shown an increase in CD prevalence over time by studying the occurrence of the condition in recent general population cohorts compared to historic cohorts (Lohi et al., 2007, Rubio-Tapia et al., 2009, Catassi et al., 2010). This increase in prevalence over time has triggered a debate on whether these changes are merely reflecting higher detection rates due to improvements in diagnostic techniques, as well as greater awareness among both physicians and patients, or are instead showing a true increase in CD autoimmunity due to the interplay of morphing environmental factors and genetic predisposition to the condition (Lohi et al., 2007, Ludvigsson et al., 2013b, Kivela et al., 2015).

The hypothesis that CD is becoming more common seems to be corroborated by studies which have reported an increase in incidence over time which runs parallel to the increase in prevalence previously reported (Murray et al., 2003, Riddle et al., 2012, Ludvigsson et al., 2013b). Such a rapid change in disease occurrence cannot be attributed to changes in gene frequencies in the population, but could plausibly be due to shifting environmental factors. Moreover, these changes in CD incidence have happened simultaneously to alterations in the incidence of related conditions such as inflammatory bowel disease (IBD) and type 1 diabetes mellitus (T1DM). Notably, since T1DM diagnostics have remained relatively stable over time, and gene frequencies in large populations do not change quickly, variation in environmental exposures are the most likely cause of changed incidence (Harjutsalo et al., 2008, Gunesh et al., 2008, Maahs et al., 2010, Shivashankar et al., 2016).

Study of environmental factors in CD has mainly focused on infant feeding practices. This approach has been occasioned by the so-called Swedish epidemic of CD which took place in the late 1980s and early 1990s, when a sharp increase in symptomatic coeliac disease autoimmunity occurred among infants less than 2 years of age in Sweden. This increase coincided with a series of changes in infant feeding practices; a decrease in the rates of breastfeeding, increased amounts of gluten provided to infants, and change in the age of gluten introduction (Ivarsson et al., 2000, Ivarsson, 2005).

However, recently published studies that have followed infants predisposed genetically to develop CD have found that although the timing of gluten introduction can delay the disease, it does not prevent it. The same studies reported that breastfeeding had no protective effect on the development of the disease. Interestingly, however, the amount of gluten consumed by infants does seem to play a role in the process to tTGA conversion (Lionetti et al., 2014, Aronsson et al., 2015, Andren Aronsson et al., 2016b). However, as pointed out in the introduction, breastfeeding could play a protective role when other factors come into play such as certain HLA genotype and rotavirus infection (Kempainen et al., 2017).

Although gluten is the primary environmental factor necessary to trigger CD, disease initiation is likely to be a multifactorial process involving other environmental exposures. Intestinal infections (rotavirus among infants and *Campylobacter* among adults), intestinal microbiota and host microbiota interaction, the hygiene hypothesis which links geographical areas of high hygiene to greater occurrence of CD, type of delivery at birth, certain medications and even iron supplementation during pregnancy have all been suggested as factors which may influence CD development (Stene et al., 2006, Kondrashova et al., 2008, Decker et al., 2010, Riddle et al., 2013, Sjoberg et al., 2013, Marild et al., 2013, Lebwohl et al., 2014b). Recently it has been found that reovirus infections may be a triggering event in CD (Bouziat et al., 2017).

The clinical presentation of CD is highly heterogeneous, ranging from severe intestinal symptoms such as diarrhoea and weight loss, to subclinical or silent forms of the disease (Rubio-Tapia et al., 2013). In 2013, the Oslo definitions of coeliac disease attempted to achieve a consensus on the different terms used to describe the condition. This paper defined symptomatic CD as having clinically evident gastrointestinal or extra-intestinal symptoms, and divided it into classical and non-Classical CD. Classical CD is when malabsorption is present, thus diarrhoea, steatorrhoea, weight loss or growth failure are required at presentation. On the other hand, non-classical CD is diagnosed when no signs and symptoms of malabsorption are evident at presentation (i.e. patients with constipation and abdominal pain) or when patients present with monosymptomatic disease other than diarrhoea or steatorrhoea. The term subclinical CD was used in the same paper to describe CD without signs or symptoms that would lead to screening for the disease (Ludvigsson et al., 2013a, Volta et al., 2014, Kivela et al., 2015) (supplementary Table 1, appendix III).

The reported trends towards increased prevalence and incidence have been accompanied by a pronounced shift of the clinical phenotype towards a later age at

diagnosis and higher prevalence of non-classical and subclinical CD both in the paediatric and adult population

(Murray et al., 2003, Volta et al., 2014, Kivela et al., 2015). Gender differences in disease presentation and prognosis have also been reported, such as later age at diagnosis for male patients and more co-existent comorbidities in female patients (Vivas et al., 2008). It is also now widely accepted that, contrary to the traditional view of CD patients presenting underweight at diagnosis, more recently most CD patients have a normal or high body mass index (BMI) at diagnosis (Dickey and Kearney, 2006, Olen et al., 2009, Cheng et al., 2010). Once more, the fact that clinical presentation has considerably mutated over time points towards changing environmental factors as the driving force behind this transformation.

CD patients' have a predisposition to develop other immune-mediated conditions (IMC), most notably autoimmune thyroid disease (ATD) and type I diabetes mellitus (T1DM). Sharing of genetic susceptibility variants presumably contributes to the high prevalence of these co-existent IMCs in CD (Gutierrez-Achury et al., 2011, Lundin and Wijmenga, 2015). However, gene-environment interactions likely also contribute, as highlighted by studies suggesting a protective effect of the gluten free diet (GFD) on comorbid IMCs (Cosnes et al., 2008, Mones, 2009). Therefore, a timely CD diagnosis may be important in improving both patient quality of life and disease prognosis (Snyder and Murray, 2015). However, diverse studies continue to show high proportions of undiagnosed patients and long delays in diagnosis, sometimes over 10 years from the onset of symptoms (Cranney et al., 2007, Gray and Papanicolas, 2010, Norstrom et al., 2011, Ditah et al., 2014)

Studies describing the clinical phenotype and disease evolution in large, well characterized cohorts of CD patients, are few (Kocsis et al., 2015). There is a particular paucity of data regarding current patterns of presentation in Ireland, traditionally an area of high diagnosis due to high awareness of the condition since the 1970's (Mylotte et al., 1973, Tajuddin et al., 2011, Saleem et al., 2012). The aims of the present chapter were to carry out a cross-sectional analysis of the clinical characteristics of the disease in a cohort of CD patients (n=749) attending secondary referral centres in Ireland, with a view to documenting the evolution of the condition over a 55 year time period in a country with relatively high awareness of the condition.

## 2.2 Aims

- Investigation of the clinical presentation and demographic characteristics of our sample of 749 patients
- Analysis of potential differences between patients at disease presentation classified according to gender, BMI, degree of damage to the small bowel mucosa, age at diagnosis in adulthood and geographical place of diagnosis.
- To explore potential changes in disease presentation over time, both using the full sample, and segregating it by geographical location at diagnosis.
- To explore weight changes in CD patients after diagnosis as well as changes in their BMI at diagnosis over time.

## **2.3 Methods**

### **2.3.1 Study design and study population**

A retrospective analysis of medical records was performed using a CD patient cohort (n=749) attending five secondary referral centres in Ireland and diagnosed from 1960-2015. Patients with suspected CD are normally referred to secondary referral centres with expertise in CD diagnosis and management; our patients are therefore likely to be a relatively unselected and representative sample. The majority of the patients were identified based on their participation in a previous study on the genetics of CD. All patients are Caucasians of Irish ancestry. 79% of patients were diagnosed in the five participating referral centres with the remainder diagnosed at other secondary referral centres across Ireland. Data was collected from medical records, hospital information systems, and a structured questionnaire completed by the patient and the physician or research nurse for patients currently attending the coeliac clinic (Appendices I and II). All patients gave written informed consent.

CD was diagnosed by serology and histology following current guidelines (Ludvigsson et al., 2014a). Serology included anti-Gliadin antibodies (AGA) for patients diagnosed during the 1960's to the 1990's; anti-Tissue Transglutaminase Antibodies (tTGA) were introduced in the late 1990's and have since been the first-line screening method for CD diagnosis together with anti-Endomysial antibodies (EMA). Small intestinal biopsies were taken from the jejunum or the second duodenal portion, and most were classified according to the modified Marsh classification by a pathologist. Where the degree of damage to the bowel mucosa was provided as follows: partial, subtotal and total villous atrophy, these were re-classified as Marsh 3a, Marsh 3b and Marsh 3c respectively (Dickson et al., 2006).

Sixteen patients with abnormal tTGA and/or EMA and Marsh 1 biopsy were included based on improvement on the GFD and serology normalisation (Ludvigsson et al., 2014a). Two patients with abnormal AGA and Marsh 1 biopsy were included in the analysis on the basis of co-existent dermatitis herpetiformis (DH), family history of CD, and improvement on the GFD with serology normalisation (Dickson et al., 2006). Fourteen seronegative patients with no IgA deficiency but abnormal biopsy at diagnosis, were included in the analysis based on evidence of improvement on the GFD and/or biopsy (Ierardi et al., 2015). Patients were excluded if the diagnosis of CD was not clearly reported on the medical records. The study was approved by the



Research Ethics Committees of St James's Hospital Dublin and all other contributing centres.

### **2.3.2 Data collection and definition of study variables**

The following information was collected; age, gender, age at diagnosis, symptoms at diagnosis, serology and histology data at diagnosis, ethnicity, family history of CD and of other IMCs. Patient age was considered to be the age at last follow-up. The following manifestations at diagnosis were collected; diarrhoea, steatorrhea, abdominal pain, constipation, vomiting, heartburn or indigestion, bloating, flatulence, weight loss, fatigue or tiredness and iron deficiency anaemia (IDA). A category classified as 'other manifestations' was also created for signs and symptoms present at diagnosis which had not been included initially.

The clinical presentation at diagnosis was classified according to the Oslo classification as classical, non-classical and subclinical using the previously recorded manifestations (Ludvigsson et al., 2013a) (supplementary Table 1, appendix III). Patients diagnosed at less than 18 years of age were excluded from analyses relating to clinical presentation over time due to the small sample size and information deficits on their clinical presentation at diagnosis. Manifestations at diagnosis were available for 602 patients diagnosed in adulthood.

The total prevalence of the following CD extra-intestinal manifestations and co-existent IMCs was determined in the study cohort; bone disorders (osteoporosis and/or osteopenia), IDA, vitamin deficiency (B12 and/or folic acid deficiency), lymphoma, depression, ataxia, neuropathy, dermatitis herpetiformis (DH), T1DM, autoimmune thyroid disease (ATD), rheumatoid arthritis (RA), inflammatory bowel disease (IBD), pernicious anaemia, vitiligo, psoriasis, systemic lupus erythematosus and Reiter's syndrome. A category 'other' was created to include any IMCs that had not been included initially (Gutierrez-Achury et al., 2011, Leffler et al., 2015a).

### **2.3.3 Assays utilised to determine coeliac serology**

During the time period covered by this study, tTGA value calculations were determined using a variety of tTGA diagnostic tests with different cut-off points. In order to compare results from different tests, tTGA values were expressed as the fold increase above the cut-off value by calculating the ratio of the observed tTGA at diagnosis divided by the cut-off value for an equivocal result (Singh et al., 2014). The resulting values were corrected for normality by log transformation, and a value of 1 was added to remove negative values (John and Draper, 1980). To test the resulting tTGA variable we

correlated it with the degree of mucosal damage at diagnosis (Marsh criteria) and observed a significant correlation as reported in previous studies (Singh et al., 2014, Rahmati et al., 2014, Donaldson et al., 2007) (Figure 2.10). tTGA values at diagnosis were available for 346 patients diagnosed in adulthood; 113 of these were analysed using Phadia Celikey ELISA assay with anti-tTGA titre of 1.9-5 U/ml for an equivocal result; 111 used Thermo Fisher ELISA assay 5-10 U/ml; 59 used Thermo Fisher Celikey EliA assay 4-10 U/ml; 38 used Thermo Fisher Celikey EliA 7-10 U/ml, and 25 used Phadia Celikey ELISA assay 2-8 U/ml.

## **2.3.4 Statistical analysis**

### **2.3.4.1 Statistical analysis of changes in clinical presentation over time**

A categorical variable was created for the age at diagnosis with coeliac disease coded as follows; 1) 0-17 years of age, 2)  $\geq 18$ -35 years, 3)  $>35$ -55 years and 4)  $>55$  years. A second categorical variable was created for the age at diagnosis with coeliac disease coded as follows; 1) 0-17 years, 2)  $\geq 18$ -40 years, 3)  $>40$ -64 and 4)  $\geq 65$  years. These two variables were created to capture potential differences in presentation of the condition when diagnosed at different stages in adult life; early adulthood, middle-adulthood and the elderly. Of special interest are those diagnosed in the elderly group (patients  $\geq 55$  or  $\geq 65$  years at diagnosis) as defined by previous papers, due to the paucity of data in clinical presentation of CD in this demographic group (Vilppula et al., 2009, Rashtak and Murray, 2009, Casella et al., 2012).

A categorical variable was also created for the year or time period in which the diagnosis was made, as follows; 1)  $\leq 1985$ , 2) 1986-1995, 3) 1996-2005 4) 2006-2010 and 5)  $\geq 2011$ . These periods are similar to those of Kivela *et al.* and were selected to capture potential changes in CD presentation and associated complications, and the introduction of new diagnostic procedures. Thus before 1985, all biopsies were jejunal capsule biopsies which were widely replaced by duodenal pinch biopsies from 1986. EMAs and tTGAs were introduced in 1994 and 1997 respectively substituting previously used AGAs (Kivela et al., 2015).

Differences in clinical presentation segregating the sample by gender, BMI, damage to the small bowel mucosa, age at diagnosis in adulthood and geographical place of diagnosis have also been explored (Tables 2.3, 2.4, 2.5, 2.6 and 2.7).

#### **2.3.4.2 Statistical analysis of changes in clinical presentation over time per geographical area**

The sample in the present study can be geographically divided between CD patients diagnosed in the west of Ireland in Galway University Hospital (GUH), and CD patients diagnosed in several secondary referral centres situated in Dublin in the east of Ireland (supplementary Figure 1, appendix IV). Therefore, we also performed a comparison of the patients diagnosed in adulthood over time segregated by geographical area (Table 2.9 and Figure 2.5), due to low numbers of patients in the period from 1986 to 1995 (n=6) in the Galway sample we decided to collapse the first two time periods into one  $\leq 1995$  to carry out comparisons between the sample diagnosed in the west and in the east. Supplementary tables carrying out analysis using the 5 time periods segregated by geographical area and using the full sample diagnosed in adulthood with the 4 time periods ( $\leq 1995$ ; 1996-2005; 2006-2010;  $\geq 2011$ ) have been also provided (supplementary Figure 2 and supplementary Figure 3, appendix V and VI).

#### **2.3.4.3 Statistical analysis of BMI changes over time**

BMI defined as  $\text{body weight(kg)}/\text{height(m}^2\text{)}$  was collected for a subcohort of patients (n=402) at diagnosis with CD in adulthood, before being prescribed a GFD by a dietician. The World Health Organisation defines a BMI  $<18.5$  as underweight, 18.5-24.9 as normal,  $\geq 25$  as overweight and  $\geq 30$  as obese. 166 patients diagnosed with CD in adulthood had their BMI at diagnosis and a year later available (mean follow-up 13 months). 29 of these patients were assessed retrospectively for adherence to the GFD by Dr Eva Grace, an expert coeliac dietician from the department of Clinical Nutrition in Saint James Hospital, using an adherence scale described by Leffler *et al.* (2009) (supplementary Table 2, appendix VII). Certain serum biomarkers including; serum haemoglobin, ferritin, folate and B12 were collected for a small subcohort of patients by Ms Melanie Martin from the department of Clinical Nutrition in Saint James Hospital. These were correlated with BMI at diagnosis as well as with the degree of damage to the small bowel mucosa at diagnosis.

#### **2.3.4.4 Regression analysis of clinical changes over time per geographical area.**

Multivariate binary logistic regression was carried out in the CD patients' cohort diagnosed in adulthood using time period of CD diagnosis and geographical place of diagnosis as independent variables. The dependant binary variable in the regression analysis using geographical place and time period of CD diagnosis analysis was having classical or non-classical CD at presentation with the condition. Regression analysis

was also carried out selecting first for those patients who presented classically and secondly for those who presented non-classically at diagnosis using time period of diagnosis as the independent variable and the geographical place of diagnosis as the dependant variable.

Statistical analysis were carried out using SPSS version 24.0 (SPSS, Inc., Chicago, IL). Discrete variables are presented as percentages and continuous variables as medians and ranges. The relationship between discrete variables was analysed using the  $\chi^2$  test. Differences between groups were assessed using Mann-Whitney U, Kruskal-Wallis H and Wilcoxon non-parametric tests. The Jonckheere-Terpstra test was used to carry out trend analysis. A threshold p-value of <0.05 was considered significant in all analyses.

## 2.4 Results

### 2.4.1 Characteristics of the sample

Demographic and clinical characteristics of the patient population are shown in Table 2.1.

**Table 2.1 Characteristics of coeliac patients (n=749)**

| Characteristic                                                                | % <sup>†</sup> | Median (range)   |
|-------------------------------------------------------------------------------|----------------|------------------|
| <b>Age of diagnosis (years)<sup>‡</sup>, n 734</b>                            |                | 44 (0.5-86.0)    |
| <b>Age of diagnosis &lt;18 years (years)<sup>‡</sup>, n 82</b>                |                | 4.5 (0.5-17.0)   |
| <b>Age of diagnosis in adulthood (years), n 652</b>                           |                | 46.0 (18.0-86.0) |
| <b>Females/Males n 749 (%)</b>                                                | 68.4/31.6      |                  |
| <b>Family history of celiac disease, n 749 (%)</b>                            |                |                  |
| Yes                                                                           | 36.4           |                  |
| <b>Relative affected by celiac disease<sup>§</sup>, n 273 (%)</b>             |                |                  |
| First degree                                                                  | 56.4           |                  |
| Distant                                                                       | 19.8           |                  |
| Both                                                                          | 21.6           |                  |
| Not Available                                                                 | 2.2            |                  |
| <b>Diagnosis by age range, n 749 (%)</b>                                      |                |                  |
| < 18 years                                                                    | 13.0           |                  |
| ≥18-35 years                                                                  | 22.0           |                  |
| >35-55 years                                                                  | 37.7           |                  |
| >55                                                                           | 27.4           |                  |
| <b>Clinical presentation at diagnosis<sup>**</sup>, n 602 (%)</b>             |                |                  |
| Classical CD                                                                  | 55.6           |                  |
| Non-classical CD                                                              | 38.7           |                  |
| Subclinical CD                                                                | 5.6            |                  |
| <b>Body mass index at diagnosis<sup>‡</sup>, n 391 (%)</b>                    |                |                  |
| Underweight (<18.5 kg/m <sup>2</sup> )                                        | 6.4            |                  |
| Normal weight (18.5-24.9kg/m <sup>2</sup> )                                   | 52.7           |                  |
| Overweight (25.0-29.9 kg/m <sup>2</sup> )                                     | 28.4           |                  |
| Obese (≥ 30 kg/m <sup>2</sup> )                                               | 12.5           |                  |
| <b>Manifestations at diagnosis<sup>‡</sup>, n 602 (%)</b>                     |                |                  |
| Diarrhoea                                                                     | 42.9           |                  |
| Abdominal Pain                                                                | 34.1           |                  |
| Bloating                                                                      | 25.0           |                  |
| Constipation                                                                  | 13.6           |                  |
| Steatorrhea                                                                   | 6.8            |                  |
| Vomiting                                                                      | 6.2            |                  |
| Fatigue/Tiredness                                                             | 31.4           |                  |
| Iron Deficiency Anaemia                                                       | 29.6           |                  |
| Weight Loss                                                                   | 28.3           |                  |
| Other                                                                         | 45.5           |                  |
| <b>Most common "Other" manifestations at diagnosis<sup>‡</sup>, n 602 (%)</b> |                |                  |
| Heartburn/Gastroesophageal Reflux                                             | 14.5           |                  |
| Flatulence                                                                    | 8.2            |                  |
| Nausea                                                                        | 7.3            |                  |
| Anorexia                                                                      | 3.7            |                  |
| Blood per Rectum/Haemorrhoids                                                 | 2.7            |                  |
| Belching                                                                      | 2.3            |                  |
| Skin Conditions <sup>λ</sup>                                                  | 8.6            |                  |
| Recurrent Mouth Ulcers                                                        | 7.7            |                  |
| B12 deficiency                                                                | 2.5            |                  |
| Folate deficiency                                                             | 2.2            |                  |

£15 patients diagnosed <18 years for whom the exact age at diagnosis was not available excluded

§ First degrees relatives are parents, siblings and offsprings

\* Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into; Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

λ Conditions included were dermatitis, eczema, rosacea and skin rashes

The median age at data collection was 56 years (range 18-91). 97 patients were diagnosed <18 years of age, and 652 were diagnosed in adulthood. A higher prevalence of female patients was found (female/male ratio = 2.2:1). Classical CD was the most common presentation at diagnosis (55.6%). As shown in Table 2.1, diarrhoea and fatigue/tiredness were the most commonly reported intestinal and extra-intestinal symptoms (42.9%) and (31.4%) respectively, and 40.9% of patients presented either overweight or obese at diagnosis (Table 2.1).

**Table 2.2 Prevalence of coeliac disease extra-intestinal symptoms and co-existent immune-mediated conditions (n=749).**

| Characteristic                                                                         |                                  | %    |
|----------------------------------------------------------------------------------------|----------------------------------|------|
| <b>Celiac Patients</b>                                                                 |                                  |      |
| <b>Prevalence of celiac disease extra-intestinal manifestations<sup>‡</sup>, n 749</b> | Yes                              | 66.1 |
| <b>Celiac disease extra-intestinal manifestations, n 749</b>                           | Bone Disorders <sup>§</sup>      | 40.6 |
|                                                                                        | Iron Deficiency                  | 35.4 |
|                                                                                        | Vitamin Deficiency <sup>*</sup>  | 16.8 |
|                                                                                        | Depression                       | 9.1  |
|                                                                                        | Dermatitis herpetiformis         | 7.2  |
|                                                                                        | Reflux Oesophagitis              | 4.1  |
|                                                                                        | Ataxia                           | 1.1  |
|                                                                                        | Lymphoma                         | 1.1  |
|                                                                                        | Neuropathy                       | 0.5  |
| <b>Presence of co-existent immune-mediated conditions , n 749</b>                      | Yes                              | 31.2 |
| <b>Co-existent immune-mediated conditions, n 749</b>                                   | Autoimmune thyroid disease       | 19.9 |
|                                                                                        | Type 1 diabetes mellitus         | 3.6  |
|                                                                                        | Psoriasis                        | 2.7  |
|                                                                                        | Inflammatory bowel disease       | 2.1  |
|                                                                                        | Rheumatoid arthritis             | 1.6  |
|                                                                                        | Pernicious Anaemia               | 1.1  |
|                                                                                        | Systemic lupus erythematosus     | 0.4  |
|                                                                                        | Other                            | 5.5  |
| <b>Other co-existent immune-mediated conditions, n 749</b>                             | Alopecia areata                  | 1.3  |
|                                                                                        | Sarcoidosis                      | 0.9  |
|                                                                                        | Immunoglobulin A deficiency      | 0.5  |
|                                                                                        | Multiple sclerosis               | 0.4  |
|                                                                                        | Scleroderma                      | 0.4  |
|                                                                                        | Lichen planus                    | 0.4  |
|                                                                                        | Autoimmune hepatitis             | 0.4  |
|                                                                                        | Autoimmune adrenalitis           | 0.3  |
|                                                                                        | Primary biliary cholangitis      | 0.3  |
|                                                                                        | Myasthenia gravis                | 0.1  |
|                                                                                        | Immune-mediated thrombocytopenia | 0.1  |
|                                                                                        | Vitiligo                         | 0.1  |
|                                                                                        | Autoimmune neutropenia           | 0.1  |

<sup>‡</sup>Manifestations occurring either at celiac disease diagnosis, or prior or posterior to celiac disease diagnosis

<sup>§</sup> Osteoporosis and/or Osteopenia

<sup>\*</sup>B12 or Folic acid deficiency

As shown in Table 2.2, over 31% of patients had a co-existent immune-mediated condition (IMC), the most prevalent of which were ATD (19.9%) and T1DM (3.6%). The most prevalent CD extra-intestinal manifestations were bone disorders (40.9%), followed by iron deficiency (35.4%). DH was present in 7.2% of cases.

**Table 2.3 Characteristics of the sample segregated by gender (n=749).**

| Characteristic                                                                             | Median or %† |         | P-value          |
|--------------------------------------------------------------------------------------------|--------------|---------|------------------|
|                                                                                            | Males        | Females |                  |
| <b>Age of diagnosis<sup>£</sup>, n 734 (years)</b>                                         | 47.0         | 42.0    | <b>0.004</b>     |
| <b>Diagnosis by age range, n 749 (%)</b>                                                   |              |         |                  |
| < 18 years                                                                                 | 11.4         | 13.7    |                  |
| ≥18-35 years                                                                               | 14.8         | 25.4    |                  |
| >35-55 years                                                                               | 39.7         | 36.7    |                  |
| >55                                                                                        | 34.2         | 24.2    | <b>0.002</b>     |
| <b>Clinical presentation at diagnosis<sup>*†</sup>, n 602 (%)</b>                          |              |         |                  |
| Classical coeliac disease                                                                  | 55.3         | 55.6    |                  |
| Non-classical/Subclinical coeliac disease                                                  | 44.7         | 44.4    | 0.957            |
| <b>Body mass index at diagnosis, n 391 (%)</b>                                             |              |         |                  |
| Underweight (<18.5 kg/m <sup>2</sup> )                                                     | 1.8          | 8.3     |                  |
| Normal weight (18.5-24.9kg/m <sup>2</sup> )                                                | 50.9         | 53.4    |                  |
| Overweight (25.0-29.9 kg/m <sup>2</sup> )                                                  | 30.7         | 27.4    |                  |
| Obese (≥ 30 kg/m <sup>2</sup> )                                                            | 16.7         | 10.8    | <b>0.046</b>     |
| <b>Manifestations at diagnosis<sup>‡</sup>, n 602 (%)</b>                                  |              |         |                  |
| Diarrhoea                                                                                  | 41.5         | 43.5    | 0.648            |
| Steatorrhea                                                                                | 6.4          | 7.0     | 0.779            |
| Abdominal Pain                                                                             | 30.3         | 35.7    | 0.193            |
| Constipation                                                                               | 6.9          | 16.4    | <b>0.002</b>     |
| Vomiting                                                                                   | 4.3          | 7.0     | 0.193            |
| Bloating                                                                                   | 18.1         | 28.0    | <b>0.009</b>     |
| Weight Loss                                                                                | 31.4         | 26.8    | 0.248            |
| Fatigue/Tiredness                                                                          | 27.7         | 33.1    | 0.183            |
| Iron Deficiency Anaemia                                                                    | 23.4         | 32.4    | <b>0.026</b>     |
| Other                                                                                      | 45.2         | 45.7    | 0.930            |
| <b>Family history of coeliac disease, n 749 (%)</b>                                        |              |         |                  |
| Yes                                                                                        | 33.3         | 37.9    | 0.228            |
| <b>Presence of coeliac disease extra-intestinal manifestations <sup>λ</sup>, n 749 (%)</b> |              |         |                  |
| Yes                                                                                        | 57.8         | 72.7    | <b>&lt;0.001</b> |
| <b>Family history of other immune-mediated conditions, n 749 (%)</b>                       |              |         |                  |
| Yes                                                                                        | 20.3         | 30.9    | <b>0.003</b>     |
| <b>Presence of other immune-mediated conditions , n 749 (%)</b>                            |              |         |                  |
| Yes                                                                                        | 21.5         | 35.7    | <b>&lt;0.001</b> |

£15 patients diagnosed <18 years for whom the exact age at diagnosis was not available excluded

\* Clinical presentation at diagnosis was classified according to the Oslo definition of coeliac disease into; Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

λ coeliac disease extra-intestinal manifestations considered; Osteoporosis/Osteopenia, Iron deficiency, B12/Folate deficiency and Depression

**Table 2.4 Characteristics of the CD patient sample diagnosed in adulthood segregated by degree of damage to the small bowel mucosa classified according to Marsh criteria (n=355).**

| Biopsy at diagnosis*                                       | Marsh ≤ 2 (%)        | Marsh 3a (%) | Marsh 3b (%)      | Marsh 3c (%) | P-value      |
|------------------------------------------------------------|----------------------|--------------|-------------------|--------------|--------------|
| Total number of patients (% of total) <sup>†</sup>         | 20 (5.6)             | 75 (21.0)    | 148 (41.5)        | 112 (31.4)   |              |
| <b>Demographics</b>                                        |                      |              |                   |              |              |
| Female patients (%)                                        | 13 (65.0)            | 54 (72.0)    | 98 (66.2)         | 81 (72.3)    | 0.67         |
| Median age at CD diagnosis (range) <sup>‡</sup>            | 50.5 (22-76)         | 47.0 (18-80) | 47.5 (18-85)      | 49.5 (19-81) | 0.87         |
| <b>Presentation, n=542*</b>                                |                      |              |                   |              |              |
| Classical CD (%)                                           | 8 (40.0)             | 41 (56.2)    | 82 (56.6)         | 63 (57.8)    | 0.525        |
| Non-classical or subclinical CD (%)                        | 12 (60.0)            | 32 (43.8)    | 63 (43.4)         | 46 (42.2)    |              |
| <b>Manifestations at diagnosis, 347</b>                    |                      |              |                   |              |              |
| Diarrhea                                                   | 7 (35.0)             | 27 (37.0)    | 61 (42.1)         | 48 (44.0)    | 0.74         |
| Steatorrhea                                                | 0                    | 1 (1.4)      | 12 (8.3)          | 9 (8.3)      | 0.11         |
| Abdominal Pain                                             | 10 (50.0)            | 28 (38.4)    | 46 (31.7)         | 38 (34.9)    | 0.39         |
| Constipation                                               | 3 (15.0)             | 12 (16.4)    | 18 (12.4)         | 12 (11.0)    | 0.74         |
| Vomiting                                                   | 1 (5.0)              | 5 (6.8)      | 8 (5.5)           | 7 (6.4)      | 0.97         |
| Bloating                                                   | 5 (25.0)             | 30 (41.1)    | 31 (21.4)         | 25 (22.9)    | <b>0.013</b> |
| Weight Loss                                                | 3 (15.0)             | 21 (28.8)    | 42 (29.0)         | 40 (36.7)    | 0.21         |
| Fatigue/Tiredness                                          | 6 (30.0)             | 26 (35.6)    | 40 (27.6)         | 38 (34.9)    | 0.54         |
| Iron deficiency anaemia at diagnosis (%)                   | 2 (10.0)             | 14 (19.2)    | 50 (34.5)         | 39 (35.8)    | <b>0.012</b> |
| Folate or B12 deficiency at diagnosis (%)                  | 0                    | 1 (1.4)      | 9 (6.2)           | 2 (1.9)      | 0.12         |
| Other                                                      | 8 (40.0)             | 29 (39.7)    | 70 (48.3)         | 43 (39.4)    | 0.46         |
| <b>Extra-intestinal manifestations<sup>¶</sup></b>         |                      |              |                   |              |              |
| Osteoporosis                                               | 4 (20.0)             | 16 (21.3)    | 32 (21.6)         | 34 (30.0)    | 0.34         |
| Osteopenia                                                 | 6 (30.0)             | 22 (29.3)    | 37 (25.0)         | 34 (30.4)    | 0.78         |
| Dermatitis Herpetiformis                                   | 4 (20.0)             | 5 (6.7)      | 11 (7.4)          | 5 (4.5)      | 0.097        |
| <b>Bone mineral density (g/cm<sup>2</sup>)<sup>∞</sup></b> |                      |              |                   |              |              |
|                                                            | <b>Marsh &lt; 3b</b> |              | <b>Marsh ≥ 3b</b> |              |              |
| Median lumbar spine, n=77                                  | 0.925                |              | 0.905             |              | 0.25         |
| Median femur neck, n=64                                    | 0.81                 |              | 0.795             |              | 0.195        |
| <b>Serum biomarkers at diagnosis</b>                       |                      |              |                   |              |              |
| Median serum haemoglobin (g/dl), n=33                      | 13.7                 |              | 13.45             |              | 0.81         |
| Median serum ferritin (µg/l), n=23                         | 56                   |              | 23.8              |              | 0.49         |
| Median serum folate (µg/l), n=26                           | 5.2                  |              | 7.0               |              | 0.16         |
| Median serum B12 (µg/l), n=26                              | 290.0                |              | 327               |              | 0.43         |

\* Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

\*Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into: Classical, Non-classical and Subclinical

∞ Bone mineral density at diagnosis of osteoporosis or osteopenia

¶ Manifestations occurring either at celiac disease diagnosis, or prior or posterior to celiac disease diagnosis

The only significant differences when the sample was segregated by biopsy category at diagnosis according to Marsh criteria were more bloating reported by those who presented with a Marsh score of ≤3a (p=0.013), and more IDA reported by those who presented with a Marsh score of ≥3b (p=0.012). More differences would have been expected with regard to bone mineral density (BMD) and serum biomarkers due to greater malabsorption in the group >3b biopsy, however the lack of significant differences is probably due to the small sample size for BMD and serum biomarkers. The data suggests a tendency towards higher BMD at diagnosis of bone disease in those with a Marsh biopsy <3b at diagnosis (p=0.25 & p=0.195). Also, a tendency towards higher serum folate is observed in those with a Marsh ≥3b at diagnosis (p=0.016) indicating it could be instructive to examine these measures in an extended study.



**Table 2.5 Characteristics of the sample segregated geographically by patients diagnosed in Galway University Hospital in the west of Ireland and those diagnosed in the other participating centres in Dublin in the east of Ireland (n=749).**

| Characteristic                                         | Median or %†        |                       | P-value          |
|--------------------------------------------------------|---------------------|-----------------------|------------------|
|                                                        | West Coast, n=211   | East Coast, n=538     |                  |
| Age of diagnosis <sup>£</sup> , n 734 (years)          | 42.0                | 44.5                  | 0.16             |
| Age of diagnosis <18 years (years) <sup>£</sup> , n 82 | 8.0                 | 3.0                   | 0.12             |
| Age of diagnosis in adulthood (years), n 652           | 45.0                | 47.0                  | 0.45             |
| Females/Males n 749 (%)                                | 142(67.3)/69 (32.7) | 370 (68.8)/168 (31.2) | 0.73             |
| Diagnosis by age range, n 734 (%)                      |                     |                       |                  |
| <18 years                                              | 33 (15.6)           | 64 (11.9)             |                  |
| ≥18-35 years                                           | 55 (26.1)           | 110 (20.4)            |                  |
| >35-55 years                                           | 69 (32.7)           | 213 (39.6)            |                  |
| >55                                                    | 54 (25.6)           | 151 (28.1)            | 0.11             |
| Clinical presentation at diagnosis**, n 602 (%)        |                     |                       |                  |
| Classical celiac disease                               | 94 (56.3)           | 240 (55.2)            |                  |
| Non-classical/Subclinical celiac disease               | 73 (43.7)           | 195 (44.8)            | 0.85             |
| Manifestations at diagnosis <sup>‡</sup> , n 602 (%)   |                     |                       |                  |
| Diarrhea                                               | 66 (39.5)           | 192 (44.1)            | 0.31             |
| Steatorrhea                                            | 10 (6.0)            | 31 (7.1)              | 0.72             |
| Abdominal Pain                                         | 50 (29.9)           | 155 (35.6)            | 0.21             |
| Constipation                                           | 18 (10.8)           | 64 (14.7)             | 0.23             |
| Vomiting                                               | 11 (6.6)            | 26 (6.0)              | 0.78             |
| Bloating                                               | 40 (24.0)           | 110 (25.3)            | 0.73             |
| Weight Loss                                            | 55 (32.9)           | 115 (26.5)            | 0.13             |
| Fatigue/Tiredness                                      | 51 (30.5)           | 138 (31.7)            | 0.84             |
| Iron Deficiency Anaemia                                | 46 (27.5)           | 129 (29.7)            | 0.69             |
| Other                                                  | 75 (44.9)           | 199 (45.7)            | 0.93             |
| Extra-intestinal manifestations <sup>¶</sup>           |                     |                       |                  |
| Depression                                             | 10 (4.7)            | 58 (10.8)             | <b>0.010</b>     |
| Dermatitis Herpetiformis                               | 14 (6.6)            | 39 (7.2)              | 0.875            |
| Biopsy at diagnosis, n=355 (%)                         |                     |                       |                  |
| Marsh ≤ 2                                              | 4 (3.0)             | 16 (7.2)              |                  |
| Marsh 3a                                               | 28 (21.2)           | 47 (21.1)             |                  |
| Marsh 3b                                               | 73 (55.3)           | 75 (33.6)             |                  |
| Marsh 3c                                               | 27 (20.5)           | 85 (38.1)             | <b>&lt;0.001</b> |
| Tissue transglutaminase <sup>§</sup> , n=346 (%)       | 2.0                 | 2.0                   | 0.18             |

£15 patients diagnosed <18 years for whom the exact age at diagnosis was not available excluded

\* Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into; Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

¶ Manifestations occurring either at celiac disease diagnosis, or prior or posterior to celiac disease diagnosis, data on bone disease not included as incomplete for UHG

§Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

The only significant difference between centres is that less Marsh 3c biopsies are observed in the west coast compared to the east coast, however the overall percentage of Marsh 3b and 3c biopsies in the west and east are similar (75.8% and 71.7% respectively). Also less CD patients in the west coast were found to have depression (p=0.01).

The differences between west coast and east coast when considering Marsh 3c biopsies could be due to a larger percentage of the sample from the west coast having been diagnosed >2005 (60.4%) compared to the sample coming from the east coast (46.1%) (p<0.0001) (data not shown). As shown in Table 2.8 Marsh 3c biopsies are significantly reduced in the period >2005 (p=0.003).

**Table 2.6 Characteristics of the sample diagnosed in adulthood segregated by BMI category at diagnosis (n=402).**

| BMI at diagnosis* (kg/m <sup>2</sup> )             | <18.5      | 18.5-24.9    | 25-29.9      | ≥30        | P-value <sup>§</sup> | P-value <sup>£</sup> |
|----------------------------------------------------|------------|--------------|--------------|------------|----------------------|----------------------|
| Total number of patients (% of total) <sup>†</sup> | 25 (6.2)   | 214 (53.2)   | 112 (27.9)   | 51 (12.7)  |                      |                      |
| <b>Demographics</b>                                |            |              |              |            |                      |                      |
| Female patients (%)                                | 23 (92.0)  | 153 (71.5)   | 76 (67.9)    | 31 (60.8)  | <b>0.04</b>          | 0.06                 |
| Median age at CD diagnosis (range)                 | 46 (21-68) | 48.5 (18-80) | 51.5 (19-77) | 50 (18-81) | 0.062                | 0.1                  |
| Median age at data collection (range) <sup>‡</sup> | 60 (30-88) | 59 (18-87)   | 58 (22-91)   | 55 (21-83) | 0.62                 | 0.6                  |
| <b>Presentation, n=397*</b>                        |            |              |              |            |                      |                      |
| Classical CD (%)                                   | 18 (75.0)  | 127 (59.6)   | 52 (46.4)    | 23 (47.9)  |                      |                      |
| Non-classical or subclinical CD (%)                | 6 (25.0)   | 86 (40.4)    | 60 (53.6)    | 25 (52.1)  | <b>0.019</b>         | <b>0.009</b>         |
| <b>Manifestations at diagnosis, 397</b>            |            |              |              |            |                      |                      |
| Diarrhea                                           | 14 (58.3)  | 89 (41.8)    | 46 (41.1)    | 22 (45.8)  | 0.43                 | 0.96                 |
| Steatorrhea                                        | 3 (12.5)   | 16 (7.5)     | 3 (2.7)      | 2 (4.2)    | 0.165                | <b>0.03</b>          |
| Abdominal Pain                                     | 9 (37.5)   | 69 (32.4)    | 38 (33.9)    | 19 (39.6)  | 0.79                 | 0.48                 |
| Constipation                                       | 4 (16.7)   | 28 (13.1)    | 20 (17.9)    | 10 (20.8)  | 0.49                 | 0.28                 |
| Vomiting                                           | 1 (4.2)    | 13 (6.1)     | 7 (6.3)      | 2 (4.2)    | 0.93                 | 0.77                 |
| Bloating                                           | 7 (29.2)   | 47 (22.1)    | 32 (28.6)    | 12 (25.0)  | 0.58                 | 0.27                 |
| Weight Loss                                        | 14 (58.3)  | 75 (35.2)    | 22 (19.6)    | 1 (2.1)    | <b>&lt;0.001</b>     | <b>&lt;0.001</b>     |
| Fatigue/Tiredness                                  | 8 (33.3)   | 66 (31.0)    | 36 (32.1)    | 12 (25.0)  | 0.82                 | 0.74                 |
| Iron deficiency anaemia at diagnosis (%)           | 5 (20.8)   | 66 (31.0)    | 34 (30.4)    | 8 (16.70)  | 0.185                | 0.78                 |
| Folate or B12 deficiency at diagnosis (%)          | 1 (4.2)    | 9 (4.2)      | 4 (3.6)      | 0          | 0.55                 | 0.088                |
| Other                                              | 14 (58.3)  | 89 (41.8)    | 55 (49.1)    | 28 (58.3)  | 0.1                  | 0.08                 |
| <b>Extra-intestinal manifestations<sup>¶</sup></b> |            |              |              |            |                      |                      |
| Osteoporosis                                       | 12 (48.0)  | 56 (26.2)    | 19 (17.0)    | 4 (7.8)    | <b>&lt;0.001</b>     | <b>0.002</b>         |
| Osteopenia                                         | 6 (24.0)   | 68 (31.8)    | 38 (33.9)    | 15 (29.4)  | 0.785                | 0.41                 |
| Depression                                         | 0          | 23 (10.7)    | 13 (11.6)    | 5 (9.8)    | 0.37                 | 0.82                 |
| Dermatitis Herpetiformis                           | 1 (4.0)    | 8 (3.7)      | 13 (11.6)    | 2 (3.9)    | <b>0.03</b>          | <b>0.027</b>         |
| <b>Histology and Serology, n=258</b>               |            |              |              |            |                      |                      |
| Marsh ≤ 2 (%)                                      | 0          | 2 (3.7)      | 9 (52.9)     | 6 (35.3)   |                      |                      |
| Marsh 3a (%)                                       | 10 (7.4)   | 22 (16.2)    | 52 (38.2)    | 52 (38.2)  |                      |                      |
| Marsh 3b (%)                                       | 5 (6.6)    | 22 (28.9)    | 29 (38.2)    | 20 (26.3)  |                      |                      |
| Marsh 3c (%)                                       | 2 (6.9)    | 10 (34.5)    | 12 (41.4)    | 5 (17.2)   | 0.15                 | <b>0.014</b>         |
| <b>Tissue transglutaminase<sup>§</sup>, n=254</b>  | 1.7        | 2.2          | 1.97         | 2.02       | 0.42                 | 0.29                 |
| <hr/>                                              |            |              |              |            |                      |                      |
| BMI at diagnosis (kg/m <sup>2</sup> )              | <25        |              | ≥25          |            |                      |                      |
| <b>Bone mineral density<sup>°</sup></b>            |            |              |              |            |                      |                      |
| Median lumbar spine (g/cm <sup>2</sup> ), n=67     | 0.84       |              | 0.95         |            | <b>0.014</b>         |                      |
| Median femur neck (g/cm <sup>2</sup> ), n=55       | 0.79       |              | 0.84         |            | <b>0.012</b>         |                      |
| <b>Serum biomarkers</b>                            |            |              |              |            |                      |                      |
| Median serum haemoglobin (g/dl), n=52              | 13.45      |              | 13.6         |            | 0.65                 |                      |
| Median serum ferritin (µg/l), n=39                 | 23.8       |              | 30.0         |            | 0.20                 |                      |
| Median serum folate (µg/l), n=38                   | 8.9        |              | 6.2          |            | 0.64                 |                      |
| Median serum B12 (µg/l), n=38                      | 313.5      |              | 298.5        |            | 0.55                 |                      |

\* Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into: Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

¶ Manifestations occurring either at celiac disease diagnosis, or prior or posterior to celiac disease diagnosis

§ Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

∞ Bone mineral density at diagnosis of osteoporosis or osteopenia

§ P-value using 4 BMI categories

£ P-value using 2 BMI categories <25 and ≥25 kg/m<sup>2</sup>

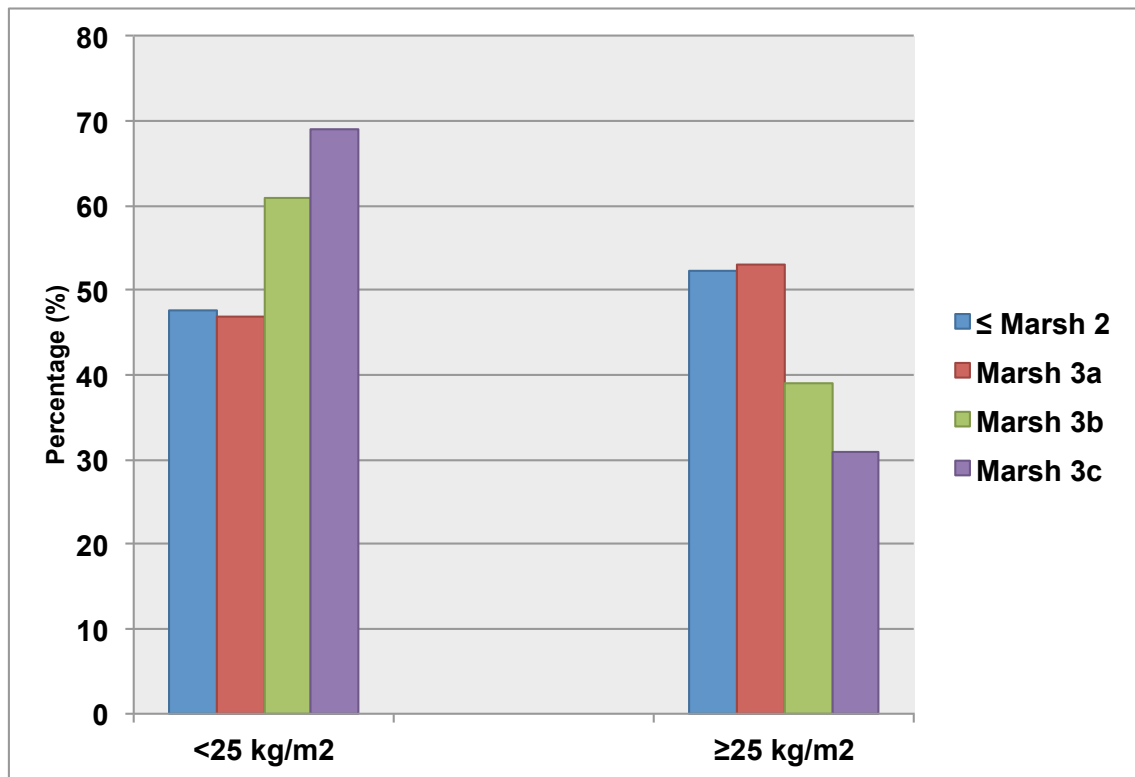
When considering the p-value obtained comparing the 4 BMI categories, overweight and obese patients presented significantly more with the non-classical or subclinical phenotype compared to underweight and normal weight patients (p=0.019). More weight loss was reported at diagnosis by underweight and normal weight patients (p<0.001). This could be due to less malabsorption in overweight and obese patients due to less severe biopsies, or instead, due to lower reporting of weight loss by overweight and obese patients leading to delayed diagnosis in this group because

traditionally weight loss was considered one of the cardinal manifestations at CD diagnosis.

There is a tendency to observe differences in the small bowel damage at diagnosis between BMI groups, with less Marsh 3c observed in overweight and obese patients ( $p=0.15$ ). This becomes a significant tendency when the BMI groups are categorised as  $<25 \text{ kg/m}^2$  and  $\geq 25 \text{ kg/m}^2$ , with less Marsh 3c observed in patients with a BMI at diagnosis  $\geq 25 \text{ kg/m}^2$  versus those with a BMI at diagnosis  $<25 \text{ kg/m}^2$  (23.8% and 37.9% respectively) ( $p=0.028$ ). At diagnosis less patients with a BMI  $\geq 25 \text{ kg/m}^2$  were found to have osteoporosis compared to those with a BMI  $<25 \text{ kg/m}^2$ . This was also observed in the case of dermatitis herpetiformis (DH). There are non-significant tendencies to observe more serum ferritin and less folate at diagnosis in patients  $\geq 25 \text{ kg/m}^2$  at diagnosis when compared to those  $<25 \text{ kg/m}^2$  ( $p=0.20$  and  $P=0.64$  respectively).

In order to reduce the possibility of not detecting real differences between BMIs due to using 4 BMI categories, and the effects of the small size underweight group, we have provided p-values when considering 2 BMI categories ( $<25 \text{ kg/m}^2$  and  $\geq 25 \text{ kg/m}^2$ ). When considering a binary BMI variable, all the previous observed significant changes are still seen, except for the differences in BMI per gender. New significant differences are also observed, such as more steatorrhea at diagnosis found in those  $<25 \text{ kg/m}^2$  compared to those  $\geq 25 \text{ kg/m}^2$ , and interestingly, a significant difference in the degree of damage to the bowel mucosa at diagnosis classified according to Marsh criteria ( $p=0.014$ ) (Figure 2.1).

**Figure 2.1 Degree of damage to the small bowel mucosa classified according to Marsh criteria in patients diagnosed with CD in adulthood classified per BMI category <25 kg/m<sup>2</sup> and ≥25 kg/m<sup>2</sup>.**



The prevalence of Marsh 3b and 3c biopsies at diagnosis was higher in those with a BMI <25 kg/m<sup>2</sup> at diagnosis compared to those with a BMI ≥25 kg/m<sup>2</sup> (69.0% and 31.0% respectively) (p=0.014).

**Table 2.7 Characteristics of the sample diagnosed in adulthood segregated by the age at diagnosis with coeliac disease; ≥18-35 years, >35-64 years and ≥65 years.**

| Characteristic                                                                              | Median or % <sup>†</sup> |                     |                 | P-value          |
|---------------------------------------------------------------------------------------------|--------------------------|---------------------|-----------------|------------------|
|                                                                                             | ≥18-35 years, n=165      | >35-64 years, n=394 | ≥65 years, n=93 |                  |
| <b>Females/Males, n=652 (%)</b>                                                             | 78.2/21.2                | 63.7/36.3           | 65.6/34.4       | <b>0.002</b>     |
| <b>Clinical presentation at diagnosis**</b> , n 602 (%)                                     |                          |                     |                 |                  |
| Classical coeliac disease                                                                   | 59.3                     | 54.8                | 52.2            | 0.51             |
| Non-classical/Subclinical coeliac disease                                                   | 40.7                     | 45.2                | 47.8            |                  |
| <b>Manifestations at diagnosis*</b> , n 602 (%)                                             |                          |                     |                 |                  |
| Diarrhoea                                                                                   | 48.3                     | 41.1                | 41.3            | 0.32             |
| Steatorrhea                                                                                 | 4.8                      | 7.9                 | 5.4             | 0.38             |
| Abdominal Pain                                                                              | 49.0                     | 30.7                | 23.9            | <b>&lt;0.001</b> |
| Constipation                                                                                | 22.1                     | 11.8                | 7.6             | <b>0.002</b>     |
| Vomiting                                                                                    | 10.3                     | 5.8                 | 1.1             | <b>0.01</b>      |
| Bloating                                                                                    | 37.2                     | 22.5                | 15.2            | <b>&lt;0.001</b> |
| Weight Loss                                                                                 | 21.4                     | 29.4                | 34.8            | 0.063            |
| Fatigue/Tiredness                                                                           | 34.5                     | 32.6                | 21.7            | 0.088            |
| Iron Deficiency Anaemia                                                                     | 21.4                     | 29.6                | 39.1            | <b>0.013</b>     |
| Other                                                                                       | 47.6                     | 47.1                | 35.9            | 0.13             |
| <b>Biopsy at diagnosis</b> , n=355 (%)                                                      |                          |                     |                 |                  |
| Marsh ≤ 2                                                                                   | 6.5                      | 5.0                 | 6.7             | 0.865            |
| Marsh 3a                                                                                    | 24.7                     | 20.6                | 18.3            |                  |
| Marsh 3b                                                                                    | 42.9                     | 42.2                | 38.3            |                  |
| Marsh 3c                                                                                    | 26.0                     | 32.1                | 36.7            |                  |
| <b>Tissue transglutaminase<sup>§</sup></b> , n=346 (%)                                      | 2.1                      | 2.01                | 2.06            | 0.51             |
| <b>Presence of coeliac disease extra-intestinal manifestations <sup>λ</sup></b> , n 652 (%) |                          |                     |                 |                  |
| Yes                                                                                         | 53.3                     | 75.4                | 78.5            | <b>&lt;0.001</b> |
| <b>Presence of other immune-mediated conditions</b> , n 652 (%)                             |                          |                     |                 |                  |
| Yes                                                                                         | 24.8                     | 32.5                | 38.7            | 0.055            |

\* Clinical presentation at diagnosis was classified according to the Oslo definition of coeliac disease into; Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

§Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

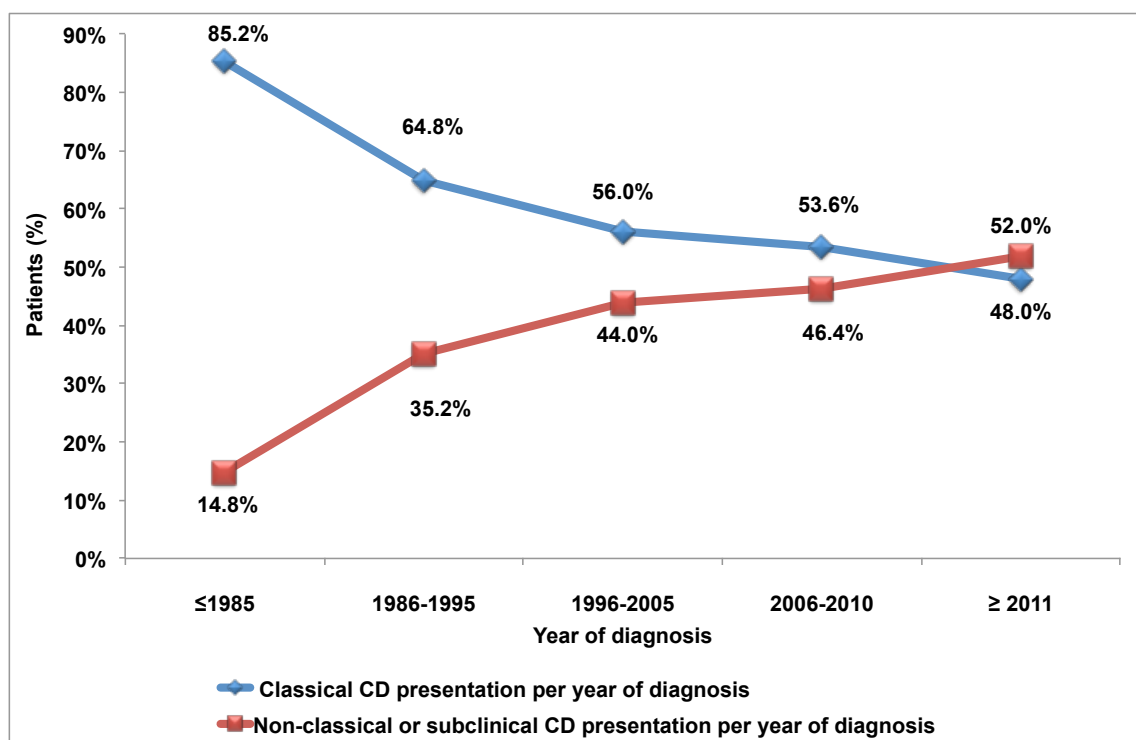
λ Coeliac disease extra-intestinal manifestations considered; osteoporosis/osteopenia, iron deficiency, B12/Folate deficiency and depression

It can be observed that the female/male ratio for those diagnosed  $\geq 65$  years 1.9:1 reducing from 3.7:1 for those diagnosed between 18-40 ( $p=0.02$ ). Coeliac disease is less symptomatic for older patients, who present with less abdominal pain, constipation, vomiting and bloating ( $p=0.001$ ;  $p=0.003$ ;  $p=0.019$  and  $p<0.001$  respectively), and presenting equally to the younger diagnosed groups with diarrhoea and weight loss ( $p=0.32$  and  $p=0.06$  respectively) as the younger diagnosed groups. Patients diagnosed over 65 present significantly more often with IDA ( $p=0.001$ ).

## 2.4.2 Characteristics of the sample over time

In this section we will examine the demographic and clinical characteristics of CD evolution over time as well as changes in CD associated immune-mediated conditions (IMCs).

**Figure 2.2 Changes in clinical presentation of CD at the time of diagnosis in adulthood over 5 time periods (n=602).**



Subjects diagnosed as adults showed a significant decrease in classical CD presentation and a concomitant increase in non-classical/subclinical presentation over time ( $p=0.006$ ) (Fig. 2.2).

**Table 2.8 Clinical characteristics at the time of coeliac disease diagnosis in adulthood and co-existence of immune-mediated conditions over 5 time periods (n=652).**

| Characteristic <sup>λ</sup>                               | n ≤ 1985, n=41  | 1986-1995, n=69  | 1996-2005, n=174 | 2006-2010, n=239 | ≥ 2011, n=128 | p-value           |
|-----------------------------------------------------------|-----------------|------------------|------------------|------------------|---------------|-------------------|
| <b>Gender</b>                                             |                 |                  |                  |                  |               |                   |
| <b>Males</b>                                              | 39.0            | 34.8             | 26.4             | 32.2             | 36.4          | p=0.309           |
| <b>Females</b>                                            | 61.0            | 65.2             | 73.6             | 67.8             | 63.6          |                   |
| <b>Symptoms at Diagnosis (n=602)</b>                      |                 |                  |                  |                  |               |                   |
| <b>Diarrhea</b>                                           | 70.4            | 46.3             | 41.9             | 42.9             | 36.7          | <b>p=0.031</b>    |
| <b>Weight loss</b>                                        | 51.9            | 44.4             | 28.1             | 26.2             | 20.3          | <b>p=0.001</b>    |
| <b>Constipation</b>                                       | 3.7             | 9.3              | 13.1             | 13.7             | 17.2          | p=0.333           |
| <b>Abdominal pain</b>                                     | 48.1            | 31.5             | 32.5             | 36.1             | 30.5          | p=0.424           |
| <b>Overweight or obese (n=391)</b>                        | 13.3            | 13.3             | 34.0             | 50.0             | 48.8          | <b>p&lt;0.001</b> |
| <b>Presence of co-existent immune-mediated conditions</b> | 46.3            | 33.3             | 31.6             | 31.4             | 26.4          | p=0.209           |
| <b>Autoimmune thyroid disease</b>                         | 36.6            | 26.1             | 20.1             | 17.6             | 17.1          | <b>p=0.039</b>    |
| <b>Type 1 diabetes mellitus</b>                           | 2.4             | 1.4              | 4.6              | 4.2              | 3.9           | p=0.802           |
| <b>Psoriasis</b>                                          | 4.9             | 2.9              | 2.3              | 2.1              | 2.3           | p=0.875           |
| <b>Biopsy at diagnosis (n=348)</b>                        |                 | <b>1986-1995</b> | <b>1996-2005</b> | <b>2006-2010</b> | <b>≥2011</b>  |                   |
| <b>Marsh ≤ 2</b>                                          | ND <sup>†</sup> | 8.7              | 3.7              | 4.5              | 9.0           | <b>p=0.003</b>    |
| <b>Marsh 3a</b>                                           | ND              | 13.0             | 9.8              | 27.3             | 23.6          |                   |
| <b>Marsh 3b</b>                                           | ND              | 26.1             | 42.7             | 43.5             | 44.9          |                   |
| <b>Marsh 3c</b>                                           | ND              | 52.2             | 43.9             | 24.7             | 22.5          |                   |
| <b>Tissue transglutaminase* (n=346)</b>                   | ND              | ND               | 2.11             | 2.07             | 1.99          | <b>p=0.019</b>    |

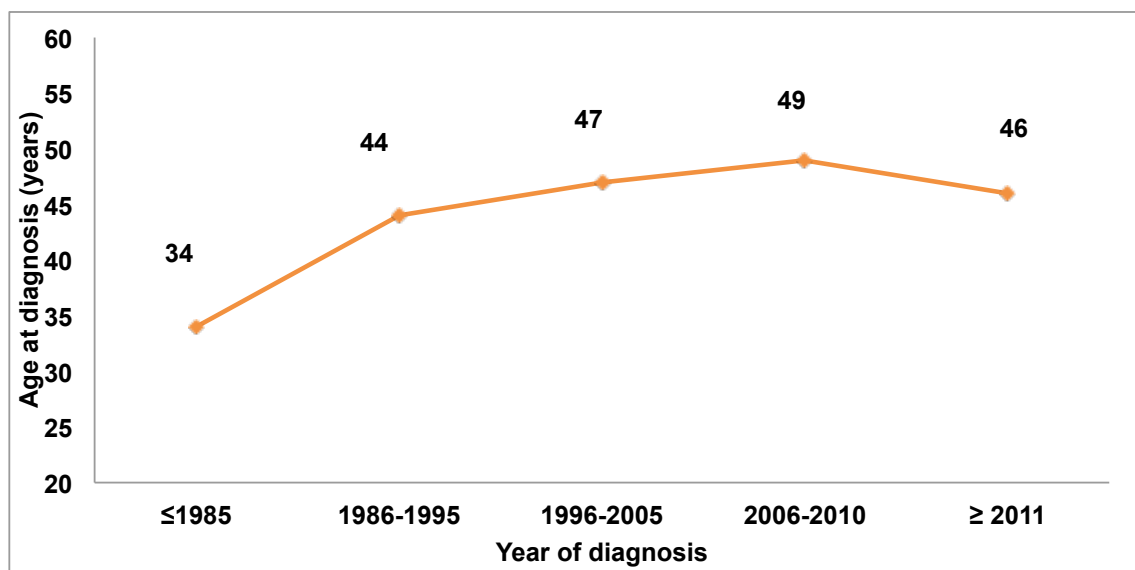
<sup>λ</sup> Valid percentage provided for each characteristic category

\*Tissue transglutaminase antibodies represented as log<sub>10</sub> of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

<sup>†</sup>ND=No or insufficient data

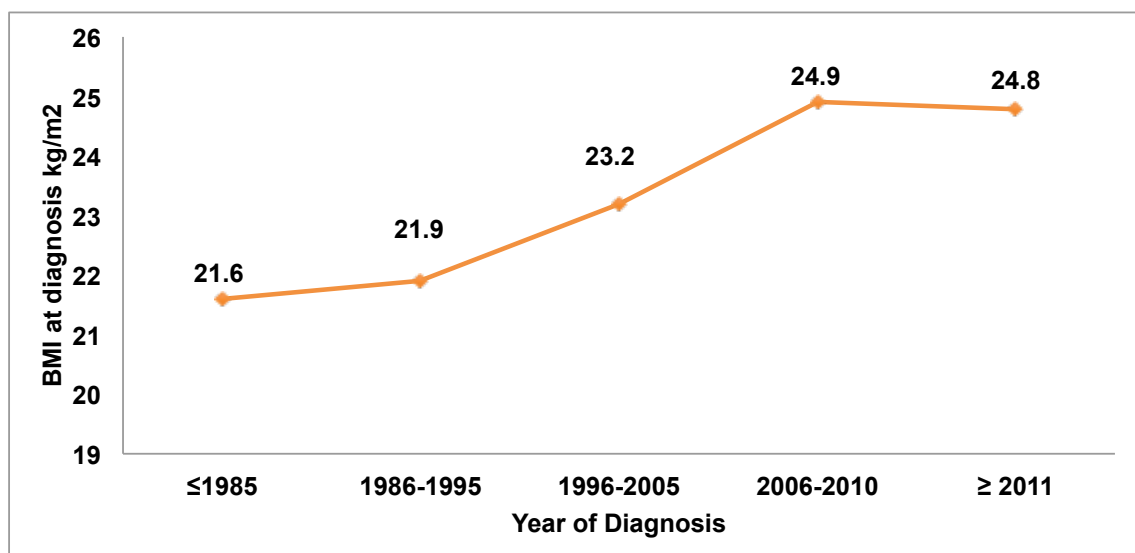
Malabsorption symptoms, such as diarrhoea and weight loss at diagnosis, were commonly reported before 1996, but decreased significantly in the period from 1996 to 2005, with continued decrease thereafter (p=0.031 and p=0.001). Milder damage to the small bowel mucosa (p=0.003) and a significant decrease in the prevalence of associated ATD (p=0.039) was observed over the study period.

**Figure 2.3 Trend analysis of age at diagnosis with CD in adulthood over 5 time periods (n=652).**



The age at diagnosis of CD was significantly lower (34 years of age) before 1986 compared with later periods (44 to 49 years of age) ( $p<0.002$ ).

**Figure 2.4 Trend analysis of BMI at diagnosis with CD in adulthood over 5 time periods (n=402).**



There was a significant increase in BMI over the study period, with median BMI at diagnosis increasing from 21.6 kg/m<sup>2</sup> before 1986, to near 25.0 kg/m<sup>2</sup> in the later periods ( $p<0.001$ ).



### **2.4.3 Comparison of characteristics of the sample segregated by geographical area.**

Our sample can be geographically divided between patients diagnosed in the west of Ireland in Galway University Hospital (GUH), and those diagnosed in the east of Ireland in the rest of the participating centres (supplementary Figure 1, appendix IV). Traditionally the west of Ireland has been an area of high awareness of CD. High incidence of the condition in adulthood was already reported in the 1970's in the paper by Mylotte et al. (1973).

Therefore, in this section, we explored the above reported changes in CD clinical presentation and associated autoimmune conditions per geographical area, to see if differences can be observed. The hypothesis being that potentially higher awareness in the west of Ireland at least in earlier periods, could lead to a different CD picture over time and particularly more non-classical CD having been observed in the earlier time

Table 2.9 Comparison of clinical characteristics and associated immune-mediated conditions between CD patients diagnosed in adulthood in GUH (Table A) and in the east of Ireland (Table B) over 4 time periods.

Table A (n=178)

| Characteristic <sup>λ</sup>                        | n ≤ 1995, n=21 | 1996-2005, n=35 | 2006-2010, n=87 | ≥ 2011, n=35 | p-value           |
|----------------------------------------------------|----------------|-----------------|-----------------|--------------|-------------------|
| Median age at diagnosis (years)                    | 35.0           | 47.0            | 46.0            | 45.0         | <b>p=0.01</b>     |
| Gender                                             |                |                 |                 |              |                   |
| Males                                              | 42.9           | 20.0            | 28.7            | 34.3         | p=0.29            |
| Females                                            | 57.1           | 80.0            | 71.3            | 65.7         |                   |
| Symptoms at Diagnosis (n=167)                      |                |                 |                 |              |                   |
| Diarrhea                                           | 46.7           | 50.0            | 35.3            | 37.1         | p=0.47            |
| Weight loss                                        | 46.7           | 46.9            | 30.6            | 20.0         | p=0.07            |
| Constipation                                       | 6.7            | 9.4             | 12.9            | 8.6          | p=0.82            |
| Abdominal pain                                     | 46.7           | 28.1            | 27.1            | 31.4         | p=0.49            |
| Overweight or obese (n=98)                         | 0.0            | 29.2            | 55.3            | 42.1         | <b>p=0.01</b>     |
| Presence of co-existent immune-mediated conditions | 33.3           | 37.1            | 32.2            | 22.9         | p=0.62            |
| Autoimmune thyroid disease                         | 23.8           | 31.4            | 17.2            | 17.1         | p=0.32            |
| Type 1 diabetes mellitus                           | 0.0            | 2.9             | 3.4             | 2.9          | p=0.86            |
| Psoriasis                                          | 4.8            | 0.0             | 2.3             | 0.0          | p=0.46            |
| Biopsy at diagnosis (n=132)                        |                |                 |                 |              |                   |
| Marsh ≤ 2                                          | 0.0            | 0.0             | 2.7             | 8.0          | <b>p&lt;0.001</b> |
| Marsh 3a                                           | 0.0            | 4.0             | 25.3            | 32.0         |                   |
| Marsh 3b                                           | 14.3           | 56.0            | 57.3            | 60.0         |                   |
| Marsh 3c                                           | 85.7           | 40.0            | 14.7            | 0.0          |                   |
| Tissue transglutaminase* (n=111)                   | ND             | 1.33            | 1.09            | 1.14         | p=0.29            |

<sup>λ</sup> Valid percentage provided for each characteristic category

\*Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

<sup>†</sup>ND=No or insufficient data

Table B (n=474)

| Characteristic <sup>λ</sup>                        | n ≤ 1995, n=89  | 1996-2005, n=139 | 2006-2010, n=153 | ≥ 2011, n=93 | p-value           |
|----------------------------------------------------|-----------------|------------------|------------------|--------------|-------------------|
| Median age at diagnosis                            | 42.0            | 47.0             | 50.0             | 48.0         | <b>p&lt;0.001</b> |
| Gender                                             |                 |                  |                  |              |                   |
| Males                                              | 34.8            | 28.1             | 34.0             | 37.6         |                   |
| Females                                            | 65.2            | 71.9             | 66.0             | 62.4         | p=0.45            |
| Symptoms at Diagnosis (n=167)                      |                 |                  |                  |              |                   |
| Diarrhea                                           | 56.1            | 39.8             | 47.0             | 37.0         | p=0.067           |
| Weight loss                                        | 47.0            | 23.4             | 23.6             | 20.7         | <b>p=0.001</b>    |
| Constipation                                       | 7.6             | 14.1             | 14.8             | 20.7         | p=0.15            |
| Abdominal pain                                     | 34.8            | 33.6             | 40.9             | 30.4         | p=0.37            |
| Overweight or obese (n=304)                        | 16.2            | 34.5             | 47.3             | 49.3         | <b>p=0.002</b>    |
| Presence of co-existent immune-mediated conditions | 39.3            | 30.2             | 30.7             | 26.9         | p=0.31            |
| Autoimmune thyroid disease                         | 31.5            | 17.3             | 17.6             | 17.2         | <b>p=0.03</b>     |
| Type 1 diabetes mellitus                           | 2.2             | 5.0              | 4.6              | 4.3          | p=0.77            |
| Psoriasis                                          | 3.4             | 2.9              | 2.0              | 3.2          | p=0.9             |
| Biopsy at diagnosis (n=223)                        |                 |                  |                  |              |                   |
| Marsh ≤ 2                                          | 8.7             | 5.3              | 6.3              | 9.5          |                   |
| Marsh 3a                                           | 17.4            | 12.3             | 28.7             | 20.6         |                   |
| Marsh 3b                                           | 21.7            | 36.8             | 30.0             | 39.7         |                   |
| Marsh 3c                                           | 52.2            | 45.6             | 35.0             | 30.2         | p=0.29            |
| Tissue transglutaminase* (n=235)                   | ND <sup>†</sup> | 2.04             | 2.06             | 1.82         | <b>p=0.015</b>    |

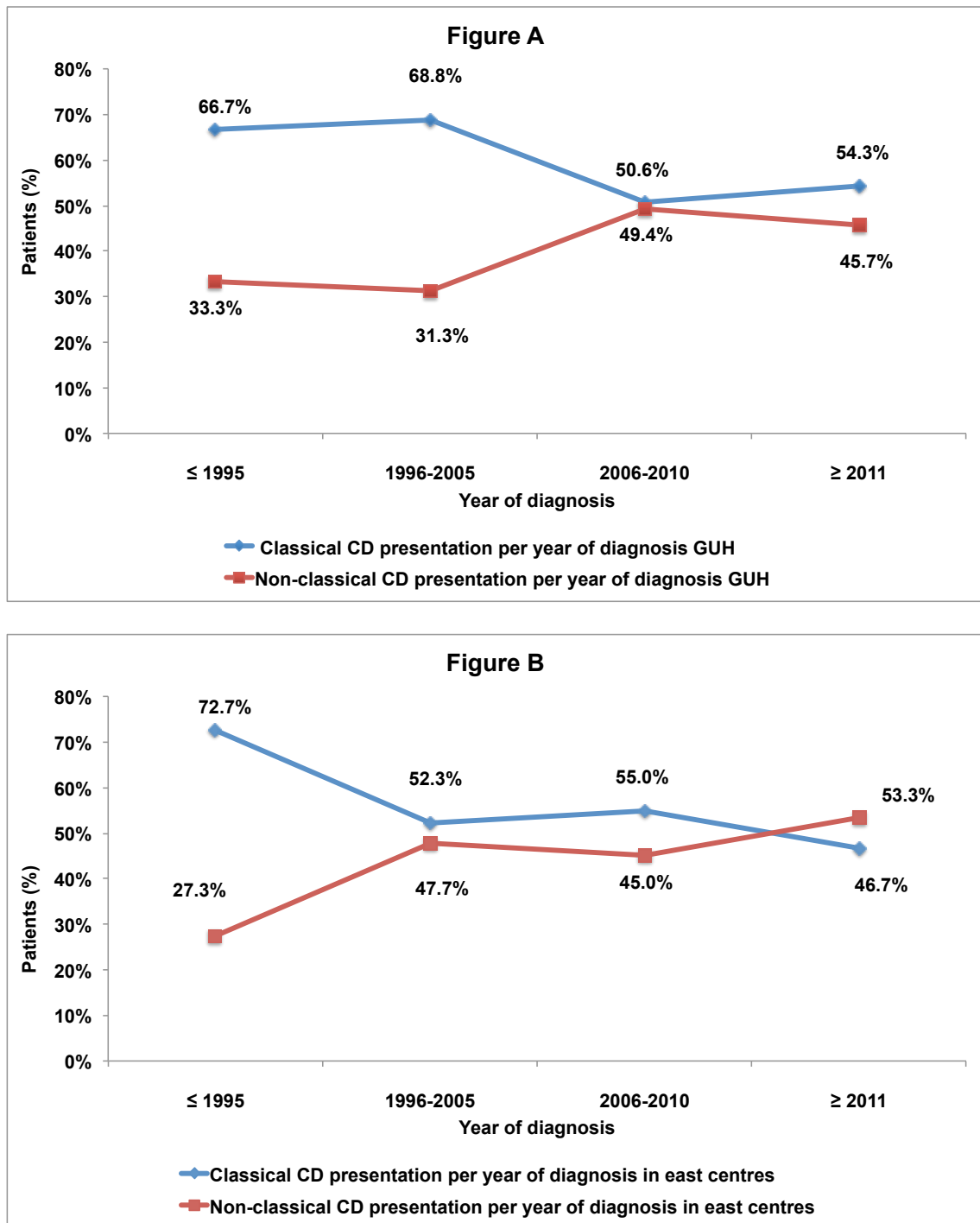
<sup>λ</sup> Valid percentage provided for each characteristic category

\*Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

<sup>†</sup>ND=No or insufficient data

We observe the same trends in the two samples over time. The decreasing prevalence of weight loss over time happens at a later stage in the Galway sample and does not reach statistical significance ( $p=0.07$ ). Co-existent ATD does not significantly decrease over time in the Galway sample, however it decreases significantly in the eastern sample ( $p=0.32$  and  $p=0.03$  respectively). Although biopsies become less severe over time for both samples, this effect does not reach statistical significance in the sample from the east.

**Figure 2.5 Changes in clinical presentation of CD patients at the time of diagnosis in adulthood in Galway University Hospital (GUH) (Figure A) ( $p=0.41$ ), and other participating centres on the east coast (Figure B) ( $p=0.004$ ) over 4 time periods.**



When we examined the trend over time for classical and non-classical CD presentation in the west and east samples, it can be observed how classical CD was predominant in the west in the first two time periods plummeting in the period after 2005 and being approximately as common as non-classical CD after that, whereas in the east it

decreases sharply after 1995 and continues decreasing gradually after that at the same time that non-classical CD increases gradually.

These results, however, are not reliable as we are splitting the sample and therefore losing power, and also we encounter a multiple testing situation. In order to solve the loss of power situation we carried regression analysis in the next table of results (Table 2.10), using time period of diagnosis and geographical place of diagnosis as independent variables with having classical CD as the outcome of interest.

**Table 2.10 Binary logistic regression using time period and geographical location of diagnosis with their unadjusted and adjusted ORs (n=602).**

Time period and geographical location of diagnosis are independent variables in this binary logistic regression, and having non-classical CD at CD presentation in adulthood in the outcome of interest and dependant variable.

| Predictors                                       | Total        | Classical CD | Non-classical or subclinical CD | Unadjusted       |          | Adjusted         |       |       |    |       |
|--------------------------------------------------|--------------|--------------|---------------------------------|------------------|----------|------------------|-------|-------|----|-------|
|                                                  |              |              |                                 |                  |          | 95% CI           |       | P     | OR | P     |
|                                                  |              |              |                                 |                  |          | Lower            | Upper |       |    |       |
| <b>Time period of diagnosis, n 652</b>           | <b>n (%)</b> | <b>n (%)</b> | <b>n (%)</b>                    | <b>OR</b>        | <b>P</b> |                  |       |       |    |       |
| ≤1995                                            | 81 (13.5)    | 58 (71.6)    | 23 (21.4)                       | 0.38             |          | 0.375            | 0.21  | 0.001 |    |       |
| 1996-2005                                        | 160 (26.6)   | 89 (55.6)    | 71 (44.4)                       | 0.76             |          | 0.755            | 0.47  | 0.24  |    |       |
| 2006-2010                                        | 234 (38.9)   | 125 (53.4)   | 109 (46.6)                      | 0.83             |          | 0.84             | 0.54  | 0.43  |    |       |
| ≥2011                                            | 127 (21.1)   | 62 (48.8)    | 65 (51.2)                       | 1.0 <sup>s</sup> | 0.013    | 1.0 <sup>s</sup> |       |       |    | 0.011 |
| <b>Geographical location of diagnosis, n 652</b> |              |              |                                 |                  |          |                  |       |       |    |       |
| East centres                                     | 435 (72.3)   | 240 (55.2)   | 195 (44.8)                      | 1.05             |          | 1.1              | 0.77  | 1.6   |    |       |
| West centre (UHG)                                | 167 (27.7)   | 94 (56.3)    | 73 (43.7)                       | 1.0 <sup>s</sup> | 0.805    | 1.0 <sup>s</sup> |       |       |    | 0.58  |

1.0<sup>s</sup>= Reference group

When we carried out regression analysis using non-classical CD as the outcome of interest, it was observed that the time period is a significant independent predictor both in the unadjusted and adjusted models (p=0.013 & p=0.011 respectively). It is 60% less likely to have non-classical CD when diagnosed ≤1995 in the adjusted model (OR 0.375; 95% CI, 0.21-0.61). On the contrary, the place of diagnosis is not a predictor in the unadjusted or adjusted model (p=0.805 & p=0.58 respectively). Therefore, although GUH is assumed to have had a high awareness of CD, the place of diagnosis is not a significant predictor of the mode of presentation of the condition, but the time period is, suggesting other factors than awareness playing a role in the change of presentation over time.

#### **2.4.4 Changes in BMI and weight over time**

In this section we explored changes in BMI after diagnosis in our cohort of CD patients. For this purpose, we analysed a subcohort of patients (n=166) for whom we had both the BMI at diagnosis and a year after diagnosis. We also analysed a small subcohort of patients (n=27) for whom adherence to the diet was assessed retrospectively by a coeliac expert dietician. Finally, we also investigated changes in BMI over time in our CD cohort compared to the general population using data from the Survey on Lifestyle and Attitudes to Nutrition (SLAN).



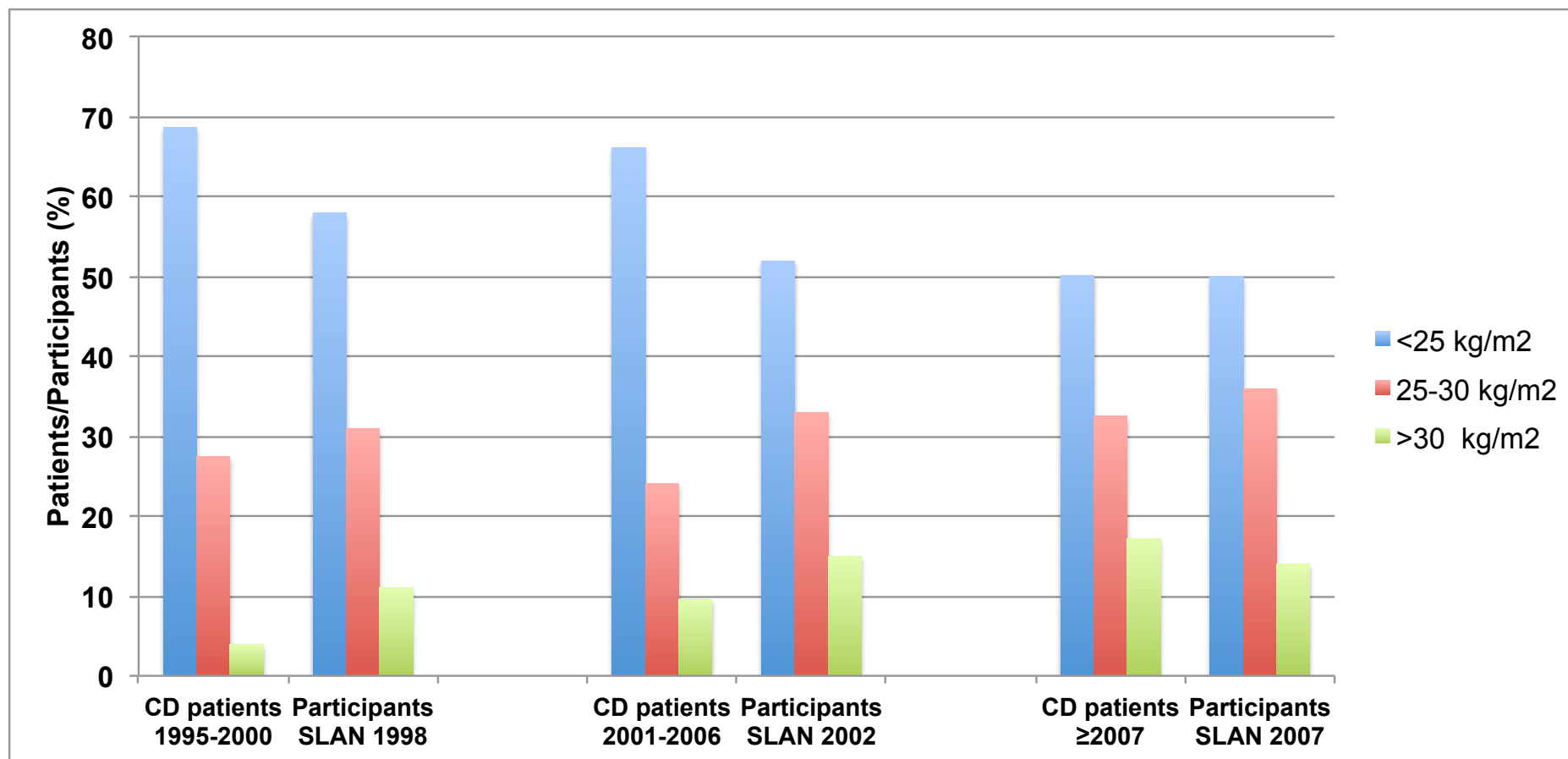
**Table 2.11 Coeliac patients BMI at diagnosis and at follow-up one year after diagnosis (kg/m<sup>2</sup>).**

| <b>Patients</b>     | <b>Median BMI at diagnosis</b> | <b>Median BMI 1 year after diagnosis</b> | <b>P-value</b>   |
|---------------------|--------------------------------|------------------------------------------|------------------|
| Whole sample, n=166 | 24.1                           | 24.8                                     | <b>0.012</b>     |
| <18.5, n=12         | 18.3                           | 19.1                                     | <b>0.008</b>     |
| 18.5-24.9, n=85     | 22.4                           | 23.6                                     | <b>&lt;0.001</b> |
| 25-29.9, n=44       | 26.6                           | 26.25                                    | 0.108            |
| ≥30, n=25           | 31.4                           | 31.6                                     | 0.056            |

Considering the whole sample there was a significant increase in BMI one year after diagnosis ( $p=0.012$ ). Considering each BMI category individually, patients presenting underweight and with normal weight significantly increased their BMI a year after diagnosis ( $p=0.008$  &  $p<0.001$  respectively), whereas those who presented overweight or obese had no significant changes in their BMI a year after diagnosis ( $p=0.108$  &  $p=0.056$  respectively).

**Figure 2.6 Changes in CD patients BMI at diagnosis from 1995 up to 2015 compared to the general population.**

Data for the general population is derived from the Survey of Lifestyles Attitudes and Nutrition (SLAN) 1998, 2002 and 2007.



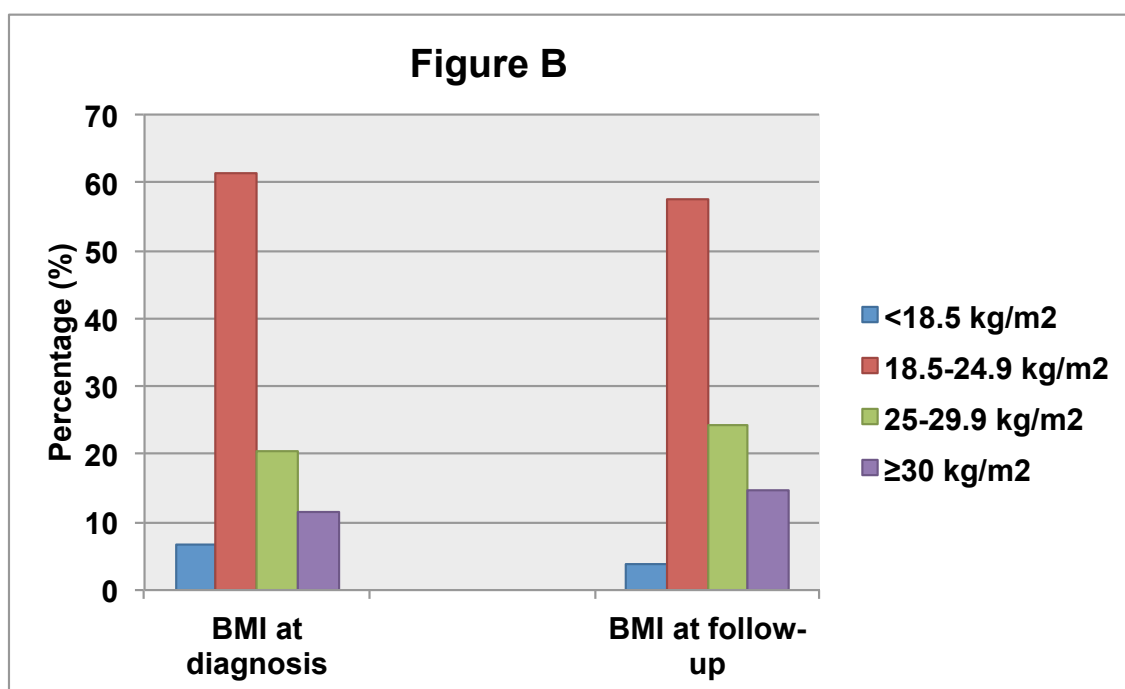
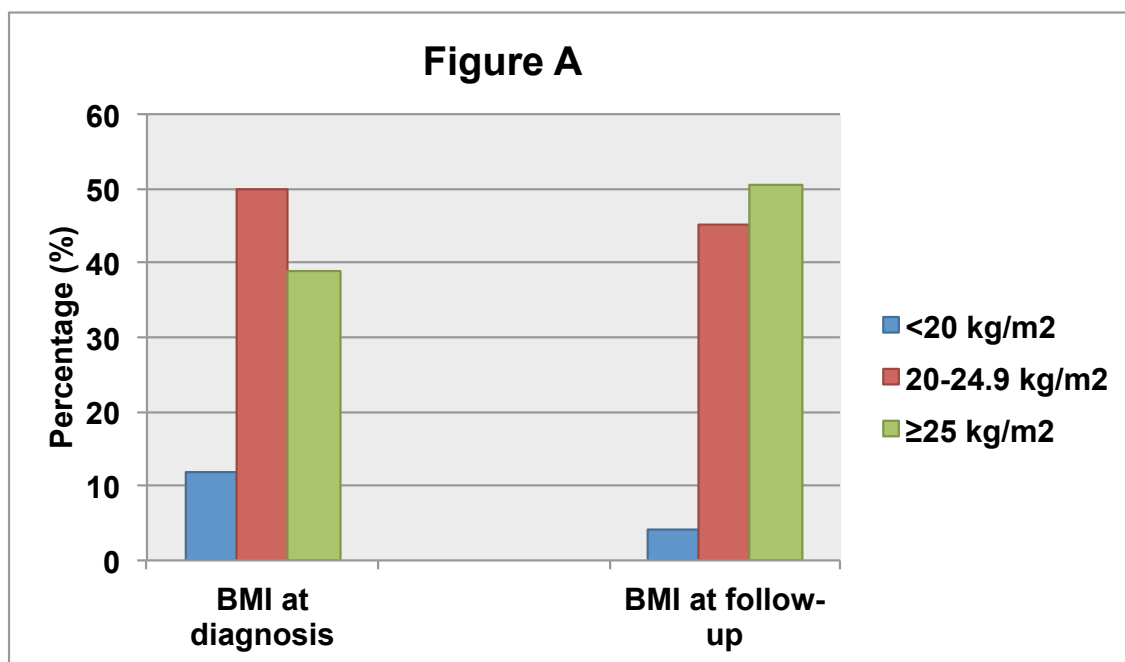
In the first two periods it can be observed how the percentage of overweight and obese patients is smaller in the CD patient cohort compared to that of the general population in SLAN (31.4% and 42.0% respectively in 1995-2000; 33.7% and 48.0% respectively in 2001-2006). However, in the last time period the percentage of overweight and obese people is practically identical in both the CD patient cohort and the SLAN participants (49.8% and 50.0% respectively), with more obese present in the CD population compared to the general population (17.2% and 14.0% respectively).

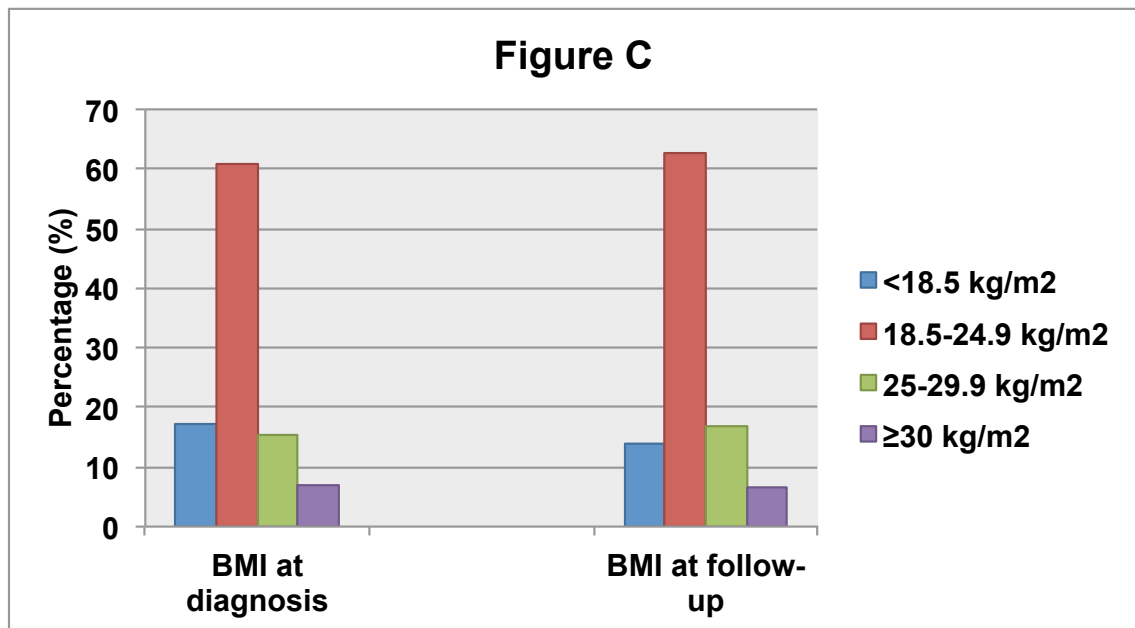
**Table 2.12 BMI and weight changes in CD patients 1 year after diagnosis (n=166).**

| Initial BMI category | No of Patients (%) | Patients Increasing Weight (%) | Median Weight Increased (kg) | BMI Category after 1 year on GFD |            |            |            |
|----------------------|--------------------|--------------------------------|------------------------------|----------------------------------|------------|------------|------------|
|                      |                    |                                |                              | <18.5                            | 18.5-24.9  | 25-29.9    | ≥30        |
| <18.5                | 12 (7.2)           | 11 (91.7)                      | 2.1                          | 2 (16.7%)                        | 10 (83.3%) | NP         | NP         |
| 18.5-24.9            | 86 (51.6)          | 65 (76.4)                      | 2.5                          | 2 (2.4%)                         | 69 (81.2%) | 12 (14.1%) | 2 (2.4%)   |
| 25-29.9              | 44 (26.3)          | 18 (40.9)                      | 1.5                          | NP                               | 6 (13.6%)  | 38 (86.4%) | NP         |
| ≥30                  | 25 (15.0)          | 8 (32.0)                       | 3.65                         | NP                               | NP         | 7 (28.0%)  | 18 (72.0%) |

76.4% of the patients who presented in the normal BMI category increased their weight, with 16.5% of these patients moving into the ≥25 kg/m<sup>2</sup> category. A large proportion of overweight and obese patients at diagnosis increased weight, and the majority remained in the ≥25 kg/m<sup>2</sup> category.

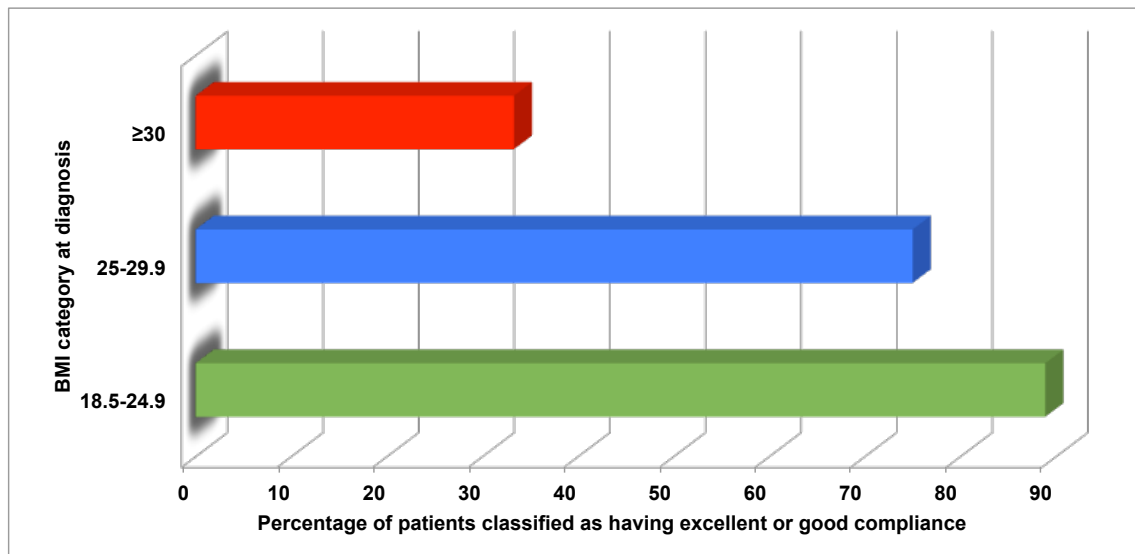
Figure 2.7 Figure A, CD patients BMI at diagnosis (n=265) and at follow-up (2 years) in the northern Ireland paper by Dickey *et al.* (n=188); Figure B, CD patients BMI at diagnosis and at follow-up in the US paper by Kabbani *et al.* (mean 3.3 years) (n=679); Figure C, CD patients BMI at diagnosis (n=369) and at follow-up (2 years) in the paper by Cheng *et al.* (n=309).





Whereas both Dickey and Kabbani *et al.* report an increase in the overweight and obese categories at follow-up, Cheng *et al.* reported a beneficial effect of the GFD on BMI, with underweight patients gaining weight and overweight and obese patients losing weight.

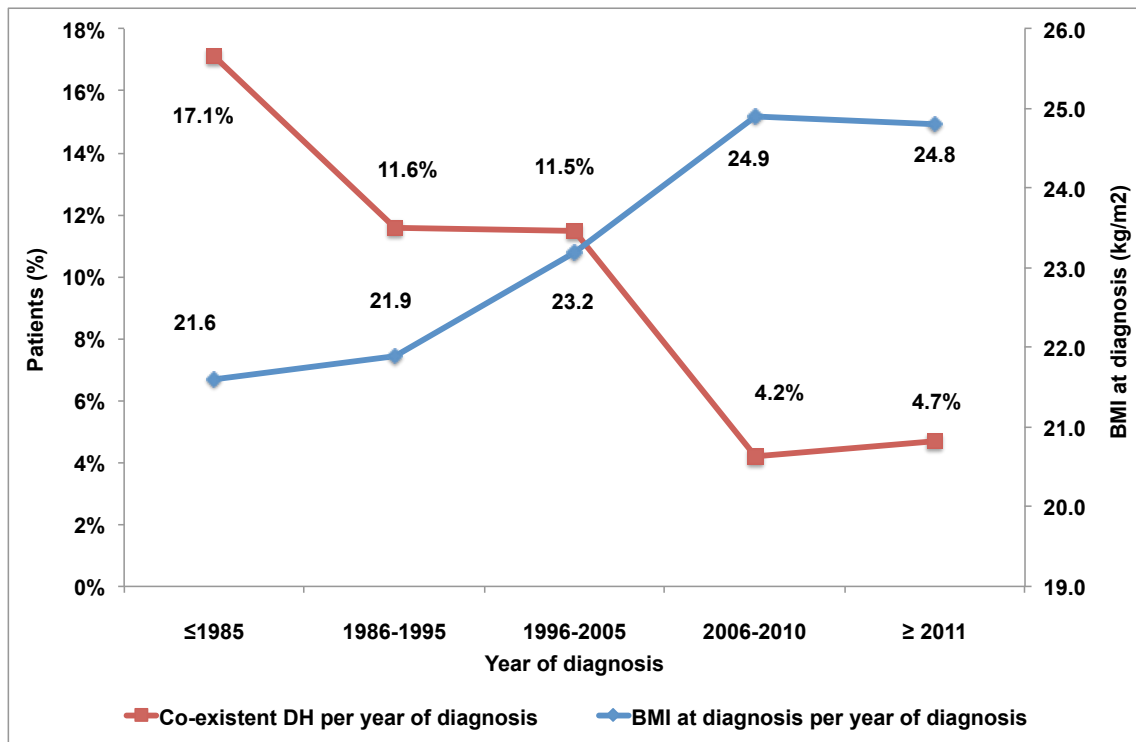
**Figure 2.8 Percentage of patients classified as having excellent or good adherence to the diet a year after diagnosis with CD classified per BMI category at diagnosis (n=27).**



29 patients were assessed retrospectively for compliance with the GFD a year after diagnosis; 17 classified as excellent, 3 as good, 5 as fair and 4 as poor. Of these, 27 had their BMI at diagnosis available. A larger percentage of patients with a BMI in the normal range at diagnosis were classified as having excellent or good compliance compared to overweight and obese patients ( $p=0.062$ ). This finding suggests that those CD patients presenting overweight or obese at diagnosis have a tendency to be less adherent to the diet when assessed a year after diagnosis. However, this finding should be taken with care given the small sample of patients and the lack of statistical significance.

Due to the higher prevalence of DH in overweight and obese patients at CD diagnosis observed in this study and previously reported in other studies (Zingone et al., 2009), and the increase in BMI at diagnosis over time also reported in this study (Figure 2.4), we thought it was interesting to explore whether this increase in BMI in CD patients at diagnosis has coincided with the presence of more DH (Figure 2.9).

**Figure 2.9 CD patients diagnosed in adulthood BMI at diagnosis and DH co-existence over five time periods.**

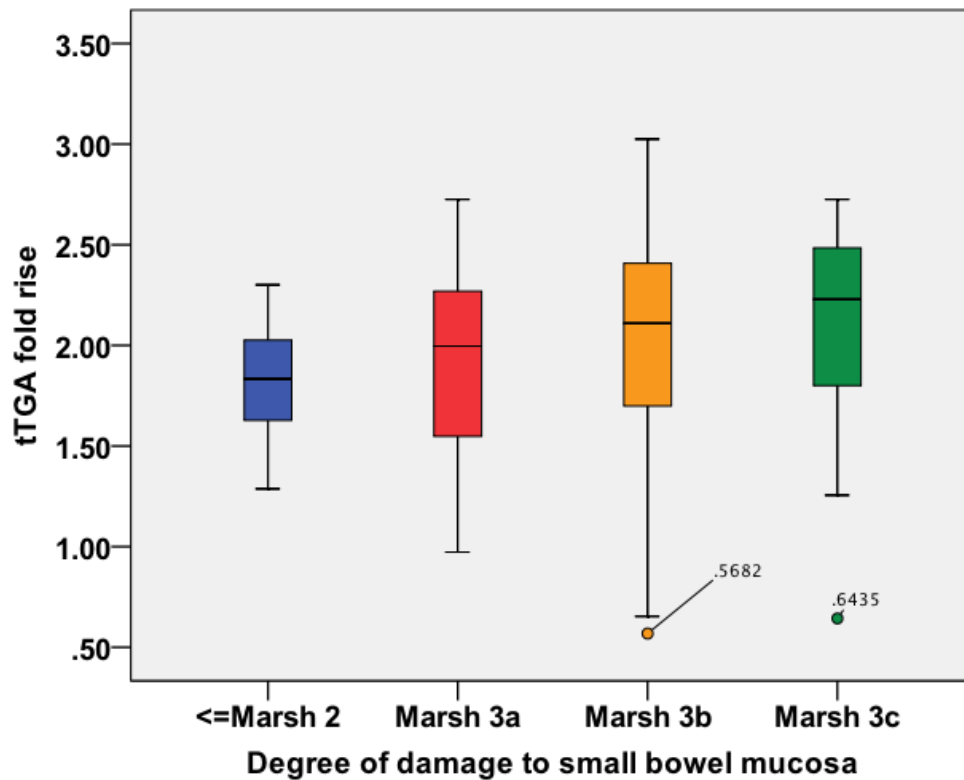


Despite of the observed significant increase in BMI at CD diagnosis observed (Figure 2.4), this does not translate in a higher prevalence of DH over time, on the contrary, co-existent DH decreases significantly over time ( $p=0.003$ ).



## 2.4.5 Correlation of serology with biopsy at diagnosis

Figure 2.10 Median tissue transglutaminase antibodies fold rise in patients with coeliac disease having different levels of small bowel damage classified according to Marsh criteria.



During the time period covered by this study, tTGA value calculations at diagnosis were determined using a variety of tTGA diagnostic tests with different cut-off points. In order to compare results from different tests, tTGA values were expressed as the fold increase above the cut-off value by calculating the ratio of the observed tTGA at diagnosis divided by the cut-off value for an equivocal result. The ratios derived to compare tTGA values at diagnosis correlate well with the degree of small bowel damage at diagnosis classified according to Marsh criteria ( $p < 0.002$ ) (Figure 2.10).

## 2.5 Discussion

This study describes the clinical features of CD in a large cohort, and represents the only study of its kind in Ireland. Covering a period from 1960-2015, it provides a valuable 55 year perspective on the clinical features of the condition in a region where it has historically been well identified and has had a high rate of diagnosis.

### 2.5.1 Changes in CD clinical presentation over time

While our sample confirms that CD can be diagnosed over a wide age range, a significant increase in the median age at diagnosis happened between 1986 and 2005 coinciding with improved diagnostic techniques (Reilly et al., 2012, Volta et al., 2014, Kocsis et al., 2015, Kivela et al., 2015). The median age at diagnosis seems to have plateaued thereafter in the late 40's (Figure 2.3) which is higher than what has been reported in similar studies internationally, but in line with previous studies in Ireland (Tajuddin et al., 2011, Saleem et al., 2012, Kocsis et al., 2013, Volta et al., 2014).

A continuous gradual decrease in classical presentation mirrored by an increase in non-classical/subclinical presentation was observed over time (Figure 2.2), corroborating findings from previous international studies (Rampertab et al., 2006, Ravikumara et al., 2006, Telega et al., 2008, Volta et al., 2014, Kivela et al., 2015, Schosler et al., 2016). This change in clinical presentation was accompanied by a significant decrease in diarrhoea and weight loss at CD diagnosis (Table 2.8). Ludwigsson *et al.* found that whereas newly diagnosed patients presenting with both diarrhoea and weight loss had decreased in the period between 2000-2010, the absolute incidence of diarrhoea had not decreased (Ludvigsson et al., 2013b). They suggest that the decrease in diarrhoea and weight loss could be due to higher awareness which has resulted in testing of new risk groups (i.e. patients with co-existent immune-mediated conditions or patients with osteoporosis), relatives of CD patients and a higher use of serology which would diagnose patients earlier thus preventing the development of classical symptoms (Ludvigsson et al., 2013b).

Concurrently to these changes in clinical presentation, a less severe damage to the small bowel mucosa is observed at diagnosis in the later time periods, which, as expected, is reflected in lower tTGA levels (Table 2.8). tTGA levels have been previously found to correlate well with the level of damage to the small bowel mucosa and we also see this correlation (Singh et al., 2014, Rahmati et al., 2014) (Figure 2.10). Comparing biopsies and serology at diagnosis, adults have a less severe presentation and a higher prevalence of non-classical symptoms compared to children (Vivas et al.,

2008, Volta et al., 2014). Taken together, this may imply that adults with unspecific or non-classical symptoms are being finally diagnosed with CD later in life due to better diagnostics and a higher awareness of CD (and particularly its variable expression in adults) by physicians (Smukalla et al., 2014). This may partly explain the large percentage of our patients diagnosed in adulthood as well as the increased median age of diagnosis and changed clinical presentation over time. However, the fact that Ireland has traditionally been an area of high awareness and diagnosis of the disease, raises the intriguing possibility of a biological shift in CD, perhaps linked to variations in environmental exposures, contributing to an increased cadre of patients with milder disease and older age at onset.

Our results on the changes in CD clinical presentation in adults over time (Table 2.8) correlate with those of Kivela *et al.* on the paediatric population. These authors report that this change towards a milder presentation, both clinically and histologically, has reached a plateau in Finland due to changes in the diagnostics of the disease, but also, potentially, to modification of environmental factors. Improved diagnostic techniques alone cannot explain the well reported increase in incidence observed in CD, as suggested by studies analysing temporal trends in CD incidence (Murray et al., 2003, Ludvigsson et al., 2013b). For instance, Ludvigsson *et al.* showed that following the introduction of better serological screening tools in North America circa the year 2000 (endomysial and tissue transglutaminase antibodies), CD incidence did not decrease as would be the expected if the new pool of cases consisted merely of previously undiagnosed cases (Ludvigsson et al., 2013b). Moreover, disease incidence can vary considerably between geographically close areas or even between socioeconomic groupings within the same country (Kondrashova et al., 2008, Kivela et al., 2015, Ramakrishna et al., 2016). Thus, the prevalence of CD is lower in Russian Karelia than in Finland, two regions whose populations share similar genotypes and cereal consumption (Kondrashova et al., 2008).

A later study comparing the cumulative incidence of CD in Finland and Estonia in a cohort of children also found a higher five year cumulative incidence in Finland compared to Estonia, even when the frequency of CD predisposing HLA alleles was slightly higher in the Estonian children (Simre et al., 2016). These observations point to a true rise in the incidence of CD presumably due to altered environmental factors rather than solely improved detection (Ludvigsson et al., 2013b). While there are no recent studies in Ireland on the prevalence and incidence of CD and its evolution over time, both the current incidence and prevalence are suspected to be high (Mylotte et al., 1973). Studies of prevalence and incidence in Ireland would be desirable to

determine if these rates are changing and in order to see how they correlate with those of other countries.

### **2.5.2 Gender differences in CD diagnosis and clinical presentation**

CD was more prevalent in female patients in our population confirming previous observations (Kocsis et al., 2013, Volta et al., 2014). Conversely, population based screening studies suggest a more equal CD gender distribution (Ludvigsson et al., 2013b). However, despite improvements in diagnosis, the female/male ratio has remained stable over time in our cohort. Male patients were diagnosed later in life than female patients which has been previously reported in Ireland (Tajuddin et al., 2011). Potential explanations for this age difference at diagnosis are, women being more likely to require health care during their fertile years, and/or gender-dependent differences in clinical presentation, including the presence of co-existent IMCs (Ludvigsson et al., 2013b, Vivas et al., 2015). Indeed, a greater proportion of women presented with certain intestinal and extra-intestinal manifestations, as well as more co-existent IMCs, supporting the latter contention and suggesting that CD presentation and prognosis was more severe in women.

### **2.5.3 CD and associated T1DM and ATD over time**

Our study confirms the association of CD with other IMCs, predominantly ATD and T1DM which has been reported previously (Gutierrez-Achury et al., 2011, Lundin and Wijmenga, 2015). ATD prevalence has decreased significantly over time in patients diagnosed in adulthood. Some studies have suggested tTGAs in CD patients may promote the development of ATD (Naiyer et al., 2008, Duntas, 2009). It could be hypothesised that with better diagnostic tools and screening of at risk patients, CD has been diagnosed at an earlier stage, thus avoiding long term exposure to tTGAs and decreasing the prevalence of ATD. Meanwhile, T1DM showed a non-significant increase in the past decade. Kivela *et al.* recently reported a similar finding in children, suggesting that the increase in T1DM is due to intensified screening for CD in T1DM patients, and the decrease in ATD could be due to the disappearance of iodine deficiency in CD patients (Kivela et al., 2015).

The explanation for the increase in T1DM is also plausible in our cohort. However, Ireland has traditionally being an area of borderline iodine deficiency, and recent studies indicate that this has not changed (Smyth et al., 1988, Nawoor et al., 2006, Smyth and O'Herlihy, 2012). Therefore, disappearance of iodine deficiency does not explain the decrease of ATD in our cohort, moreover iodine deficiency does not explain

the occurrence of autoimmune ATD (Miranda et al., 2015, Duntas, 2015). Thus, it is possible that morphing environmental factors are predisposing individuals to develop CD who have less exposure to ATD genetic susceptibility factors than heretofore, analogous to our scenario of a biological shift favouring milder, later onset disease.

#### **2.5.4 CD differences per geographical location of diagnosis**

As mentioned in the methods section, the sample in the present study can be geographically divided between CD patients diagnosed in the west of Ireland in Galway University Hospital (GUH), and CD patients diagnosed in several secondary referral centres situated in Dublin in the east of Ireland (supplementary Figure 1, appendix IV). Traditionally the west of Ireland has been an area of high awareness of CD. High incidence of the condition in adulthood was reported in the early 1970's by Mylotte *et al.* (1973). University Hospital Galway's (UHG) paediatric unit has had a special interest in CD since its beginnings in 1956, and the gastrointestinal unit established in 1969 has always been aware of adult CD (Mylotte et al., 1973). Therefore, if the changes in clinical presentation observed over the 55 year period are mainly due to increased awareness of CD, they might not be so pronounced in the GUH sample when compared to the sample from the east of the country.

Initially, when we compare the demographics and clinical presentation characteristics between east and west, we observed a lower prevalence of Marsh 3c and depression in GUH patients ( $p < 0.001$  and  $p = 0.01$  respectively) (Table 2.5). This could potentially be indicative of increased awareness and therefore earlier diagnosis of the condition (Jackson et al., 2012). In order to try to prove differences in clinical presentation over time between UHG and the centres in the east of Ireland, we carried out the same analysis over time as in Figure 2.2 and Table 2.8 independently using the CD sample diagnosed in Galway and the one diagnosed in the east. In Table 2.9A, it can be observed that when the two samples are analysed separately the significant decrease in diarrhoea, weight loss and co-existent ATD reported using the full sample is lost in the Galway cohort ( $p = 0.47$ ;  $p = 0.07$  &  $p = 0.32$  respectively). This is not the case in the east, where weight loss and co-existent ATD decrease significantly over time and diarrhoea shows a nearly significant tendency to decrease as well ( $p = 0.067$ ;  $p = 0.001$  &  $p = 0.03$  respectively) (Table 2.9B). However, although all these changes are lost in the Galway cohort still a decreasing tendency is observed over time (Table 2.9A).

When we looked at the clinical presentation over time using the Galway sample, we can see that classical presentation is predominant for a longer, decreasing sharply in the period after 2005, and reaching a plateau thereafter, however this is non-significant

(Figure 2.5A) ( $p=0.52$ ). In the case of those patients diagnosed in the east, classical presentation drops significantly in the period after 1995 reaching a plateau thereafter (Figure 2.5 B) ( $p=0.006$ ). It could be hypothesised that higher awareness of CD in the west of Ireland results in non-significant changes over time, with classical and non-classical CD reaching a plateau in the later periods due to environmental factors.

On the other hand, in the east of Ireland, increasing awareness over time, together with environmental factors, could have led in the east of Ireland to a significant and continuous decrease of classical CD accompanied by an increase in non-classical CD (Figure 2.5 A & B). However, the most likely explanation for the difference in findings between the two samples could also be the decreased power of the analysis due to the segregation of the sample, and also the lower numbers of patients in the Galway sample, especially in the earlier periods with only 15 patients in the period  $\leq 1985$  and 6 in the period between 1986-1995. These numbers are unlikely to constitute a representative sample of the clinical presentation of CD during those periods in Galway. This previous approach to capture potential changes per geographical area is also indirect as we are separating both samples and leads to the problem of multiple testing.

To salvage the previous problems of loss of power and multiple testing, and analyse the changes in clinical presentation over time per geographical area using the whole sample diagnosed in adulthood, we carried out regression analysis controlling for geographical place of diagnosis and time period of diagnosis and having non-classical CD as the outcome of interest (Table 2.10). No differences can be observed between east and west coast in the regression analysis (Table 2.10) ( $p=0.58$ ). However, the time period of CD diagnosis is a significant predictor of having classical or non-classical CD; one is 60% less likely to have non-classical CD when diagnosed  $\leq 1995$  in the adjusted model (OR 0.375; 95% CI, 0.21-0.61) ( $p=0.011$ ) (Table 2.10). Therefore, although GUH is assumed to have had a high awareness of CD, we see no evidence that the place of diagnosis predicts the mode of presentation, while the time period does, suggesting factors other than awareness are playing a role in the change of presentation over time.

The main hypothesis regarding environmental factors predisposing to the development of CD are based on infant feeding practices which were identified as the origin of the Swedish CD epidemic from 1985-1994 (Ivarsson et al., 2000). The epidemic only affected children younger than two years. In this cohort, CD incidence tripled from 65 cases per 100,000 person-years to 198 cases per 100,000 person-years. In contrast, incidence data remained stable in older children during the same time period. Although

the epidemic ended in 1995 after changes on infant feeding practices were established, the cohort of children born during this period have a continuing increased risk of developing CD (Ivarsson et al., 2000, Ivarsson, 2005). Timing of gluten introduction, gluten load in the diet due to increased gluten content of formula milk, and low breastfeeding rates were the factors identified to be involved in the origin of the epidemic. However, recent prospective studies indicate that, of all the previously identified factors, the only one involved in increasing CD risk in genetically predisposed infants is the amount of gluten consumed up to two years of age (Andren Aronsson et al., 2016b).

There are no studies on the effects of high gluten content of the diet in adults predisposed to CD. However, gluten consumption has tripled in the United States since 1977 which correlates with the time of increased CD incidence (Kasarda, 2013). This increased consumption is not only due to more ingestion of gluten rich foods such as pasta and bread, but also potentially to changes in wheat breeding which have led to increased protein content, and also to a greater density of epitopes capable of stimulating T-cells in the context of CD (Molberg et al., 2005, Kasarda, 2013, Ludvigsson et al., 2013b). Therefore, potentially increased gluten content in the diet could lead to the development of CD in adults genetically predisposed to it.

### **2.5.5 CD clinical presentation in the elderly**

A considerable proportion (12.4%) of our sample were  $\geq 65$  years at diagnosis (Table 2.7). A previous study has reported similar prevalence of CD patients diagnosed in the elderly ( $\geq 65$  years) and those diagnosed less than 18 years of age. We report the same finding in our sample, with 13.0% of patients having been diagnosed less than 18 years (Table 2.1). The incidence of CD in those over 60 years of age was reported to be 4% in 1960 whereas it ranges nowadays from 19-34% (Rashtak and Murray, 2009). Some studies have found a higher prevalence and incidence of CD in the elderly when compared to adolescents and adults, which could be partly be as a result of diagnostic delay due to a more subtle or non-classical presentation of the disease in this population group, as previously pointed out (Freeman, 2008, Vilppula et al., 2009).

In the US, It has also been reported that large percentages of elderly patients with CD have been misdiagnosed with irritable bowel syndrome delaying the correct diagnosis for an average of 17 years (Patel et al., 2005). Apart from a milder clinical presentation of the condition, other factors could result in delayed diagnosis in this age group, such as a low level of suspicion by the physician who would first consider other age-related causes as the origin of the clinical manifestations, for example colon cancer

investigations triggered by IDA in the patient (Freeman, 2008). A true increase in CD incidence in the elderly (adults over 55), taking account of delayed diagnosis, has been reported by a study conducted over a three year period in Finland (Vilppula et al., 2009).

Different studies have used various age cut-off points to define CD diagnosed in old age ranging from 55-65 years of age (Freeman, 2008, Vilppula et al., 2009, Casella et al., 2012). For this reason, we have analysed our data with different cut-off points to include the elderly diagnosed as both over 55 and over 65 years of age (supplementary Table 3 in appendix VIII and Table 2.7 respectively). When considering the elderly as those diagnosed over 65 years of age, we observe a less variable CD presentation at diagnosis when considering its intestinal manifestations, with patients presenting with less abdominal pain, constipation, vomiting and bloating and presenting equally to the younger diagnosed groups with diarrhoea and weight loss (Table 2.7). This fact has been previously reported and it has been suggested that the paucity of signs and symptoms in elderly CD patients could be due to a less extensive mucosal damage in the duodenum and jejunum in this group (Freeman, 2008, Rashtak and Murray, 2009).

Freeman et al. also reported that although weight loss and diarrhoea do occur at CD presentation in these patients, often the only manifestation at diagnosis would be IDA (Freeman, 2008). We observe no differences in biopsy at diagnosis classified according to Marsh criteria in the over 65s compared to younger groups. However, the Marsh criteria describes the immune response in the epithelium and the degree of architectural changes in the mucosa, but not the extent of the lesion (Dickson et al., 2006). We also observe diarrhoea and weight loss happening in this group of patients equally to the younger groups, however we did not collect data on the severity of the diarrhoea and weight loss. It could be that these symptoms vary in severity between the two age groups. A significant difference in IDA is observed which is more common in this age group (Table 2.7). Therefore, the hypothesis of less extensive damage to the mucosa resulting in minimised signs and symptoms and a higher prevalence of IDA is possible in our cohort of patients over 65 (Freeman, 2008).

Similar findings apply to the cohort over 55, the only difference being that a significantly higher prevalence of weight loss is observed in this cohort (supplementary Table 3, appendix VIII). In both the over 55 and over 65 cohorts, a significantly higher prevalence of extraintestinal manifestations, such as osteoporosis and osteopenia, can be observed. The risk of having osteoporosis and osteopenia increases with age and female gender, and delayed CD diagnosis could exacerbate matters as suggested by studies which have found a correlation between tTGA seropositivity and the risk of hip



fracture (Hernlund et al., 2013, Heikkila et al., 2015b). Therefore, potentially delayed CD diagnosis could work synergistically with age and gender predisposing elderly CD patients to bone disease. This theory cannot be proven with our data as the time between the onset of symptoms until diagnosis was not collected.

IDA is present in a large percentage of elderly patients, and it is the main micronutrient deficiency observed in these patients at CD diagnosis, although folate and B12 deficiency also account for a small percentage of anaemias in this group (Hankey and Holmes, 1994, Freeman, 2008, Bode and Gudmand-Hoyer, 1996, Dahele and Ghosh, 2001, Rashtak and Murray, 2009). These deficiencies have been mainly attributed to the malabsorption of micronutrients due to bowel damage, however it is believed that anaemias in coeliac disease are multifactorial with systemic inflammation being a contributing factor to their development (Harper et al., 2007). Harper *et al.* found that 12-13% of CD patients with anaemia presented with higher serum ferritin (an acute phase protein which is raised during systemic inflammation) levels than expected for their age and gender. Moreover, in the same study it was found that when serum ferritin and erythrocyte sedimentation rate (ESR) were correlated, high ESR was observed in those who had low or high ferritin. High ferritin together with high ESR suggest systemic inflammation without malabsorption and anaemia of chronic disease in these patients (Harper et al., 2007, Bermejo and Garcia-Lopez, 2009). The malabsorption hypothesis is supported in our data by the fact that IDA is more common in the group with Marsh 3b/3c biopsy at diagnosis. However, we cannot rule out that some of our patients who presented with IDA had anaemia of chronic disease because, ideally, we should have explored serum ferritin levels in all patients with anaemia.

Some studies have shown a relationship between the age of diagnosis with CD and developing a co-existent IMC. Some researchers have found that gluten exposure is a significant predictor of co-existent autoimmunity in CD patients, whereas others have found that increased age at diagnosis is the significant predictor rather than gluten exposure (Ventura et al., 1999, Sategna Guidetti et al., 2001). In our patients diagnosed over 65 years we observe a borderline significant tendency to have a higher prevalence of co-existent IMCs compared to those diagnosed at younger age (Table 2.7) ( $p=0.055$ ). This tendency is also observed when the age of diagnosis is taken at over 55 years (supplementary Table 3, appendix VIII). Potentially people diagnosed at a later age are at increased risk of a delayed diagnosis and may therefore be exposed to gluten for a longer period of time.

### **2.5.6 Body mass index (BMI) changes in CD patients**

An interesting finding in our study is how, over time, more patients have presented at diagnosis with a normal, overweight or obese BMI in line with previous reports. In 2006, Dickey and Kearney reported a non-significant increase in the median BMI of CD patients at diagnosis over a period of 10 years, from 24.0 in the period from 1995-2000 to 24.9 from 2001-2005 (Dickey and Kearney, 2006). We report a significant increase in median BMI over the 55 years covered in this study from 21.6 before 1985 to 24.8 in the most recent period. A limitation to our study is the lack of an unaffected population control sample; thus, the observed changes in CD patients' BMI over time may be simply reflecting worldwide general population trends (Finucane et al., 2011).

To address this problem, CD patients' BMI at diagnosis in the present study was compared to that of the general population in the 1998, 2002 and 2008 Survey of Lifestyle Attitudes and Nutrition (SLAN) in Ireland (Figure 2.6). It can be observed that CD patients follow the same increasing trend in BMI over time, but are always lower than the general population in SLAN 1998 and 2002. Potential reasons for these differences in BMI between CD patients at diagnosis and the general population in these two periods, are lower intestinal absorption due to active coeliac disease, as well as lower food intake due to the adverse symptoms resulting after food ingestion such as bloating, abdominal pain and nausea (Kabbani et al., 2012). However, in the latter period, corresponding to SLAN 2008, CD patients and SLAN participants show a very similar BMI profile, with 49.8% of patients having a BMI at diagnosis over or equal to 25 kg/m<sup>2</sup> and 50% of SLAN participants being in the same category, and more obese present in the CD population compared to the general population (17.2% and 14.0% respectively) (Figure 2.6).

Therefore, it seems that CD patients have followed worldwide trends in BMI increase, reaching the same levels as the general population in the most recent years even when the condition is undiagnosed. This finding is striking due to the expected lower nutrient absorption, and adverse symptomatology that would lead to lower food intake at diagnosis, as previously mentioned. Nonetheless, a potential explanation for this is that the BMI in the SLAN surveys is obtained from self-reported data, and it is well known that self-reported BMI tends to be lower than measured ones (Visscher et al., 2006).

In fact, in SLAN 2007, measured BMIs were collected for a subcohort of the participants, which showed that a large percentage of overweight and obese participants according to measured BMI had been classified as normal weight and overweight respectively using self-reported data. When we compared the measured

BMIs from this SLAN 2007 sub-cohort to the BMIs at diagnosis for patients diagnosed  $\geq 2007$  (supplementary Figure 4, appendix IX), it can be observed that the percentage of SLAN participants over 25 kg/m<sup>2</sup> is 64% in this cohort as opposed to 49.8% in CD patients. Thus we can conclude that, while CD patients' BMI at diagnosis has been continuously increasing over time following global patterns, it has remained lower than the general population in undiagnosed CD.

A recent large epidemiological study found the risk of developing DH increased by 14% per BMI unit in obese women, however the risk of CD decreased 7% per BMI unit, and increased in underweight women (Harpsoe et al., 2014). We also observe a higher prevalence of DH in patients presenting at diagnosis with a high BMI (Table 2.6). Zingone *et al.* (2009) showed a similar association in untreated CD patients. Obesity has been linked to a number of skin disorders which are thought to have their origin in altered epidermal barrier function in obese patients (Brown et al., 2004, Yosipovitch et al., 2007). Interestingly, we have demonstrated that although CD patients' BMI at diagnosis has been increasing during the 55 year period covered by this study, there has been a decrease in co-existent DH which is in line with findings from Finnish and UK studies (Figure 2.9) (Salmi et al., 2011, West et al., 2014).

Thus, while CD prevalence and incidence is increasing, the rates for DH are decreasing. As discussed above, increased CD epidemiology is thought to be due to improved diagnostic techniques and awareness, and increasing incidence. However, the decrease in DH cannot be attributed to lower detection rates because the same diagnostic techniques have been used since the middle of the 1970s, and the condition has visible skin symptoms at presentation (Collin et al., 2007, Salmi et al., 2011, West et al., 2014). Thus, there appears to be real reduction in the rates of DH which may be related to quicker detection of untreated CD, since this seems to be a pre-requisite for the development of DH (Aine et al., 1992, Salmi et al., 2011, West et al., 2014, Collin et al., 2017). However, we again consider the possibility, as with ATD, that changing environmental factors are predisposing individuals to develop CD who have less exposure to DH genetic susceptibility factors than before, analogous to our scenario of a biological shift favouring milder, later onset disease.

Another interesting and established finding is that whereas CD has a female predominance, DH has traditionally had a male preponderance which ranged from 1.5:1 to 2:1 (Bolotin and Petronic-Rosic, 2011, Salmi et al., 2011, Reunala and Lokki, 1978). However, this situation also seems to have changed over time with studies reporting a male/female ratio of approximately 1 in recent years (Salmi et al., 2011, West et al., 2014). Although it does not reach statistical significance, we observe a

similar effect, with the male/female ratio dropping from 2.5  $\leq 1985$  to 1.0  $\geq 2011$  (supplementary Figure 5, appendix X).

This study shows that only a minority of CD patients (6.2%) present underweight at diagnosis with a large percentage (40.6%) presenting overweight or obese (Table 2.6). These percentages correlate well with similar studies in the literature, indicating that the traditional idea of CD patients presenting underweight at diagnosis is no longer true (West et al., 2004, Murray et al., 2004, Dickey and Kearney, 2006, Kabbani et al., 2012). Patients overweight or obese at diagnosis presented more often with a non-classical CD phenotype and had a lower percentage of Marsh 3c biopsies. On the other hand, their underweight and normal weight counterparts presented with a more evident malabsorption scenario, correlating with significantly more weight loss reported at diagnosis, as well as a tendency towards more steatorrhea. A tendency towards more diarrhoea at diagnosis in the underweight patients can also be observed.

Harpsoe's *et al.*'s (2014) finding that CD risk decreased by 7% per BMI unit may indicate underdiagnosis of the condition in overweight and obese patients due to lack of awareness or suspicion of CD in this group, and may explain our observed tendency of a higher age at diagnosis in overweight and obese patients (Table 2.6) (Harpsoe et al., 2014).

There was also a powerful correlation between BMI and bone disease, including osteoporosis and osteopenia. Despite similar ages at data collection and higher female percentage in each BMI category, only 24.8% of overweight or obese patients at diagnosis was found to have osteoporosis compared to 74.2% of underweight or normal weight patients (Table 2.6). There was also a significant difference in bone mineral density (BMD) at diagnosis of osteoporosis or osteopenia between patients classified per BMI category, with overweight or obese patients at CD diagnosis having higher values both at the lumbar spine and the neck of the femur, indicating a milder bone disease phenotype in these patients (Table 2.6). This effect had been reported before by Dickey and Kearney (2006), who found higher BMD in overweight or obese patients who had a DEXA scan at diagnosis with CD, compared to their underweight and normal weight counterparts. These authors suggest that BMI should be considered when selecting CD patients for a DEXA scan.

CD patients of all ages are predisposed to bone disease, with the prevalence of CD adults with osteoporosis or osteopenia ranging from 26% to 74% (Kemppainen et al., 1999, Vilppula et al., 2011, Larussa et al., 2012, Di Stefano et al., 2013, Singh and Garber, 2016). Although the mechanism underlying this predisposition to bone disease

is not completely clear, it is suspected to result from a combination of impaired calcium absorption due to bowel damage and thus stimulation of parathyroid hormone, together with chronically elevated cytokines and antibodies against the antiresorptive protein osteoprotegerin (Colston et al., 1994, Fornari et al., 1998, Pazianas et al., 2005, Tilg et al., 2008, Riches et al., 2009). The risk of fractures is also increased in CD patients, above all in those non-compliant with the GFD and in undiagnosed CD (Vasquez et al., 2000, Heikkila et al., 2015b, Heikkila et al., 2015a).

Initiation of the GFD corrects BMD in CD patients diagnosed during childhood reaching eventually the same level as healthy controls. However, this is not the case for CD patients adhering to the GFD but who were diagnosed in adulthood. For these patients, though the GFD results in a significant increase in BMD, it never reaches the same level as healthy controls, and they will require calcium and vitamin D supplementation, and in some cases antiresorption treatment with bisphosphonates (Larussa et al., 2012, Ludvigsson et al., 2014a). In the present study, 29 patients were assessed retrospectively for adherence to the GFD by an expert coeliac dietician. A larger percentage of patients with normal range BMI at diagnosis were classified as having excellent or good compliance compared to overweight and obese patients (88.9%, 75% and 33.3% respectively) ( $p=0.062$ ) (Figure 2.8). This is in line with similar findings by Kabbani *et al.*, suggesting potential behavioural and diet-related differences between the two groups of patients (Kabbani et al., 2012).

Therefore, patients overweight or obese at diagnosis would hypothetically be less diet aware than their underweight and normal weight counterparts. Although adherence has been found to depend on different factors, more research is needed on the relationship between BMI and compliance with the GFD (Hall et al., 2009). Moreover, as previously highlighted, there is a tendency to diagnose overweight and obese patients later in life. Taken together, this evidence would seem to suggest that, although initially presenting with higher BMD, it seems plausible that this cohort of patients is long term at higher risk of both bone disease and fractures due to both less adherence to the GFD and delayed diagnosis.

As probably may be expected, we found that the median BMI of our sample increased a year after diagnosis. Significant changes in BMI took place both for patients presenting underweight and normal weight at diagnosis (Table 2.11). After CD diagnosis weight gain is desirable in underweight patients, and we observe that 91.7% of these patients put on weight a year after diagnosis, with the large majority (83.3%) entering the normal weight category (Table 2.12). However, weight gain is not welcome in patients who are normal weight, overweight or obese at diagnosis. We observe that

all of these categories of patients increase their weight a year after diagnosis. 76.4% of the patients presenting with normal BMI increased their weight, with 16.5% moving into the  $\geq 25$  kg/m<sup>2</sup> category. 72.9% of overweight and obese patients at diagnosis increased weight, and the majority remained in the  $\geq 25$  kg/m<sup>2</sup> category (Table 2.12).

Our results are in line with those of Dickey and Kearney (2006) and Kabbani *et al.* (2012), however they differ from those by Cheng *et al.* (2010) who reported a beneficial effect of the GFD on BMI, with underweight patients gaining weight and overweight and obese patients losing weight (Figure 2.7). Possible explanations for this difference are that in the study by Cheng *et al.* only patients adherent to the GFD were included, moreover all patients were closely followed by a dietitian, and although there was no specific advice about weight loss, the issue was frequently addressed by the dietitian (Cheng *et al.*, 2010). The study by Dickey and Kearney also only includes GFD compliant patients, using serology to measure compliance and determining that seronegative patients are adherent to the diet. However, though they mention that the diet was prescribed under the supervision of a dietitian, they do not mention close follow-up of the patients during the two year period.

A limitation to the findings on BMI and weight changes presented in this study is the fact that we are reporting BMI and weight changes a year after diagnosis which is probably not long enough to adapt to the GFD. Also we only have data on adherence for a minority of the patients which provides very little information on differences in weight changes categorising the groups into good and poor adherence. Taking into account all the previous findings we can conclude that a large percentage of CD patients present normal weight, overweight or obese at diagnosis highlighting the well reported fact that CD should no longer be considered mainly a malabsorptive condition in which patients present underweight. Later age at diagnosis in overweight/obese patients could reflect lack of awareness of CD in this cohort. Our results indicate that dietary advice on weight maintenance and weight loss is a pivotal part on coeliac dietary education. CD patients who are overweight and obese at diagnosis may require more support adhering to the GFD.

### **2.5.7 Serum biomarkers at CD diagnosis**

Certain serum biomarkers including; serum haemoglobin, ferritin, folate and B12 were collected for a small subcohort of patients. Haematological abnormalities are common in untreated CD, thus IDA, folate and B12 deficiency are common at diagnosis in these patients. The anaemia in CD is normally due to malabsorption of iron and vitamins such as folate and B12, though as previously discussed the presence of anaemia of

chronic disease in certain patients highlights the role of inflammation in the development of IDA (Halfdanarson et al., 2007, Baydoun et al., 2012). IDA due to malabsorption presents with, among other parameters, low serum ferritin (Halfdanarson et al., 2007, Baydoun et al., 2012).

Therefore, lower serum ferritin would be expected in those who presented with more damage to the small bowel mucosa, however when we correlate serum ferritin with damage to the small bowel mucosa according to Marsh criteria, we do not observe a significant difference between groups, although levels are lower in those who presented with a Marsh biopsy  $\geq 3b$  (Table 2.4). The same happens when we correlate serum folate and B12 with mucosal damage, however serum folate and B12 lack sensitivity and specificity to detect a deficiency and other parameters such as red cell folate, serum methylmalonic acid and homocysteine would be needed (Table 2.4) (Snow, 1999, Baydoun et al., 2012). Moreover, the low number of patients for which these biomarkers were available would also have an effect in the non-significant findings.

The previous serum biomarkers were also correlated with the patients' BMI at diagnosis (Table 2.6). Once more no significant differences were observed between the two BMI groups, however it is interesting to observe higher ferritin and lower folate and B12 levels in people with a BMI  $\geq 25$  kg/m<sup>2</sup>. The higher ferritin levels correlate with the fact that patients overweight or obese at diagnosis present with less classical CD and less damage to the small bowel mucosa, thus having less malabsorption (Table 2.6). On the other hand, serum folate and B12 have been found to be low in overweight and obese subjects (Kimmons et al., 2006, Baltaci et al., 2013, Bird et al., 2015)

## 2.6 Conclusion

This study reports the clinical features of a large cohort of CD patients in Ireland both cross-sectionally and over a long time frame. Our results corroborate findings internationally, but also shed light on the changing characteristics of CD in a country where there is a paucity of data, but where high incidence has been reported. This information is needed to increase awareness of CD among physicians and the general population, which may facilitate early diagnosis, improved management and prevention of associated comorbidities.

The clinical profile of the disease has changed radically over the past 55 years. Among these changes we find; an increased median age at diagnosis in adulthood, higher prevalence of non-classical CD at diagnosis, equal prevalence of patients diagnosed in childhood (<18 years) and patients diagnosed over 65 years and a large percentage of patients presenting either overweight or obese at diagnosis. Also less associated ATD and DH are observed over time. It seems plausible that these changes are due to improved detection of non-classical disease, and/or potentially shifts in environmental exposures which are engendering changes in the clinical manifestation of CD.

The clinical presentation of overweight and obese CD patients at diagnosis is different to their underweight and normal weight counterparts, with patients presenting more with a non-classical phenotype associated to less severe biopsies at diagnosis. In overweight and obese patients extraintestinal manifestations also differ, having less osteoporosis, but more DH compared to the groups presenting with a BMI less than 25 kg/m<sup>2</sup>.

The group of patients diagnosed over 65 years also differ in their clinical presentation compared to their younger counterparts. The over 65 have a less varied clinical presentation and present significantly more with iron deficiency anaemia. This group also show a tendency to have more co-existent IMCs.

We have also highlighted how dietary advice on weight maintenance and weight loss is a pivotal part on coeliac dietary education. CD patients who are overweight and obese at diagnosis may require more support adhering to the GFD.



### **Chapter 3 Investigation of Coeliac Disease Co-existence With Other Immune Mediated Conditions and Pattern of Autoimmune Thyroid Disease Over Time**

### 3.1 Introduction

The co-existence of CD with other immune-mediated conditions (IMC) has been well reported in the literature (Fasano, 2006, Neuhausen et al., 2008, Lundin and Wijmenga, 2015). It is recommended that patients with type 1 diabetes mellitus (T1DM) and autoimmune thyroid disease (ATD) are screened for CD (Bakker et al., 2016, Fallahi et al., 2016, Roy et al., 2016). Sharing of HLA haplotypes with several other immune-mediated conditions, such as type 1 diabetes mellitus (T1DM), autoimmune thyroid disease (ATD) and systemic lupus erythematosus (SLE), may in part explain the predisposition of CD patients to develop these functionally related conditions. Moreover, genome wide association studies (GWAS) have shown that many of the non-HLA CD associated risk variants are also shared with other IMCs. Non-HLA loci associated with CD have been shown to be also shared variously with rheumatoid arthritis (RA), inflammatory bowel disease (IBD), T1DM, ankylosing spondylitis (AS) and multiple sclerosis (MS) (Gutierrez-Achury et al., 2011).

Therefore, the co-occurrence of CD with other immune-mediated conditions (IMCs) may be caused by the sharing of genetic susceptibility factors; however, it has also been suggested that active CD may potentiate the development of these other IMCs (Fasano, 2006). This hypothesis is of interest clinically, as it is based on the proposition that gluten induced autoantibodies help promote the development of co-existent IMCs, and thus eliminating gluten would stop the development of the comorbidity (Ventura et al., 1999, Ventura et al., 2000, Toscano et al., 2000). Indeed, studies in animals and humans have linked gluten exposure to a higher risk of developing T1DM (Serena et al., 2015).

Similarly, in the case of ATD, anti-tissue transglutaminase antibodies (tTGAs) have been suggested as a potential mechanism triggering ATD since they have been found to bind thyroid follicles in active CD before introduction of a gluten free diet (GFD) (Naiyer et al., 2008, Duntas, 2009). However, other studies have not found an association between gluten exposure or adherence to the GFD and the posterior development of ATD in CD patients (Meloni et al., 2009, Metso et al., 2012). They have, however, found a higher prevalence of

ATD in CD patients compared to age matched controls, and reported a gender bias with more women being affected, suggesting a role of sex hormones predisposing to ATD (Chiovato et al., 1993, Meloni et al., 2009). In this study, we have found ATD to be

the most common co-existent IMC with CD (Chapter 2, Table 2.2). We have also reported a decreased prevalence of ATD over a 55 year period in Ireland (Chapter 2, Table 2.8), paired with a change in CD clinical presentation at diagnosis in adulthood. We also observed a female bias, with a female/male ratio of 2.2:1 (Chapter 2, Table 2.1).

Studies analysing the link between gluten exposure and the development of co-existent IMCs in CD patients have yielded inconsistent results. Some suggest that reduced gluten exposure can prevent the development of co-existent IMCs (Ventura et al., 1999, Ventura et al., 2000, Cosnes et al., 2008), while other studies do not show this protective effect (Sategna Guidetti et al., 2001, Viljamaa et al., 2005, Ouaka-Kchaou et al., 2008, Elli et al., 2012, Imperatore et al., 2016). Therefore, despite all the above studies reporting an increased risk of co-existent IMCs, whether the GFD can prevent or delay the development of these associated IMCs is still controversial.

Several studies have analysed the risk of developing different co-existent IMCs individually in CD patients compared to controls. These studies have showed a significantly increased prevalence of conditions such as T1DM, SLE, ATD and psoriasis in these patients (Hakanen et al., 2001, Ludvigsson et al., 2006, Ludvigsson et al., 2011, Ludvigsson et al., 2012). However, few studies have calculated this risk using prevalence ratios (PRs), derived by comparing the prevalence of IMCs in the CD population vs. the general population (Yang et al., 2005, Neuhausen et al., 2008). Studies like those of Yang *et al.* and Neuhausen *et al.* allow to explore how large is the risk of CD patients to develop these conditions. Moreover, studying the timing of presentation of these co-existent IMCs in relation to CD diagnosis is important due to the considerable on-going debate as to whether a GFD protects against the development of IMCs (Cosnes et al., 2008, Ventura et al., 1999, Viljamaa et al., 2005, Sategna Guidetti et al., 2001).

In addition, not only are CD patients at increased risk of developing IMCs, their relatives also have a higher predisposition to do so (Neuhausen et al., 2008, Emilsson et al., 2015). However, estimates of CD prevalence among first degree relatives (FDRs) have been inconsistent between studies, potentially due to the use of different diagnostic methods (serology versus biopsy testing) and also different geographical locations. A systematic review by Dube *et al.* reported that the pooled CD prevalence in FDRs was 16% (Dube et al., 2005). Two studies have reported that distant relatives (DRs; defined in section 3.2.2 below) have an IMC prevalence of 2.6%-5.5% (Korponay-Szabo et al., 1998, Fasano et al., 2003). Shared genetics and environmental factors, as well as more active case finding, account for this high

prevalence of CD among relatives of cases (Fasano et al., 2003, Neuhausen et al., 2008, Emilsson et al., 2015). Moreover, CD patients' relatives also have a higher predisposition to develop other IMCs (Neuhausen et al., 2008, Emilsson et al., 2015). Given these observations, it is of interest to further explore the prevalence of IMCs in patients and their relatives (Petaros et al., 2002, Cataldo and Marino, 2003, Neuhausen et al., 2008, Emilsson et al., 2015).

Therefore the aim of this chapter is to investigate the prevalence and timing of presentation of co-existent IMCs, as well as their burden, using prevalence ratios (PRs) in a CD patient population. We report a descriptive analysis of the amount of CD patients having a relative affected with CD or other IMC, as well as the type of patient affected. We also calculate the burden of IMCs in the parents of CD patients. Moreover, in the previous chapter we have reported a reduction in ATD over time among CD patients diagnosed in adulthood, concomitant with an increased age at CD diagnosis. Thus we wished to investigate this in greater detail, specifically the timing of ATD presentation over time, and the role of the time period of CD diagnosis, and age at diagnosis with CD, in predicting ATD. We will also discuss, for each of the co-existent IMCs, shared genetics and potential environmental factors with CD.

## 3.2 Aims

- Investigation of the prevalence and timing of presentation of co-existent IMCs in CD patients.
- Calculation of the burden of IMCs in CD patients using prevalence ratios (PRs).
- Descriptive analysis of CD relatives affected by IMCs.
- Calculation of IMC burden in the parents of CD patients using PRs.
- Investigation of the role of time period of CD diagnosis and age at diagnosis with CD in predicting ATD.
- Discussion of shared genetics and potential environmental factors between CD and other IMCs

## **3.3 Methods**

### **3.3.1 Study design and study population**

A retrospective analysis of medical records was performed using a CD patient cohort (n=749) attending five secondary referral centres in Ireland and diagnosed from 1960-2015. Study design is as described in Chapter 2, section 2.2.1.

### **3.3.2 Data collection and definition of study variables**

The following information was collected for CD patients; age, gender, age at diagnosis with CD, year of diagnosis with CD, presence of co-existent IMCs, age at diagnosis with co-existent IMCs, family history of CD and family history of other IMCs. If family history of CD and/or other IMCs was found then the number of family members affected and the type of relative affected was also collected either from the medical records or using a structured questionnaire. Relatives were divided into first degree relatives (FDRs) and distant relatives (DRs). FDRs are parents, siblings and offspring, and DRs are cousins, aunts and uncles, nieces and nephews, grandparents and grandchildren.

There were 86 patients that were found to have both a family history of CD and also a family history of other IMC, 25 of these patients reported the same category of relative (e.g. brother) both in CD history and IMC history. As the identity of the relative could not be established (i.e. whether it was the same brother with CD and an IMC or two different brothers one with the IMC and the other with CD) these 25 individuals were excluded in the analysis related to FDRs and DRs and IMC burden. This was done to avoid the possible confounding effects of including potential CD patients in these analysis where we are looking to establish solely the burden of IMCs in patients relatives rather than those suffering from CD (Neuhausen et al., 2008).

The presence of the following CD associated IMCs was determined in the study cohort and their relatives; T1DM, ATD, RA, IBD, psoriasis, and SLE. A category 'other' was created to include any IMCs that had not been included initially. The presence of co-existent IMCs could be corroborated in the medical records in the majority of CD patients' cases. The presence of co-existent ATD could be corroborated in the medical records in the majority of cases either by the presence of thyroid peroxidase antibodies at diagnosis or by specific ATD medication taken by the patient. In a minority of cases, certain data was accepted as self-reported as it had not been systematically collected in the medical records, and also due to its potential to have changed since first

reported in the records, such as family history of CD and family history of other IMCs. The age of diagnosis with co-existent IMC was also self-reported in a minority of cases where diagnosis had not been made in the referral centre.

### **3.3.3 Statistical analysis**

Two groups were analysed for the prevalence and burden of IMCs; the cohort of 749 CD patients and their parents' cohort. The latter was done because the information on how many relatives existed per patient was not available in the medical records, thus we were only sure of the sample of parents, (two per patient:  $n=1498$ ). To calculate the burden of associated IMCs, prevalence ratios (PRs) were calculated as the ratio of the number of observed cases compared to the number of expected cases using prevalence data from the literature for the different IMCs (Table 3.1). 95% confidence intervals (CIs) for the PRs were also calculated (Neuhausen et al., 2008). The criteria to choose the studies reporting prevalence data in the general population was to give priority to any existing meta-analysis, or in their absence, try to find epidemiological studies carried out in the Irish population or a similar cohort of patients (white ethnicity and northern European).

A categorical variable was created for the year of diagnosis as follows; 1)  $\leq 1985$ , 2) 1986-1995, 3) 1996-2005 4) 2006-2010 and 5)  $\geq 2011$  as described in Chapter 2, section 2.2.4.1. In order to control for patients' age at data collection when analysing changes of ATD over time, analysis included at first, all patients diagnosed with CD in adulthood ( $>18$  years of age) ( $n=652$ , 442 females; median age 57 years, range 18-91), and following, subgroup analysis was carried out with patients  $>55$  years at data collection ( $n=354$ , 226 females; median age 67 years, range 56-91) to control for age as a confounder and allow for the development of co-existent ATD (Reid and Wheeler, 2005, Garber et al., 2012).

Multivariate binary logistic regression was carried out in the CD patients' cohort diagnosed in adulthood using time period at CD diagnosis, age at CD diagnosis, patients' age at data collection, gender and having a family history of a second

IMC as independent variables. The dependent binary variable in the regression analysis was having co-existent ATD.

Statistical analysis were carried out using SPSS version 24.0 (SPSS, Inc., Chicago, IL). Discrete variables are presented as percentages and continuous variables as medians and ranges. The relationship between discrete variables was analysed using the  $\chi^2$  test. Differences between groups were assessed using the Mann-Whitney U and

Kruskall-Wallis H non-parametric tests. The Jonckheere-Terpstra test was used to carry out the trend analyses. A threshold p-value of  $<0.05$  was considered significant in all analyses.



## 3.2 Results

### 3.2.1 General population prevalence studies of immune mediated conditions (IMCs) and prevalence of IMCs among CD patients and their relatives.

Table 3.1 Disease prevalence studies for comparison.

| Study                             | Date | Type of Study                      | Age      | Population | Prevalence |
|-----------------------------------|------|------------------------------------|----------|------------|------------|
| <b>ATD</b>                        |      |                                    |          |            |            |
| Garmendia Madariaga <i>et al.</i> | 2014 | meta-analysis                      | All ages | Europe     | 3.82%      |
| <b>Hypothyroidism</b>             |      |                                    |          |            |            |
| Garmendia Madariaga <i>et al.</i> | 2014 | meta-analysis                      | All ages | Europe     | 3.05%      |
| <b>Hyperthyroidism</b>            |      |                                    |          |            |            |
| Garmendia Madariaga <i>et al.</i> | 2014 | meta-analysis                      | All ages | Europe     | 0.75%      |
| <b>T1DM</b>                       |      |                                    |          |            |            |
| Maahs <i>et al.</i>               | 2010 | epidemiological                    | ≤ 18     | US         | 0.30%      |
| Forouhi <i>et al.</i>             | 2006 | epidemiological (population-based) | All ages | UK         | 0.34%      |
| <b>Psoriasis</b>                  |      |                                    |          |            |            |
| Gelfand <i>et al.</i>             | 2005 | epidemiological                    | All ages | UK         | 1.50%      |
| Springate <i>et al.</i>           | 2016 | epidemiological                    | All ages | UK         | 2.80%      |
| <b>IBD</b>                        |      |                                    |          |            |            |
| Burisch <i>et al.</i>             | 2013 | literature review                  | All ages | Europe     | 0.30%      |
| Molodecky <i>et al.</i>           | 2012 | literature review                  | All ages | Europe     | 0.87%      |
| <b>Chrons Disease</b>             |      |                                    |          |            |            |
| Loftus <i>et al.</i>              | 2007 | epidemiological (population-based) | All ages | US         | 0.17%      |
| Lucendo <i>et al.</i>             | 2014 | epidemiological                    | adults   | Spain      | 0.14%      |
| <b>Ulcerative colitis</b>         |      |                                    |          |            |            |
| Loftus <i>et al.</i>              | 2007 | population-based                   | All ages | US         | 0.21%      |
| Rubin <i>et al.</i>               | 2000 | epidemiological                    | All ages | UK         | 0.24%      |

**Table 3.1 (cont.)**

|                               |      |                                    |                            |             |            |  |
|-------------------------------|------|------------------------------------|----------------------------|-------------|------------|--|
| <b>RA</b>                     |      |                                    |                            |             |            |  |
| Symmons <i>et al.</i>         | 2002 | epidemiological                    | ≥ 16                       | Europe      | 0.89%      |  |
| Power <i>et al.</i>           | 1999 | epidemiological (population-based) | All ages                   | Ireland     | 0.50%      |  |
| <b>SLE</b>                    |      |                                    |                            |             |            |  |
| Gourley <i>et al.</i>         | 1997 | epidemiological                    | All ages                   | Ireland     | 0.25%      |  |
| Rees <i>et al.</i>            | 2016 | epidemiological                    | All ages                   | UK          | 0.13%      |  |
| <b>Alopecia Areata</b>        |      |                                    |                            |             |            |  |
| Villasante-Fricke & Miteva    | 2015 | meta-analysis                      | All ages                   | Worldwide   | 2.0%       |  |
| <b>Sarcoidosis</b>            |      |                                    |                            |             |            |  |
| Arkema <i>et al.</i>          | 2016 | epidemiological                    | adults                     | Sweden      | 0.15-0.21% |  |
| Barbaros <i>et al.</i>        | 2012 | epidemiological                    | All ages                   | US          | 0.16-0.33% |  |
| <b>MS</b>                     |      |                                    |                            |             |            |  |
| Brola <i>et al.</i>           | 2016 | epidemiological                    | All ages                   | Poland      | 0.11%      |  |
| Simonsen <i>et al.</i>        | 2016 | epidemiological                    | All ages                   | Norway      | 0.21%      |  |
| Albor <i>et al.</i>           | 2016 | epidemiological                    | All ages (white ethnicity) | East London | 0.18%      |  |
| Dilokthornsakul <i>et al.</i> | 2016 | epidemiological                    | All ages                   | US          | 0.15%      |  |

Presented in Table 3.1 are the observed prevalences of ATD, T1DM, psoriasis, IBD, RA, SLE, alopecia areata (AA), sarcoidosis, pernicious anaemia, IgA deficiency and MS in the general population as reported by different studies. These prevalences were compared later with the observed prevalences for the same conditions in our cohort of CD patients, as explained below.

**Table 3.2 Hypothyroidism, Type 1 diabetes mellitus (T1DM) and rheumatoid arthritis (RA) prevalence studies for comparison used by Neuhasen *et al.***

| <b>Study</b>                              | <b>Date</b> | <b>Type of Study</b> | <b>Age</b> | <b>Population</b>  | <b>Prevalence</b>                       |
|-------------------------------------------|-------------|----------------------|------------|--------------------|-----------------------------------------|
| <b>Hypothyroidism</b>                     |             |                      |            |                    |                                         |
| National Health and Nutrition Examination | 1988-1994   | Survey               | ≥12        | US                 | 4.60%                                   |
| Canaris <i>et al.</i>                     | 1995        | Health fair          | >18        | US                 | 9.50%                                   |
| Bjoro <i>et al.</i>                       | 1995-1997   | Questionnaire        | >19        | Norway             | Females 4.8%<br>Males 0.9%              |
| <b>T1DM</b>                               |             |                      |            |                    |                                         |
| Jacobson <i>et al.</i>                    | 1965-1996   | Epidemiological      | <20        | US, Canada, Europe | 0.20%                                   |
| Forouhi <i>et al.</i>                     | 2001        | Population-based     | All        | UK                 | 0.30%                                   |
| <b>RA</b>                                 |             |                      |            |                    |                                         |
| Linoss <i>et al.</i>                      | 1950-1974   | Record review        | >15        | US                 | Females 1.3%<br>Males 0.6%<br>Both 1.2% |
| Jacobson <i>et al.</i>                    | 1965-1996   | Literature review    | >16        | US, Canada, Europe | 0.86%                                   |

Prevalence ratios for hypothyroidism, T1DM and RA for a cohort of CD patients and their relatives were calculated using the above prevalences obtained by different studies in the general population by Neuhausen *et al.*, (2008). Due to the similarities between our and Neuhausen's studies, we have used the same data to calculate PRs in our CD patient population, and to allow comparison of our results with those obtained by Neuhausen *et al.* (see Table 3.6).

**Table 3.3 CD patients' IMCs prevalence and prevalence of CD and IMCs among relatives categories.**

| Characteristic                                                          | % (n)                   |
|-------------------------------------------------------------------------|-------------------------|
| <b>Ratio Females/Males CD patients</b> n 749                            | 68.4 (237)/31.6 (512)   |
| <b>CD patients with co-existent IMC</b> n 749                           |                         |
| Yes                                                                     | 31.2 (234)              |
| No                                                                      | 68.8 (515)              |
| <b>Ratio Females/Males CD patients with co-existent IMC</b> , n 234     | 78.5 (183)/21.9 (51)    |
| <b>Family history of CD</b> , n 749                                     |                         |
| Yes                                                                     | 36.4 (273)              |
| No                                                                      | 63.6 (476)              |
| <b>Relative category affected by CD per patient<sup>§</sup></b> , n 273 |                         |
| First degree                                                            | 56.4 (154)              |
| Distant                                                                 | 20.1 (55)               |
| Both                                                                    | 21.2 (58)               |
| Not Available                                                           | 2.2 (6)                 |
| <b>Total Relatives affected with CD</b> , n 458                         |                         |
| First Degree                                                            | 59.4 (272)              |
| Distant                                                                 | 40.6 (186)              |
| <b>First Degree Relative affected with CD</b> , n 272                   |                         |
| Mother                                                                  | 9.9 (27)                |
| Father                                                                  | 7.0 (19)                |
| Sister                                                                  | 44.0 (120)              |
| Brother                                                                 | 18.0 (49)               |
| Daughter                                                                | 14.0 (38)               |
| Son                                                                     | 7.0 (19)                |
| <b>Distant relative affected with CD</b> , n 186                        |                         |
| Cousin                                                                  | 36.6 (68)               |
| Aunt                                                                    | 28.5 (53)               |
| Uncle                                                                   | 8.1 (15)                |
| Niece                                                                   | 12.4 (23)               |
| Nephew                                                                  | 9.1 (17)                |
| Grandmother                                                             | 1.6 (3)                 |
| Grandfather                                                             | 1.1 (2)                 |
| Granddaughter                                                           | 1.6 (3)                 |
| Grandson                                                                | 1.1 (2)                 |
| <b>Ratio Females/Males relatives CD<sup>£</sup></b> , n 337             | 63.5 (214) / 36.5 (123) |
| <b>Family history of other IMC</b> , n 749                              |                         |
| Yes                                                                     | 26.8 (201)              |
| No                                                                      | 73.2 (548)              |
| <b>Relative category affected by other IMC per patient</b> , n 193      |                         |
| First degree                                                            | 74.6 (150)              |
| Distant                                                                 | 10.4 (21)               |
| Both                                                                    | 12.4 (25)               |
| Not Available                                                           | 2.5 (5)                 |
| <b>Total Relatives affected with co-existent IMC</b> , n 283            |                         |
| First degree                                                            | 76.0 (215)              |
| Distant                                                                 | 24.0 (68)               |

**Table 3.3 (cont.)****First Degree Relative affected with IMC, n 215**

|          |           |
|----------|-----------|
| Mother   | 30.7 (66) |
| Father   | 13.5 (29) |
| Sister   | 25.1 (54) |
| Brother  | 18.6 (40) |
| Daughter | 8.4 (18)  |
| Son      | 3.7 (8)   |

**Distant Relative affected with IMC, n 68**

|               |           |
|---------------|-----------|
| Cousin        | 19.1 (13) |
| Aunt          | 25.0 (17) |
| Uncle         | 10.3 (7)  |
| Niece         | 7.35 (5)  |
| Nephew        | 7.35 (5)  |
| Grandmother   | 8.8 (6)   |
| Grandfather   | 10.3 (7)  |
| Granddaughter | 4.4 (3)   |
| Grandson      | 2.9 (2)   |
| Unknown       | 4.4 (3)   |

**Ratio Females/Males relatives with IMC<sup>£</sup>, n 270**

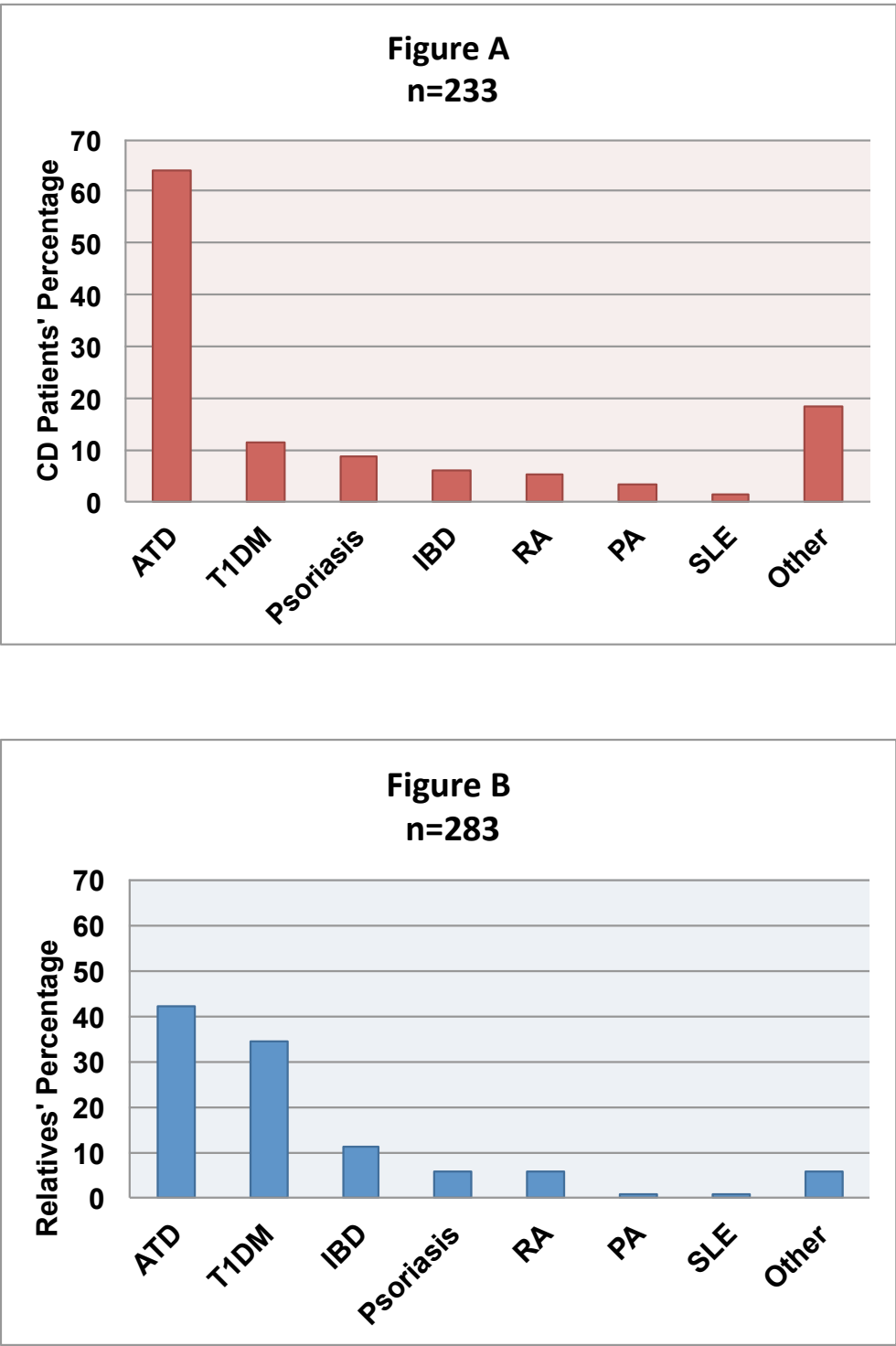
62.6 (169) / 36.3 (98)

§ First degrees relatives are parents, siblings and offsprings

£ Cousins not included as gender was not specified by patients

A total of 749 CD patients were included in the analysis. A higher prevalence of females was found among CD patients (female/male ratio = 2.2/1). Of the 749 CD patients, 31.1% (n=233) had a co-existent IMC. Among the 233 CD patients with co-existent IMCs, a higher prevalence of females was found (female/male = 3.6/1). A total of 458 relatives of the 749 CD patients were found to have CD, and 59.4% of these were FDRs. Higher female prevalence was also found among relatives with CD (female/male = 1.7:1). A total of 283 CD relatives of the 749 CD patients were found to have IMCs, 76% of these were FDRs. Higher female prevalence was also found among relatives with IMCs (female/male= 1.7:1).

Figure 3.1 Prevalence and type of co-existent IMCs in CD patients (Figure A) and their relatives (Figure B).



ATD and T1DM were the most prevalent IMCs within the cohort of CD patients with a co-existent IMC and also the cohort of relatives found to have an IMC (75.5% and 76.0% combined frequencies for the two conditions and cohorts respectively). Other co-existent IMCs reported in either patients or their relatives are listed in Table 3.4.

**Table 3.4 Prevalence and type of other co-existent IMCs in CD patients and their relatives.**

| Characteristic                                         |                                  | % (n)      |
|--------------------------------------------------------|----------------------------------|------------|
| <b>Coeliac Patients</b>                                |                                  |            |
| Other co-existent IMCs in CD patients, n 43            | Alopecia areata                  | 23.25 (10) |
|                                                        | Sarcoidosis                      | 16.3 (7)   |
|                                                        | Immunoglobulin A deficiency      | 9.3 (4)    |
|                                                        | Multiple sclerosis               | 7.0 (3)    |
|                                                        | Scleroderma                      | 7.0 (3)    |
|                                                        | Lichen Planus                    | 7.0 (3)    |
|                                                        | Autoimmune hepatitis             | 7.0 (3)    |
|                                                        | Lupus                            | 4.65 (2)   |
|                                                        | Autoimmune adrenalitis           | 4.65 (2)   |
|                                                        | Primary biliary cholangitis      | 4.65 (2)   |
|                                                        | Immune-mediated thrombocytopenia | 2.3 (1)    |
|                                                        | Myasthenia Gravis                | 2.3 (1)    |
|                                                        | Vitiligo                         | 2.3 (1)    |
|                                                        | Autoimmune neutropenia           | 2.3 (1)    |
| Other co-existent IMCs in CD patients' relatives, n 13 | Multiple sclerosis               | 38.5 (5)   |
|                                                        | Lupus                            | 15.4 (2)   |
|                                                        | Sarcoidosis                      | 7.7 (1)    |
|                                                        | Immunoglobulin A deficiency      | 7.7 (1)    |
|                                                        | Autoimmune hepatitis             | 7.7 (1)    |
|                                                        | Autoimmune adrenalitis           | 7.7 (1)    |
|                                                        | Primary sclerosing cholangitis   | 7.7 (1)    |
|                                                        | Primary biliary cirrhosis        | 7.7 (1)    |

### **3.2.2. Immune mediated conditions (IMCs) prevalence ratios in CD patients and their parents.**

In order to investigate whether the frequency of IMCs is greater among our CD patients compared to the general population, we examined our sample in reference to the previous general population studies listed in Table 3.1. Table 3.5 shows the number of cases and frequency of each autoimmune comorbidity observed in our CD cohort, as well as the female:male ratio, median age at diagnosis with co-existent IMC, and the number of cases expected based on the prevalence reported by the studies in Table 3.1.



**Table 3.5 Prevalence ratios of co-existent IMCs in CD cases using prevalence studies in Table 3.1.**

| <b>Disease</b>                  | <b>No.<br/>Observed (%)</b> | <b>Female/Male<br/>ratio</b> | <b>Median age<br/>of diagnosis<br/>(years)</b> | <b>No.<br/>Expected (n)</b> | <b>Standardized ratio<br/>(95% CI)</b> | <b>Presentation</b> | <b>n</b> |
|---------------------------------|-----------------------------|------------------------------|------------------------------------------------|-----------------------------|----------------------------------------|---------------------|----------|
| <b>Thyroid Disorders, n 749</b> | 149 (19.9)                  | 6.8:1                        | 47 (15-79)                                     |                             |                                        | Prior CD            | 50       |
|                                 |                             |                              |                                                | 28.6                        | 5.2 (4.5-6.0)                          | Same time CD        | 9        |
|                                 |                             |                              |                                                | 37.8                        | 3.9 (3.4-4.5)                          | After CD            | 66       |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 24       |
| <b>Hypothyroidism</b>           | 110 (14.7)                  | 5.9:1                        | 48 (15-79)                                     |                             |                                        | Prior CD            | 31       |
|                                 |                             |                              |                                                | 22.8                        | 4.8 (4.0-5.7)                          | Same time CD        | 7        |
|                                 |                             |                              |                                                |                             |                                        | After CD            | 53       |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 19       |
| <b>Hyperthyroidism</b>          | 37 (4.9)                    | 17.5:1                       | 42 (15-70)                                     |                             |                                        | Prior CD            | 17       |
|                                 |                             |                              |                                                | 5.6                         | 6.6 (4.5-8.7)                          | Same time CD        | 2        |
|                                 |                             |                              |                                                |                             |                                        | After CD            | 13       |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 5        |
| <b>Type 1 Diabetes, n 749</b>   | 27 (3.6)                    | 0.8:1                        | 21 (1-61)                                      |                             |                                        | Prior CD            | 19       |
|                                 |                             |                              |                                                | 2.2                         | 12.3 (7.7-16.8)                        | Same time CD        | 2        |
|                                 |                             |                              |                                                | 2.5                         | 10.8 (6.8-14.8)                        | After CD            | 2        |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 4        |
| <b>Psoriasis, n 749</b>         | 20 (2.7)                    | 1.8:1                        | 38.5 (7-65)                                    |                             |                                        | Prior CD            | 6        |
|                                 |                             |                              |                                                | 10.7                        | 1.7 (1.06-2.7)                         | Same time CD        | 1        |
|                                 |                             |                              |                                                | 21.0                        | 0.95(0.54-1.4)                         | After CD            | 7        |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 6        |
| <b>IBD, n 749</b>               | 14 (1.9)                    | 2.5:1                        | 42 (21-65)                                     |                             |                                        | Prior CD            | 4        |
|                                 |                             |                              |                                                | 2.2                         | 6.4 (3.1-9.5)                          | Same time CD        | 3        |
|                                 |                             |                              |                                                | 6.2                         | 2.3 (1.1-3.4)                          | After CD            | 6        |
|                                 |                             |                              |                                                |                             |                                        | Not known           | 1        |

**Table 3.5 (cont.)**

|                           |          |     |              |          |                             |              |   |
|---------------------------|----------|-----|--------------|----------|-----------------------------|--------------|---|
| <b>Crohns disease</b>     | 6 (0.8)  | 6:0 | 39.5 (21-65) |          |                             | Prior CD     | 1 |
| <b>Study 1</b>            |          |     |              | 1.3      | 4.6 (0.9-8.3)               | Same time CD | 2 |
| <b>Study 2</b>            |          |     |              | 1.05     | 5.7 (1.14-10.2)             | After CD     | 3 |
|                           |          |     |              |          |                             | Not known    | 0 |
| <b>Ulcerative colitis</b> | 8 (1.1)  | 1:1 | 42 (25-60)   |          |                             | Prior CD     | 3 |
| <b>Study 1</b>            |          |     |              | 1.6      | 5 (1.5-8.5)                 | Same time CD | 1 |
| <b>Study 2</b>            |          |     |              | 1.8      | 4.4 (1.4-7.5)               | After CD     | 3 |
|                           |          |     |              |          |                             | Not known    | 1 |
| <b>RA, n 749</b>          | 12 (1.6) | 3:1 | 57.5 (51-71) |          |                             | Prior CD     | 3 |
| <b>Study 1</b>            |          |     |              | 6.6      | 1.8 (0.8-2.8)               | Same time CD | 0 |
| <b>Study 2</b>            |          |     |              | 3.7      | 3.2 (1.41-5.06)             | After CD     | 5 |
|                           |          |     |              |          |                             | Not known    | 3 |
| <b>SLE, n 749</b>         | 3 (0.4)  | 2:1 | ND           |          |                             |              |   |
| <b>Study 1</b>            |          |     |              | 1.9      | 1.5(0.19-3.35)              | ND           |   |
| <b>Study 2</b>            |          |     |              | 1.0      | 3(0.37-6.4)                 |              |   |
| <b>AA, n 749</b>          | 10 (1.3) | 9:1 | ND           |          |                             |              |   |
| <b>Study 1</b>            |          |     |              | 15.0     | 0.7(0.25-1.1)               | ND           |   |
| <b>Sarcoidosis</b>        | 7 (0.9)  | 7:0 | ND           |          |                             |              |   |
| <b>Study 1</b>            |          |     |              | 1.12/1.5 | 6.25(1.6-10.9)-4.6(1.2-8.1) | Prior CD     | 2 |
| <b>Study 2</b>            |          |     |              | 1.2/2.5  | 5.8(1.5-10.2)-2.8(0.7-4.9)  | Same time CD | 2 |
|                           |          |     |              |          |                             | After CD     | 1 |
|                           |          |     |              |          |                             | Not known    | 2 |
| <b>MS</b>                 | 3 (0.4)  | 2:1 | ND           |          |                             |              |   |
| <b>Study 1</b>            |          |     |              | 0.8      | 3.7 (0.5-8.0)               | ND           |   |
| <b>Study 2</b>            |          |     |              | 1.6      | 1.9 (0.2-4)                 |              |   |
| <b>Study 3</b>            |          |     |              | 1.3      | 2.3 (0.3-4.9)               |              |   |
| <b>Study 4</b>            |          |     |              | 1.1      | 2.7 (0.34-5.8)              |              |   |

We derived prevalence ratios (PRs) based on prevalence estimates from the two different studies considered for each of ATD, T1DM and IBD (Table 3.1). Thus we observe a significantly increased prevalence of ATD with 149 cases compared to 28.6/37.8 expected, corresponding to prevalence ratios (PR) of 5.2 (CI 4.5-6.0) and 3.9 (CI 3.4-4.5).

For T1DM, we observed 27 cases compared to 2.2/2.5 expected, with PR of 12.3 (CI 7.7-16.8) and 10.8 (CI 6.8-14.8). For IBD, 16 cases were observed compared to 2.2/6.2 expected giving PR 7.3 (CI 3.7-10.8) and 2.6 (CI 1.3-3.8).

ATD is diagnosed equally frequently prior to, or at the same time as CD, compared to after CD diagnosis. T1DM, on the contrary, is diagnosed mainly before CD diagnosis. Females are more likely to be affected by all conditions except T1DM and ulcerative colitis.

**Table 3.6 Prevalence ratios and presentation of co-existent T1DM, Hypothyroidism and RA in CD cases.**

Table 3.6 represents PRs for T1DM, hypothyroidism and RA in our CD sample using prevalence studies in Table 3.2, and comparison to PRs and 95% confidence intervals in the study by Neuhausen *et al.*

| Disease present study (n=749) | No. Observed present study | No. Expected present study | Prevalence Ratio (95% CI) | Disease Neuhausen <i>et al.</i> | No. Observed Neuhausen <i>et al.</i> (n) | No. Expected Neuhausen <i>et al.</i> (n) | Prevalence Ratio Neuhausen <i>et al.</i> 9 (95% CI) |
|-------------------------------|----------------------------|----------------------------|---------------------------|---------------------------------|------------------------------------------|------------------------------------------|-----------------------------------------------------|
| <b>Hypothyroidism</b>         |                            |                            |                           | <b>Hypothyroidism</b>           |                                          |                                          |                                                     |
| Study 1                       | 110                        | 34.45                      | 3.2 (2.6-3.75)            | Study 1, n=381                  | 40.0                                     | 17.5                                     | 2.3 (1.6-3.0)                                       |
| Study 2                       | 110                        | 71.15                      | 1.55 (1.3-1.8)            | Study 2, n=343                  | 39.0                                     | 32.6                                     | 1.2 (0.8-1.6)                                       |
| Study 3                       |                            |                            |                           | Study 3                         |                                          |                                          |                                                     |
| Females, n=512                | 94                         | 25                         | 3.8 (3.1-4.6)             | Females, n=191                  | 27                                       | 9.2                                      | 2.9 (1.9-4.2)                                       |
| Males, n=237                  | 16                         | 2.1                        | 7.6 (4.5-11.3)            | Males, n=143                    | 11                                       | 1.3                                      | 8.5 (4.1-14.1)                                      |
| <b>Type 1 Diabetes</b>        |                            |                            |                           | <b>Type 1 Diabetes</b>          |                                          |                                          |                                                     |
| Study 1                       | 27                         | 1.5                        | 18 (11.5-24.5)            | Study 1, n=408                  | 12                                       | 1.2                                      | 10.9 (5.1-19.8)                                     |
| <b>RA</b>                     |                            |                            |                           | <b>RA</b>                       |                                          |                                          |                                                     |
| Study 1                       | 12                         | 8.9                        | 1.35 (0.6-2.1)            | Study 1, n=355                  | 2                                        | 4.3                                      | 0.5 (0.0-1.3)                                       |
| Females, n=512                | 9                          | 6.1                        | 1.5 (0.5-2.4)             | Females, n=203                  | 2                                        | 2.6                                      | 0.8 (0.7-2.3)                                       |
| Males, n=237                  | 3                          | 1.4                        | 2.1 (0.0-4.6)             | Males, n=152                    | 0                                        | 1.9                                      | No reported cases                                   |
| Study 2                       | 12                         | 6.7                        | 1.8 (0.8-2.8)             | Study 2, n=353                  | 2                                        | 3.0                                      | 0.7 (0.2-1.9)                                       |

The prevalence ratios (PRs) obtained for T1DM and hypothyroidism in the present study are similar to those in Neuhausen *et al.* (2008). Our CD sample size is larger than that of Neuhausen *et al.*, and consequently our 95% CIs are narrower. We detect a significant increased risk of hypothyroidism and T1DM for both studies considered as per their 95% CI. While we observe a nominally increased risk for RA, this is not significant as they contain the null hypothesis similar to the results of Neuhausen *et al.*

**Table 3.7 Prevalence ratios (PRs) of co-existent immune mediated conditions IMCs in parents of CD patients using prevalence studies in Table 3.1.**

| Disease                          | N <sub>o.</sub><br>Observed (%) | Female/Male<br>ratio | N <sub>o.</sub><br>Expected (n) | Standardized ratio<br>(95% CI) |
|----------------------------------|---------------------------------|----------------------|---------------------------------|--------------------------------|
| <b>Thyroid Disorders, n 1498</b> | 57 (3.8)                        | 5.3:1                |                                 |                                |
| Study 1                          |                                 |                      | 57.2                            | 1.0 (0.7-1.25)                 |
| Study 2                          |                                 |                      | 75.6                            | 0.75 (0.55-0.95)               |
| <b>Type 1 Diabetes, n 1462*</b>  | 25 (1.7)                        | 1.7:1                |                                 |                                |
| Study 1                          |                                 |                      | 4.4                             | 5.7 (3.3-7.6)                  |
| Study 2                          |                                 |                      | 5.0                             | 5.0 (2.9-6.7)                  |
| <b>Psoriasis, n 1498</b>         | 6 (0.4)                         | 1:1                  |                                 |                                |
| Study 1                          |                                 |                      | 22.5                            | 0.3 (0.06-0.46)                |
| Study 2                          |                                 |                      | 41.9                            | 0.14 (0.04-0.25)               |
| <b>IBD, n 1498</b>               | 10 (0.7)                        | 1.5:1                |                                 |                                |
| Study 1                          |                                 |                      | 4.5                             | 2.2 (1.0-3.6)                  |
| Study 2                          |                                 |                      | 13.0                            | 0.8 (0.3-1.3)                  |
| <b>RA, n 1498</b>                | 9 (0.6)                         | 2:1                  |                                 |                                |
| Study 1                          |                                 |                      | 13.3                            | 0.7 (0.2-1.1)                  |
| Study 2                          |                                 |                      | 7.5                             | 1.2 (0.4-2.0)                  |
| <b>SLE, n 1498</b>               | 2 (0.4)                         | 0:2                  |                                 |                                |
| Study 1                          |                                 |                      | 3.7                             | 0.5 (0.0-1.2)                  |
| Study 2                          |                                 |                      | 1.9                             | 1.05 (0.37-2.4)                |

This analysis shows that the occurrence of T1DM among parents of CD patients (PR=5.7 and PR=5.0 respectively). PRs of 1 or less were observed for all other autoimmune conditions indicating that these observations are not significant. Where a report of a parent with diabetes did not distinguish between type 1 or type 2 diabetes, these individuals (n=36) were excluded (thus n is reduced to 1462 under the type 1 diabetes heading above).

### **3.2.3. Regression analysis to predict the development of any second autoimmunity or ATD in CD patients.**

Ventura *et al.* (1999) reported that the prevalence of co-existent IMCs in CD patients increased with increasing age at diagnosis with CD, using age at diagnosis as a surrogate marker of gluten exposure. As reported in Chapter 2, we also see a continuous increase in the median age of diagnosis of CD patients in our cohort diagnosed in adulthood, (Figure 3.2), suggesting that they may potentially have experienced longer gluten exposure.

Therefore, we wanted to test whether age at diagnosis in patients diagnosed in adulthood predicted co-existent ATD while controlling for other factors such as age of the patient at data collection, gender, family history of IMCs and time period of diagnosis. We have carried out this regression analysis using ATD as dependant variable. We chose to examine ATD as it is the most prevalent IMC in our sample and also has been shown to decrease markedly in CD patients diagnosed in adulthood over time in the bivariate analysis presented in Chapter 2, (Table 2.8).

**Table 3.8 Regression analysis to predict the probability of CD patients diagnosed in adulthood of developing co-existent ATD (n=652).**

| Predictors                                     | Model using having co-existent ATD as dependent variable |        |        |       | Adjusted <sup>£</sup> |                   |        |       |
|------------------------------------------------|----------------------------------------------------------|--------|--------|-------|-----------------------|-------------------|--------|-------|
|                                                | Unadjusted <sup>*</sup>                                  |        | 95% CI |       | P                     | OR                | 95% CI |       |
|                                                | OR                                                       | P      | Lower  | Upper |                       |                   | Lower  | Upper |
| <b>Age of diagnosis with CD (years), n 749</b> |                                                          |        |        |       |                       |                   |        |       |
| ≥18-35                                         | 1.0 <sup>\$</sup>                                        |        |        |       |                       | 1.0 <sup>\$</sup> |        |       |
| >35-55                                         | 1.30                                                     |        | 0.76   | 2.10  |                       | 1.76              | 0.82   | 3.76  |
| > 55                                           | 1.605                                                    | 0.2    | 0.95   | 2.71  | 0.16                  | 2.60              | 0.97   | 6.99  |
| <b>Gender, n 749</b>                           |                                                          |        |        |       |                       |                   |        |       |
| Male                                           | 1.0 <sup>\$</sup>                                        |        |        |       |                       | 1.0 <sup>\$</sup> |        |       |
| Female                                         | 3.99                                                     | <0.001 | 2.33   | 6.85  | <0.001                | 4.70              | 2.65   | 8.28  |
| <b>Patients' age (years), n 749</b>            |                                                          |        |        |       |                       |                   |        |       |
| 18-35                                          | 1.0 <sup>\$</sup>                                        |        |        |       |                       | 1.0 <sup>\$</sup> |        |       |
| >35-55                                         | 1.51                                                     |        | 0.71   | 3.21  |                       | 0.89              | 0.33   | 2.40  |
| >55                                            | 2.69                                                     | 0.003  | 1.33   | 5.41  | 0.63                  | 1.22              | 0.38   | 3.92  |
| <b>Calendar year of CD diagnosis, n 733</b>    |                                                          |        |        |       |                       |                   |        |       |
| ≤1985                                          | 1.0 <sup>\$</sup>                                        |        |        |       |                       | 1.0 <sup>\$</sup> |        |       |
| 1986-1995                                      | 0.61                                                     |        | 0.27   | 1.41  |                       | 0.38              | 0.14   | 1.00  |
| 1996-2005                                      | 0.44                                                     |        | 0.21   | 0.91  |                       | 0.225             | 0.09   | 0.59  |
| 2006-2010                                      | 0.37                                                     |        | 0.18   | 0.76  |                       | 0.23              | 0.08   | 0.627 |
| ≥2011                                          | 0.36                                                     | 0.046  | 0.16   | 0.78  | 0.043                 | 0.225             | 0.075  | 0.67  |
| <b>Family history of second IMC, n 749</b>     |                                                          |        |        |       |                       |                   |        |       |
| Yes                                            | 1.0 <sup>\$</sup>                                        |        |        |       |                       | 1.0 <sup>\$</sup> |        |       |
| No                                             | 0.35                                                     | <0.001 | 0.235  | 0.525 | <0.001                | 0.33              | 0.22   | 0.515 |

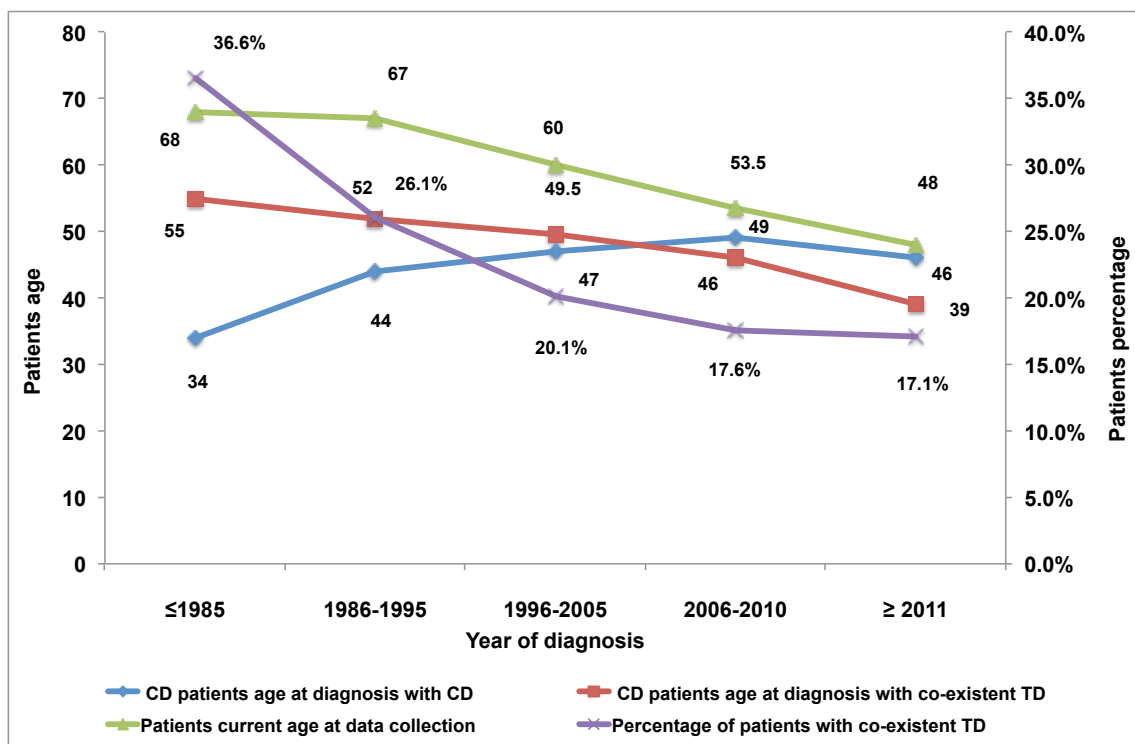
\*Binary logistic regression carried out with each variable individually as predictor of dependent variables.

£ Binary logistic regression carried out with full model adjusted for all variables as predictors of dependent variables; model adjusted for patients' age of diagnosis, gender, age at data collection, calendar year of diagnosis, and family history of co-existent IMC as independent variables, and having co-existent ATD as the dependant variable.

In the unadjusted analysis, female gender ( $p<0.001$ ), being over 35 years at data collection ( $p=0.003$ ), being diagnosed  $\leq 1985$  ( $p=0.046$ ) and having a family history of a second IMC (including ATD) ( $p<0.001$ ) are all significant predictors of co-existent IMC in our full sample of CD patients.

In the full model adjusting for all independent variables, only female gender ( $p<0.001$ ), being diagnosed  $\leq 1985$  ( $p=0.043$ ) and having a family history of a second IMC ( $p<0.001$ ) are significant predictors of co-existent ATD). This could be interpreted as implying either that better diagnostics have led to reduced gluten exposure and therefore prevented co-existent IMCs, or that changing environmental factors are predisposing to CD people with genetic architectures that are linked to CD only and are not linked to other IMCs.

**Figure 3.2 Analysis of age at diagnosis with CD, age at diagnosis with co-existent ATD, median age at data collection and prevalence of co-existent ATD for patients diagnosed in adulthood over 5 time periods (n=652).**



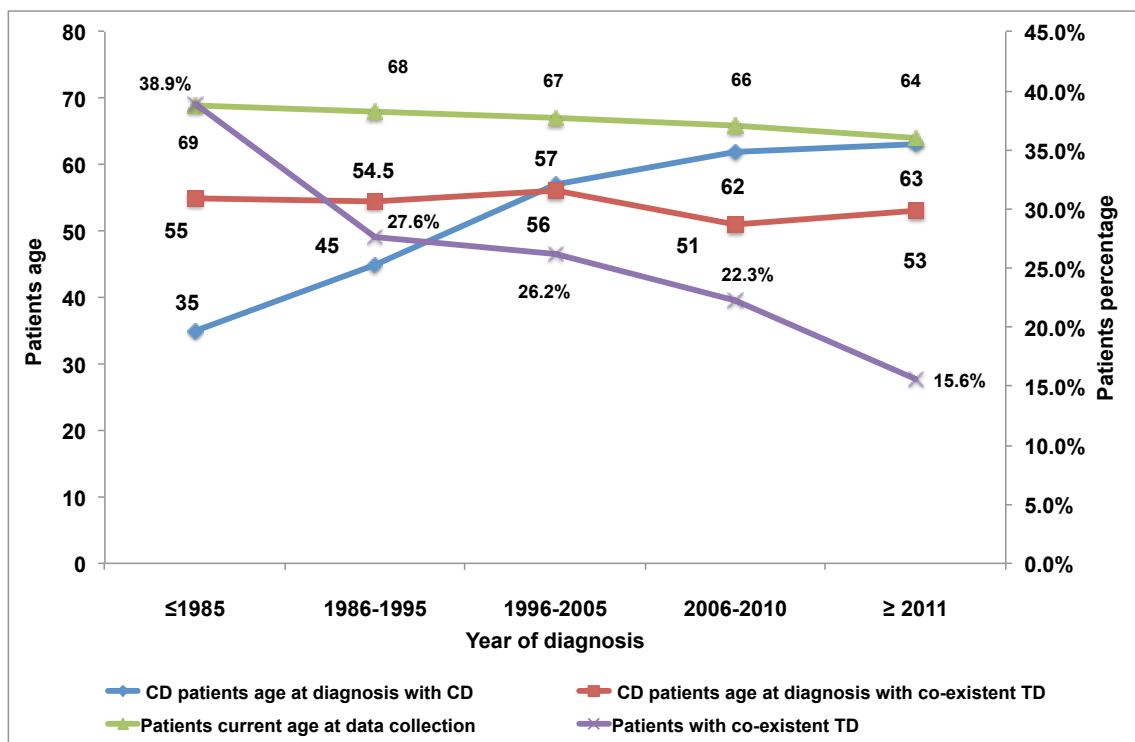
As discussed in Chapter 2, the age at diagnosis of CD was significantly lower before 1986 compared with later periods ( $p=0.002$ ), while there is a significant decrease in the prevalence of associated ATD over time ( $p=0.039$ ). However this may be skewed by the fact that patients in later time periods are younger - i.e. CD patients current age at



data collection reduces significantly over the five time periods, with patients from the two last periods being younger at the time of data collection (53.5 and 48 years respectively) ( $p<0.001$ ). Patients age at diagnosis with co-existent ATD is significantly higher before 1986 compared with later periods ( $p=0.004$ ).

Given the possibility that the decreasing age of patients in later time periods may be depressing the observed prevalence of ATD, we attempted to adjust for this by selecting patients over 55 years of age at data collection in each of the time periods as shown in Figure 3.3.

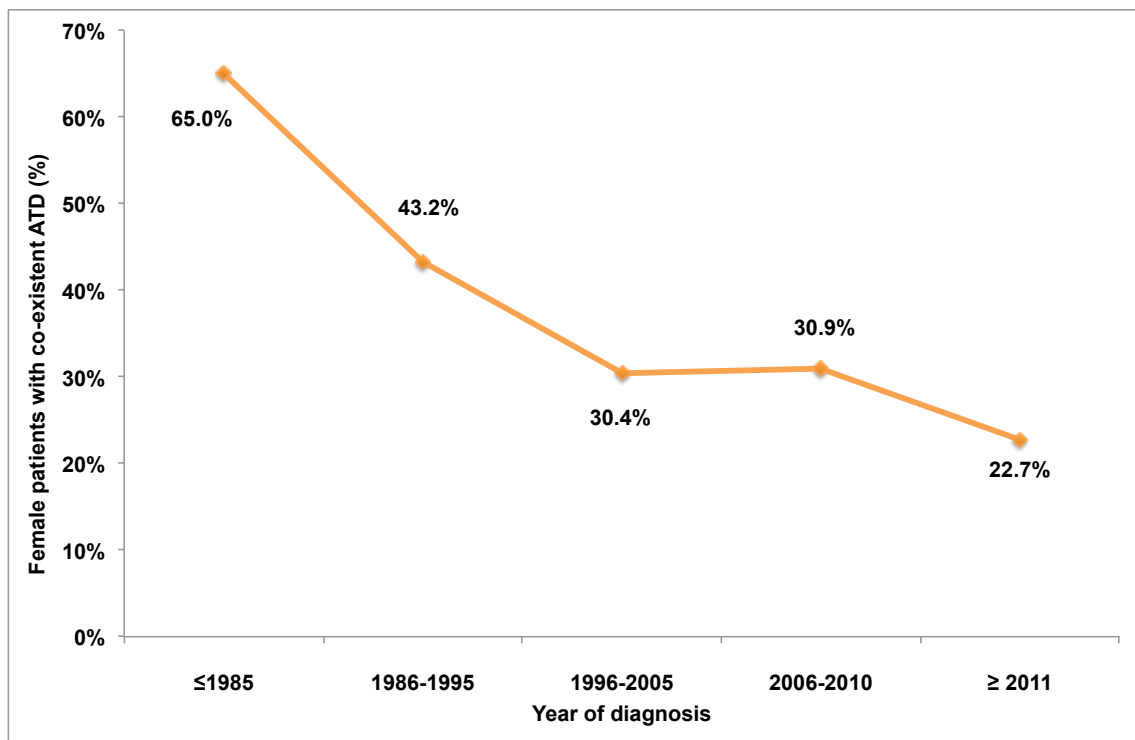
**Figure 3.3 Analysis of parameters in Figure 3.2 with subcohort of patients diagnosed in adulthood and >55 years at data collection over 5 time periods (n=354).**



For those >55 years at data collection, 89 patients had co-existent ATD (79 females); a non-significant trend towards less co-existent ATD over time is still observed, falling from 38.9% ≤1985 to 15.6% ≥2011 ( $p=0.163$ ). However, when stratified by gender, ATD significantly decreases in female patients from 65% ≤1985 to 22.7% ≥2011 ( $p=0.019$ ; not significant in males  $p=0.61$ ). Patient age at data collection is similar over the five time periods (decreasing from 69 to 64 over the 5 time periods) ( $p=0.15$ ). The age at ATD diagnosis remains stable over the five time periods ( $p=0.77$ ), therefore the timing of ATD presentation relative to CD changes completely, presenting mainly prior to CD in the latter time periods.

Therefore, in Figure 3.2, CD patients are diagnosed with ATD at a younger age in the latter periods as a result of the decreasing age of the patients in the most recent periods. However, the prevalence of co-existent ATD still decreases over time when controlling for age (Figure 3.3), indicating that this represents a real decline in ATD prevalence in the latter time periods.

**Figure 3.4 Prevalence of co-existent ATD for female CD patients diagnosed in adulthood and >55 years at data collection divided into 5 periods according to the year of diagnosis (n=226).**



ATD significantly decreases in female patients from 65% ≤1985 to 22.7% ≥2011 ( $p=0.019$ ).

### **3.2.4 Genetics, environmental factors, and epidemiology of co-existent IMCs in CD patients.**

All the co-existent IMCs in our sample share genetic susceptibility factors with CD, as well as potential causative environmental factors. Moreover, many conditions also are similar to CD in that they show an increasing incidence over time. We have summarised all of this information in Table 3.9.

**Table 3.9 Epidemiology, shared genetics and potential environmental factors of co-existent IMCs and CD.**

| <b>Disease</b>                    | <b>HLA<sup>§</sup></b>                                           | <b>Non-HLA<sup>§</sup></b>                                                                                                                                                | <b>Epidemiology</b>                                                                                                                                                                                  | <b>Environmental factors<sup>§</sup></b>                                                                                 |
|-----------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| <b>Type 1 diabetes mellitus</b>   | HLA DQ2 & DQ8                                                    | CTLA4, CCR5, ADAA1, IL21, IL2, TAGAP, SH2B3, NAA25, CTSH, PTPN2                                                                                                           | <1% Prevalence<br>Incidence varies per country ranging from 0.1/100000 to 60/100000<br>Incidence increasing 3-4% per calendar year                                                                   | Rotavirus infection<br>Microbiome dysbiosis<br>Hygiene hypothesis<br>Increased intestinal permeability<br>Dietary gluten |
| <b>Autoimmune Thyroid disease</b> | HLA DQ2 & DQ8                                                    | LPP, ICOS, CTLA4                                                                                                                                                          | ≈ 4% prevalence in Europe<br>0.7/100000 to 2.2/100000 (males)<br>99/100000 to 498.4/100000 (females)<br>Incidence stable over time                                                                   | Dietary gluten exposure                                                                                                  |
| <b>Inflammatory bowel disease</b> | HLA-C, HLA-DQA1, HLA-DRA, HLA-DRB1, HLA-DQB1, HLA-DRB5, HLA-DQA2 | IL3, IL21, ICOSLG, IL18RAP, IL18R1, PTPN2, STAT4, SOCS1, IFNG, IL1RL1, IL1RL2, BACH2, FASLG, UB2L3, ZMIZ, KIF21B, ZFP36L1, PUS10, REL, IRF4, FUT2, IFNG, TNFRSF14, CXCL16 | <1% prevalence<br>24.3/100000 ulcerative colitis & 12.7/100000 crohns disease in Europe.<br>19.2/100000 ulcerative colitis & 20.2/100000 crohns disease in the US.<br>Incidence increasing worldwide | Viral infections<br>Hygiene hypothesis<br>Microbiome dysbiosis<br>Increased intestinal permeability                      |
| <b>Rheumatoid arthritis</b>       | HLA-DRB1                                                         | TNFRSF14, MMEL1, CTLA4, CD28, IL21, IL2, TAGAP, TNFAIP3, ELMO1, CXCR5, DDX6, SH2B3, PTPN11, UBASH3A, IRAK1                                                                | ≈ 1% prevalence<br>Incidence; 53.1/100000 women & 27.7/100000 men in the US<br>Incidence increasing in women since 1995 in the US                                                                    | Viral infections<br>Microbiome dysbiosis                                                                                 |

**Table 3.9 (cont.)**

|                                     |                              |                                            |                                                                                                                                                                                                                                                                                                                                                                            |                                                                |
|-------------------------------------|------------------------------|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|
| <b>Sarcoidosis</b>                  | HLA DR3, HLA DQ2             | CCR2, CCR5                                 | 0.005% - 0.04% prevalence<br>Incidence; 10/100000 US<br>11.5/100000 Sweden<br>Incidence stable over time                                                                                                                                                                                                                                                                   | Not Known                                                      |
| <b>Multiple sclerosis</b>           | HLA-DRB1, HLA DQA1, HLA-DQA2 | RGS1, MMEL1, PTPRK,<br>TAGAP, ZMIZ1, IL12A | > 0.1% prevalence in Europe and North America<br>Incidence; 3.6/100000 women &<br>2.0/100000 men<br>Incidence increasing over time<br>in lower latitudes and in women<br>≈ 2% prevalence in Europe and the US<br>Incidence; ≈ 80/100000 in adults in<br>the US & ≈ 100/100000 for all ages in Europe<br>Incidence increasing over time in children and<br>adults in the US | Viral infections<br>Hygiene hypothesis<br>Microbiome dysbiosis |
| <b>Psoriasis</b>                    | HLA-C, HLA-B                 | TAGAP, ZMIZ1, UBE2L3                       |                                                                                                                                                                                                                                                                                                                                                                            | Viral infections                                               |
| <b>Alopecia areata</b>              | HLA-DQB1                     | SH2B3, ATXN2                               | ≈ 2% prevalence worldwide<br>Incidence; 20.2/100000 in the US<br>Incidence stable up to 1989 in the US                                                                                                                                                                                                                                                                     | Viral infections                                               |
| <b>Systemic Lupus Erythematosus</b> | HLA-DRB1                     | TNFAIP3, SH2B3, UBE2L3                     | < 1% prevalence                                                                                                                                                                                                                                                                                                                                                            | Viral infections                                               |

\$ HLA loci, Non-HLA loci and potential environmental factors shared with CD

### 3.3 Discussion

In this chapter we have investigated the frequency of co-existent IMCs (other than CD) as well as their burden in CD patients compared to the general population. Due to the decrease in prevalence of ATD over time reported in Chapter 2 (Table 2.8), we have also studied the association of factors such as age of diagnosis with CD, time period of CD diagnosis and patient's age at data collection on ATD prevalence. The age of diagnosis with CD is important given the hypotheses pointing to gluten exposure as a predisposing agent in the development of co-existent IMCs in CD patients (Ventura et al., 1999, Ventura et al., 2000, Cosnes et al., 2008). Finally, the age of the patients at data collection is also of interest since the frequency of ATD increases in older age groups (Reid and Wheeler, 2005, Garber et al., 2012).

We have also analysed the prevalence of IMCs and CD in parents of CD patients. Earlier studies have confirmed higher prevalence of both IMCs and CD in the relatives of CD patients (Neuhausen et al., 2008, Emilsson et al., 2015, Fasano et al., 2003). Genome wide association studies (GWAS) have demonstrated shared genetic predisposing factors between various IMCs, including CD which provides a logical explanation for the observed higher prevalence of these IMCs in both CD patients and their relatives (Immunobase; [www.immunobase.org](http://www.immunobase.org)) (Gutierrez-Achury et al., 2011). However, shared environmental factors are also likely to play a role in the development of both IMCs and CD in these population groups (Emilsson et al., 2015).

The prevalence of co-existent IMCs among CD cases in this study is 31.2%, in line with previous findings (Volta et al., 2014, Kocsis et al., 2013). This is higher than the estimated prevalence in the general population of these IMCs of 3-9.4% (Jacobson et al., 1997, Eaton et al., 2007, Cooper et al., 2009). As with similar studies in adult CD, we find the most common co-existent IMCs are the endocrine type disorders, T1DM and ATD (Kocsis et al., 2013, Volta et al., 2014, Collin et al., 2002, Fasano, 2006). The total number of observed cases of T1DM (n=27) and ATD (n=149) we observed among CD cases were higher than expected based on prevalence studies in the literature. The prevalence ratios (PRs) and 95% confidence intervals (CI) we obtain for T1DM and ATD are in line with those found in a comparable analysis of a North American CD cohort (Table 3.6) (Neuhausen et al., 2008). We found that in the majority of cases, T1DM precedes CD, whereas ATD occurs equally either prior to, or after, CD diagnosis, in agreement with other studies (Elfstrom et al., 2008, Greco et al., 2013). We also found that the number of parents with T1DM was significantly greater than expected, as shown previously by Neuhausen *et al.* (PR=5.7 and PR=5.0 respectively)

(Table 3.7). CD patients' parents were not significantly associated with other IMCs. This could be because autoimmune conditions are relatively rare and our study is insufficiently powered to detect them. The presence of Type I diabetes in parents is certainly important and leads to the conclusion of perhaps children of Type I diabetes should be checked for celiac disease.

### **3.3.1 CD and Type 1 Diabetes Mellitus (T1DM).**

The predisposition of CD patients to develop T1DM is well established (Ludvigsson et al., 2006, Neuhausen et al., 2008). GWAS have shown that shared susceptibility loci between T1DM and CD are not limited to the well described HLA class II associations, but include at least eight non-HLA loci documented on the ImmunoBase database encoding candidate genes including among others, *CTLA4* (encoding cytotoxic lymphocyte antigen), the chemokine gene *CCR5*, and *SH2B3* and *PTPN2* which encode molecules involved in intracellular signalling (Table 3.9) (Immunobase.org) (Gutierrez-Achury et al., 2011). Therefore shared susceptibility genes seem to be a logical explanation for the co-existence of both conditions. However, common environmental factors associated with the initiation of both conditions are also widely believed to be important (Table 3.9) (Honeyman et al., 2000, Stene et al., 2006, Okada et al., 2010, Hu et al., 2015).

Some studies point to dietary gluten, the main environmental trigger in CD, playing a role in the pathogenesis of T1DM, though its mechanism is not fully understood (Serena et al., 2015). As in the case of CD patients, it has been proposed that gliadin may provoke an opening of intercellular tight junctions (perhaps mediated by zonulin) causing a 'leaky' gut in T1DM patients, leading to an abnormal ingress of luminal antigens to the sub mucosal layer (Visser et al., 2009). This abnormal influx of antigens, leads to the activation of the intestinal immune system response, and has been suggested to be an important catalyst in the development of T1DM (Secondulfo et al., 2004, Sapone et al., 2006, Vaarala, 2008). In fact, zonulin upregulation has been observed to precede the onset of T1DM (Vaarala, 2008). This implies that gliadin activated T-cells may be important in the molecular pathogenesis of T1DM and potentially playing a role in the destruction of pancreatic beta cells (Visser et al., 2009).

Gluten introduction during infancy may also be a factor in T1DM development in genetically predisposed individuals. Notably, the cumulative incidence and the onset of T1DM has been shown to be reduced in mice born of NOD mothers fed with a GFD. There is evidence that the protection of the NOD mouse from diabetes by a gluten-free diet may be in part mediated through alteration in the microbiome (Marietta et al.,

2013). Moreover a GFD prevents and reduces the incidence of T1DM in NOD mice and BB rats (Schmid et al., 2004, Hansen et al., 2014, Antvorskov et al., 2016). While this has not been proven in humans, a GFD has been shown to improve insulin sensitivity in T1DM patients (Pastore et al., 2003, Fuchtenbusch et al., 2004). A case study reported remission of T1DM in a child without CD and on a GFD without insulin therapy, however this effect has not been tested in large cohorts of patients (Sildorf et al., 2012).

Differences in disease distribution in close or neighbouring geographical areas, as well as increasing incidence of CD and T1DM cannot easily be explained by variation in allele frequencies or improved diagnostic techniques alone, but may well be explained by variable environmental exposures (Ludvigsson et al., 2013b, Tuomilehto, 2013, Kang et al., 2013, Borchers et al., 2010). Interestingly, it has been demonstrated that the proportion of newly diagnosed T1DM patients with a higher risk HLA susceptibility genotype has decreased over time although the frequency of these higher risk alleles has remained stable. On the other hand, the percentage of new T1DM patients with lower risk genotypes has increased, especially in younger patients. These findings point strongly towards a role of environmental factors which are now able to cause the development of T1DM on a genotypic background that would not previously have been sufficient to cause disease (Fourlanos et al., 2008, Vehik et al., 2008, Steck et al., 2011). No studies similar to these intriguing findings exist in CD, which would help clarify the role of environmental factors in the increasing incidence of CD.

### **3.3.2 CD and autoimmune thyroid disease (ATD).**

In this retrospective study on adult CD patients we confirm previous reports that ATD is also observed frequently with CD (Sategna-Guidetti et al., 2001, Collin et al., 2002, Neuhausen et al., 2008, Meloni et al., 2009, Metso et al., 2012). CD and ATD shared genetic susceptibility loci (Table 3.9) (Immunobase.org) (Ch'ng et al., 2007, Hadithi et al., 2007a). As far as common environmental factors for ATD and CD are concerned, some studies have suggested tTGAs in CD patients may promote the development of ATD (Naiyer et al., 2008, Duntas, 2009). Naiyer *et al.* observed how tTGAs bind to tissue transglutaminase II in thyroid tissue, and that tTGA titers correlate with those of antithyroidperoxidase antibodies (TPO-AB). Although they did not show proof of thyrocyte damage, it suggests that tTGAs deposition could lead eventually to thyroid dysfunction.

Various studies have shown binding of tTGAs to tissues such as the liver, the brain and the heart, which were associated with extraintestinal manifestations in CD such as

autoimmune hepatitis and gluten ataxia (Sategna-Guidetti et al., 2004, Korponay-Szabo et al., 2004, Hadjivassiliou et al., 2006, Hadjivassiliou et al., 2013, van Gerven et al., 2014). These findings point towards a pathogenic role of humoral immunity in CD patients. Naiyer *et al.* (2008) observed that CD patients positive for thyroid antibodies were older than patients with CD only. Similarly, in a study of ATD, the risk of co-existent CD was higher in patients aged over 65 years (peak incidence is between the 5<sup>th</sup> and 6<sup>th</sup> decades of life) compared to younger patients and healthy controls (Ravaglia et al., 2003). Thus, it is interesting to speculate whether long term exposure to gluten/tTGA has increased the rate of comorbidity among this older age group.

However, two studies exploring ATD development in CD patients have failed to see any protective effect of the GFD (Meloni et al., 2009, Metso et al., 2012). Metso *et al.*, followed a cohort of newly diagnosed CD patients in Finland for a year. The development of overt or subclinical ATD was analysed at baseline and one year later, and compared to that of non-coeliac controls on a gluten-containing diet. Adherence to the gluten free diet (GFD) was comprehensively measured by serology, a control biopsy and review by an experienced dietician (Ludvigsson et al., 2014b). At baseline there was no significant difference in the number of patients with raised TPO-AB levels between the CD group and controls. However, TPO-AB titers were significantly higher in CD patients at baseline and continued increasing during follow-up, whereas those of controls remained stable. Metso *et al.* found no improvement in the progression of ATD in the CD cohort despite adherence to the GFD; on the contrary, median thyroid volume decreased in CD patients when compared to controls (Metso et al., 2012). An earlier prospective study found that whilst non-immune-mediated subclinical hypothyroidism normalised for most patients on adoption of a strict GFD, this effect was not observed for ATD (Sategna-Guidetti et al., 2001).

Here, we report for the first time a significant reduction in ATD over a 55 year period in a cohort of CD patients diagnosed in adulthood (Chapter 2, Table 2.8). One confounder in our sample of CD patients diagnosed in adulthood is that there is an age bias, with the patients in the more recent time periods being significantly younger at data collection compared to those in the earlier time periods (Figure 3.2). ATD has a peak incidence between the 5<sup>th</sup> and 6<sup>th</sup> decade (Eber and Langsteger, 1994). This age bias could explain our previously reported changes in ATD (Chapter 2, Table 2.8). Thus we would be observing less ATD over time simply because the latter cohorts are younger. However, we still observe a decrease in ATD when selecting for the subcohort of CD patients diagnosed >55 years of age at data collection (Figure 3.3). In this subgroup segregated by gender, ATD decreases significantly for women over time (Figure 3.4).



ATD is reportedly 5-10 times more common in women than men (Garber et al., 2012). We observe a similar female bias (Table 3.5), which is hypothesised to be due to sex hormones predisposing to ATD (Chiovato et al., 1993, Ngo et al., 2014).

In a logistic regression analysis adjusting for age at CD diagnosis, gender, age at data collection, and time period of CD diagnosis, we also observe that the age of the patients at data collection is not a predictor of co-existent ATD in the adjusted model ( $p=0.63$ ) (Table 3.8). The only predictors of co-existent ATD in our fully adjusted model are female gender, family history of second IMC and time period of CD diagnosis (Table 3.8). Based on Naiyer's findings, it could be hypothesised that better diagnostic tools and screening of at-risk patients has lead to CD being diagnosed at an earlier stage, thus avoiding long term exposure to tTGAs and decreasing the prevalence of ATD. However, considering the change in genotypic makeup for T1DM patients, it could also be that an analogous situation has occurred in CD leading to a progressive diminution of patients with a genotype predisposing to both CD and ATD as discussed in Chapter 2 (Section 2.4.3). The latter may be supported by the fact that whereas CD incidence is increasing, this does not seem the case for ATD (McLeod and Cooper, 2012, Jozkow et al., 2017).

Therefore, in the present study, we do not observe a protective effect of reduced gluten exposure in the unadjusted or adjusted model; the independent variable, age at diagnosis with CD, is not a significant predictor of co-existent ATD ( $p=0.19$  and  $p=0.17$  respectively) (Table 3.8). We have used the age at diagnosis with CD as an indirect measure of gluten exposure, meaning that in theory, gluten exposure ended when the patient was diagnosed with CD and from that point on followed a GFD. As mentioned earlier, this approach was used by Ventura *et al.* (1999) who found that the prevalence of various co-existent IMCs, including ATD, in CD patients correlated with gluten exposure, with risk increasing with the age at diagnosis of CD.

Potential explanations for this difference in findings are first of all, that the cohort in Ventura's paper is young (mean age  $16.1 \pm 3.8$ ) whereas our patients were diagnosed in adulthood (median age 57, range 18-91). Therefore, it is possible that the age at CD diagnosis plays a role in preventing co-existent IMCs in childhood due to GFD adoption, but not in adulthood (Metso et al., 2012). Moreover, in Ventura's study, all CD patients were EMA negative, suggesting compliance with the diet; however, we did not collect GFD adherence data over time, and thus our age at CD diagnosis is potentially a less accurate measure of gluten exposure.

In fact, gluten exposure is a multidimensional concept, and several parameters should be considered; namely, delay in diagnosis from the onset of symptoms, age at diagnosis, compliance with GFD after diagnosis, serology, biopsy and assessment by an expert CD dietician during the follow-up period (Leffler et al., 2007b, Metso et al., 2012). It could be that those diagnosed in adulthood in the earlier time periods had a longer delay between the onset of symptoms and diagnosis due to less awareness among physicians and less specific diagnostic techniques. Patients may also have been less adherent to the GFD due to poor availability of gluten free products. On the other hand, high CD awareness in Ireland and regular monitoring at secondary referral centres would potentially promote adherence to GFD and minimise delayed diagnosis.

Increased gluten consumption worldwide, together with other environmental factors, may underlie a true rise in CD incidence in recent years (Kasarda, 2013, Lebwohl et al., 2014b). In fact, the amount of gluten consumed by infants <2 years of age seems to play a role in the process of tTGA seroconversion in genetically predisposed infants, increasing the risk of the condition by 2-fold in those consuming larger amounts (Andren Aronsson et al., 2015, Szajewska et al., 2016). A recent study in Indian adults found that regional variations in CD prevalence were correlated to wheat intake despite a similar genetic background (Ramakrishna et al., 2016).

### **3.3.3 CD and inflammatory bowel disease (IBD).**

There is a significant increased risk of co-existent IBD in our cohort with an IBD prevalence of 1.9%, and PRs 6.4 and 2.3 for study 1 and 2 respectively (Table 3.5). This is in line with similar studies which have reported a prevalence range 2.2% to 3.3%, or a range of 2 to 10 fold increased risk of patients with CD having co-existent IBD (Bulger et al., 1988, Yang et al., 2005, Leeds et al., 2007, Kocsis et al., 2015). Nearly 200 non-HLA loci are now associated with IBD development, compared with more than 40 non-HLA loci for CD (Jostins et al., 2012, Withoff et al., 2016). A large proportion of the CD non-HLA risk loci are shared with IBD which provides an evident shared genetic background between both conditions (Table 3.9) (Pascual et al., 2014).

Both diseases also share common potential environmental factors involved in disease initiation such as, viral infections, the hygiene hypothesis and microbiome dysbiosis (Stene et al., 2006, Bosca-Watts et al., 2015, Gent et al., 1994, Kondrashova et al., 2008, Sanz, 2015, Elson and Cong, 2012). This is supported by familial studies which have found an increased risk of developing IBD not only in first degree relatives of CD patients, but curiously, also in their spouses (Mayberry et al., 1986, Emilsson et al., 2015). In summary, sharing of genetic susceptibility and environmental factors are

likely important causes for the co-existence of both diseases. Moreover, in common with CD and T1DM, IBD patients have increased intestinal permeability which plays a pivotal role in the development of the condition (Mankertz and Schulzke, 2007, Jager et al., 2013, Antoni et al., 2014).

Diet, among other factors, seems to play a role both in the pathogenesis of IBD and in the induction of flares after development of the condition. Nutritional factors such as high fibre diets and vitamin D seem to be protect against IBD while others, such as linoleic acid in red meat, predispose to it (Dutta and Chacko, 2016). Although gluten is not known to play a role in pathogenesis, a study of IBD patients by Herfarth *et al.* (2014) reported large percentages of them voluntarily adopting a GFD after diagnosis: of 1647 IBD patients nearly 314 had ever followed a GFD after diagnosis, with more than 60% of this subcohort reporting an improvement of their symptoms and nearly 40% reporting less IBD flares. Herfarth *et al.* concluded that there is a subcohort of IBD patients for whom a GFD is beneficial, possibly composed of patients carrying HLA-DQ2 or HLA-DQ8 haplotype. (Herfarth et al., 2014). Interestingly, There is also data on patients with diarrhea predominant IBS reporting that those patients carrying HLA DQ2/8 may be more likely to respond to gluten withdrawal (Vazquez-Roque et al., 2013).

Their assumption is based on studies which showed a favourable effect of the GFD on patients with irritable bowel syndrome who carried HLA-DQ2 but did not have CD (negative tTGA and normal biopsy) (Wahnschaffe et al., 2007). They conclude that, as hypothesised in the case of patients with non-coeliac gluten sensitivity, gluten might have an effect on intestinal permeability and the mucosal immune system in IBD patients carrying HLA-DQ2 or HLA-DQ8 genotype (Herfarth et al., 2014). Recently a Japanese study found a high prevalence of IBD patients with positive tTGA compared to controls, however none had CD according to biopsy examination however only a small minority had high risk CD HLA genotypes. Nevertheless, patients who followed a GFD reported an improvement of their symptoms (Watanabe et al., 2014).

It is notable that, similar to T1DM and CD, IBD incidence is increasing worldwide (Maahs et al., 2010, Tuomilehto, 2013, Ludvigsson et al., 2013, Catassi et al., 2014, Benchimol et al., 2009, M'Koma, 2013). However, whereas for T1DM and IBD the increasing incidence is affecting more children, in CD this increasing incidence is mainly observed in groups >45 years of age (Ye et al., 2015, Ludvigsson et al., 2013) as we have also seen (Chapter 2, Figure 2.3).

### 3.3.4 CD comorbidities with other immune-mediated conditions (IMCs).

In this section we will discuss the association of CD with the remaining IMCs in Table 3.5, namely rheumatoid arthritis (RA), sarcoidosis, multiple sclerosis (MS), systemic lupus erythematosus (SLE) and alopecia areata (AA). All of these are characterized by sharing non-HLA susceptibility loci with CD, and in some cases, they also share HLA associations (Table 3.9).

Although we observe an increased risk of co-existent RA in our CD patients cohort, only the 95% CI for study 2 is significant (Table 3.5). Using the same studies as Neuhausen *et al.*, there is still an apparent increase in risk, however none of the 95% CIs are significant (Table 6.3). Unlike us however, Francis *et al.* (2002) and Neuhausen *et al.* (2008) have shown lesser risk of developing co-existent CD in RA patients and RA in CD patients respectively (Francis *et al.*, 2002, Neuhausen *et al.*, 2008). Our significant association may be due to increased power since our sample is larger than those of Neuhausen *et al.* (n=408) and Francis *et al.* (n=160).

The Neuhausen *et al.* study is the most comparable to ours, as it also comprises a cohort of CD patients and the authors compare the prevalence of RA to that in the general population using published studies. Our different findings could be due to power, as mentioned, but also Neuhausen's CD sample is younger than ours (median ages 42 vs. 56 years). Although RA can occur at any age, its prevalence increases with age and incidence has been increasing in recent decades in women >50 years of age (Rasch *et al.*, 2003, Neuhausen *et al.*, 2008, Myasoedova *et al.*, 2010, Peschken and Hitchon, 2012, Innala *et al.*, 2014). Therefore, the fact that our sample is older and female biased could be skewing our results.

It might be expected that we would see an increased level of comorbidity between CD and RA given several issues. The incidence for both conditions is increasing and both putatively share common environmental factors (Table 9.3) (Hafstrom *et al.*, 2001, Elkan *et al.*, 2008, Lerner and Matthias, 2015b). RA also has a strong association with class II HLA genes, in particular HLA-DRB1\*04 alleles which are present on the HLA-DR4-DQ8 haplotype associated with CD (Laivoranta-Nyman *et al.*, 2004). Both conditions also share a substantial number of non-HLA genes thought to affect immune cell activation and differentiation (Koning *et al.*, 2015, Lerner and Matthias, 2015b).

Sarcoidosis has been associated with a variety of other immunological disorders including CD: having CD predisposes to sarcoidosis and *vice versa* (Ludvigsson *et al.*, 2007, D'Ercole *et al.*, 2012). Apart from the shared genetic predisposition between CD

and sarcoidosis (Table 3.9) (Hwang et al., 2008, Fingerlin et al., 2015), it is possible that a common antigen triggers both conditions, and it has been suggested that once one condition has developed, increased expression of HLA class II molecules predisposes to the other (D'Ercole et al., 2012).

Estimates of sarcoidosis prevalence vary considerably because approximately 90% of cases are asymptomatic, or present with symptoms which correlate with other more common conditions (i.e. asthma) (Hwang et al., 2008). This explains its lower prevalence in population-based data compared to data obtained from autopsies. It is for this reason that we have included a prevalence range for both of the sarcoidosis general population studies above (Table 3.5). We observe a sarcoidosis prevalence of 0.9% which approaches the 1.15% reported by Hwang *et al.* (2008) in a sample of 866 CD patients. Similarly to earlier studies, we observe an increased risk of developing co-existent sarcoidosis with 3 of the 4 95% CIs being significant. Unlike for CD, T1DM, IBD and RA, the incidence of sarcoidosis is reported to be stable over time (Arkema et al., 2016, Ungprasert et al., 2016), although Ungprasert et al. note that the age of peak incidence in women and men is increasing, similar to CD.

There is also a nominally increased risk of MS in our cohort, however none of the 95% CIs are significant. The relationship between MS and CD has been reported in several studies, however the results are still controversial with some reporting higher levels of CD antibodies, such as tTGAs, in MS patients compared to controls and others not seeing this association (Salvatore et al., 2004, Reichelt and Jensen, 2004, Pengiran Tengah et al., 2004, Nicoletti et al., 2008, Shor et al., 2009, Rodrigo et al., 2011). Neurological symptoms are observed in approximately 10-36% of CD patients diagnosed in adulthood, with cerebellar ataxia being the most common, but other symptoms such as dementia, seizures and occipital calcifications are also found in these patients.

Moreover, neurological manifestations often present on their own at CD diagnosis in the absence of any gastrointestinal signs or symptoms (Bushara, 2005, Burk et al., 2009). IgA antibodies against gluten have a high affinity for the blood brain barrier circulatory system which potentially can cause an increase in permeability (Pratesi et al., 1998, Reichelt and Jensen, 2004). Transglutaminase 6 (TG6) has specifically been implicated in the genesis of gluten ataxia (Hadjivassiliou et al., 2008, Hadjivassiliou et al., 2013). The beneficial effects of a GFD to reverse, or improve such neurological manifestations is controversial as neurological symptoms in CD can develop even with strict adherence to the GFD (Kieslich et al., 2001, Hadjivassiliou et al., 2002, Burk et al., 2009).

Again there is a genetic overlap between MS and CD (Table 3.9) (Immunobase.org) (Gutierrez-Achury et al., 2011, Withoff et al., 2016). A familiar scenario of MS incidence increasing over time is also observed, here however driven by an increase in the southern latitudes where it was previously less prevalent, and an increase in the female to male ratio (Alonso and Hernan, 2008). Again this is attributed to altered environmental exposures, some of which overlap with CD (Table 3.9) (Alonso and Hernan, 2008). Despite the genetic and potential environmental overlap with MS, as well as possible noxious effects of gluten in the central nervous system, whether MS and other CD linked neurological manifestations are part of the myriad of CD manifestations or an epiphenomenon is still under debate (de Oliveira et al., 2016, Freedman et al., 2017).

The association of CD and psoriasis has yielded contradictory results in population and clinical studies. Some have reported increased risk of CD patients developing psoriasis and *vice versa*. Increased CD serology in psoriatic patients has also been reported; however other studies did not find this connection. However, the studies showing a negative association tended to have smaller samples and in some cases did not include controls (Bhatia et al., 2014). In our case, study 1 for psoriasis yielded an increased psoriasis risk of (1.7) in our patients which was borderline significant (95%CI: 1.06-2.7). Two meta analyses have reported a 2.4-fold increased risk of having CD antibody positivity and 3-fold increase in the risk of developing CD respectively in patients with psoriasis (Bhatia et al., 2014, Ungprasert et al., 2017).

Shared predisposing genes are again a feature of the CD/psoriasis relationship (Abenavoli et al., 2006, Trynka et al., 2011, Tsoi et al., 2015, Withoff et al., 2016). Increased intestinal permeability is also found in psoriasis (Humbert et al., 1991, Montalto et al., 2002, Vaarala, 2008). Moreover, vitamin D deficiency in CD patients has been proven to predispose to psoriasis which responds positively to treatment with topical vitamin D analogues (Orgaz-Molina et al., 2012, Tavakkoli et al., 2013). Two US studies have reported increased incidence of psoriasis over a 30 year period both in adults and children (Icen et al., 2009, Tollefson et al., 2010).

CD has also been associated with alopecia areata (AA) with some studies reporting that the GFD can lead to hair regrowth while others do not see this association (Corazza et al., 1995, Volta et al., 1997, Bondavalli et al., 1998, Fessatou et al., 2003). We do not observe an increased risk of AA in our cohort of CD patients (Table 3.5),

and we are not aware of similar studies looking for AA prevalence in CD patients - most studies have been done in AA patients looking for the presence of CD. This lack of association could be due to insufficient power of our sample or under reporting of the condition. Like CD, AA is a polygenic autoimmune condition with both HLA and non-HLA associations, some of which overlap with T1DM, RA and CD (Table 3.9) (Immunobase.org) (Colombe et al., 1999, de Andrade et al., 1999, Petukhova et al., 2010, Petukhova and Christiano, 2016).

Interestingly, as in the case of psoriasis, lower levels of vitamin D have been found in patients with AA compared to controls (Bakry et al., 2016, Ghafoor and Anwar, 2017). In the case of CD patients, Tavakkoli *et al.* (2013) found psoriasis, but not AA, to be increased in CD patients with vitamin D deficiency (Tavakkoli et al., 2013). However, it should be noted that only 55 of their patients had a second autoimmune condition (with the majority being psoriasis, n=19). Therefore, it cannot be ruled out that low or deficient vitamin D levels may predispose CD patients to develop AA.

Systemic lupus erythematosus (SLE) and CD comorbidity is rare, mainly arising in case reports (Criscov et al., 2015, Alves et al., 2016). The largest population study on the co-existence of CD and SLE was carried out in 29,000 biopsy proven CD patients, which showed that CD patients were 3 times more likely to have SLE. This increased risk was found 5 years after diagnosis, however the absolute risk was low (Ludvigsson et al., 2012). In our study, although we observe an increased prevalence ratio for the two studies considered, both of the CIs are non-significant (Table 3.5). Studies looking for positive CD serology in SLE patients have reported low numbers with positive serology or, and in most of these the biopsy did not support a diagnosis of CD (Marai et al., 2004).

### **3.4 Conclusion**

The results of this study support previous findings that CD patients have an increased occurrence of IMCs. This is true as well for CD patients parents and T1DM in line with previous studies. Sharing of genetic predisposing loci partially explain the co-existence of CD with other IMCs, however shared potential environmental factors could also play a role in this association. The latter seems to be supported by the fact that many of the co-existent IMCs have an increasing incidence over time.

Time period of CD diagnosis, and not age at CD diagnosis or age of patients at data collection, together with female gender and family history of autoimmunity have a significant effect in predisposing to co-existent ATD in our CD cohort. These findings may point to a change in CD susceptibility genotypes over time as seen with T1DM, with maybe the pool of patients with a predisposition only to CD and not to ATD having enlarged in recent decades in response to changing environmental factors.



## **Chapter 4 Genetic Burden and Coeliac Disease Phenotype**

## 4.1 Introduction

While there have been tremendous developments in the genetics of coeliac disease (CD) there is a lack of understanding of how genotypes affect clinical phenotypes of the disease. As previously discussed, CD is a highly heterogeneous condition at disease presentation, with symptoms ranging from sub-clinical to severe malabsorptive syndromes at diagnosis (Rubio-Tapia et al., 2013). It can occur at any age, with a considerable proportion of patients being diagnosed over 50 years of age (Rashtak and Murray, 2009). This variable clinical picture can lead to delayed diagnosis if it is not the first option considered for investigation (Vilppula et al., 2008). Undiagnosed CD has been linked to increased mortality. Therefore, early detection is important, especially for individuals who are at high risk, such as those who have an affected first degree relative, or an associated immune-mediated condition (IMC), such as type 1 diabetes mellitus (T1DM) or autoimmune thyroid disease (ATD) (Abraham et al., 2015).

Currently, CD diagnosis in adults is based on the presence of raised anti-tissue transglutaminase antibodies (tTGA), followed by a confirmatory small bowel biopsy to show villous atrophy (Ludvigsson et al., 2014a). However, this approach, whilst highly accurate in detecting active CD, does not provide any insight on the prognosis of the condition, or on the risk of developing CD in a person without active disease (Abraham et al., 2015). Genetic risk prediction represents a promising tool to salvage the previous mentioned disadvantages of the current diagnostic method (Romanos et al., 2014, Abraham et al., 2015). A characteristic feature of CD is the need of susceptibility genes encoding for HLA-DQ2 and/or DQ8 which are seen in approximately 99.6% of all individuals with CD (Karell et al., 2003).

Therefore, due to the pivotal role of the HLA genes, the most recent CD diagnostic guidelines recommend HLA genotyping as a first-line investigation for patients who are self-treated on the gluten free diet (GFD) at diagnosis and those classified as high risk, such as first degree relatives of CD patients (Husby et al., 2012, Ludvigsson et al., 2014a). Moreover, in 2012 new guidelines from the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) allow the omission of a biopsy in children with CD symptoms based on the presence of tTGA ten or more times the normal upper limit and who are positive for HLA DQ2/DQ8 (Husby et al., 2012). Nevertheless, HLA genotyping *per se* in CD diagnosis is inadequate, given that approximately 40% of the general population express HLA-DQ2 and DQ8 and do not develop CD (Sollid, 2002, Karell et al., 2003). Thus, whereas HLA genotyping has high sensitivity and negative predictive value when the risk genotypes are absent in an

individual, they are poor predictors and have low specificity when present (Megiorni et al., 2009).

Since 2007, genome wide association studies (GWAS) have found many non-HLA single nucleotide polymorphisms (SNPs) associated with CD (Dubois et al., 2010, Trynka et al., 2011, Withoff et al., 2016). These non-HLA SNPs only contribute a small overall increased risk of CD development (Dubois and van Heel, 2008). Although these GWAS SNPs account for only a small proportion of disease heritability, evidence in other conditions has been accumulating showing that a larger proportion of phenotypic heterogeneity can be explained by the grouping of these SNPs with modest effects using polygene risk scores (PRS) (Dudbridge, 2013, Purcell et al., 2009, Bush et al., 2010). PRS scores have been built in most cases from the SNPs found to be highly associated with CD, using stringent p-values due to multiple-testing corrections performed in GWAS. However they have also been designed to allow SNPs which fail to achieve genome wide significance (GWS) to be included in genetic models of complex disease, where PRS including SNPs at lower p-values have been shown to increase predictive capacity (Purcell et al., 2009, Dudbridge, 2013, Abraham et al., 2014).

PRS have been proven to differentiate between clinical subtypes and different types of presentations in certain conditions such as bipolar disorder, inflammatory bowel disease (IBD), and rheumatoid arthritis (RA) (Chibnik et al., 2011, Ananthakrishnan et al., 2014, Charney et al., 2017). In CD, several studies have used PRS formed from an ensemble of HLA and non-HLA SNPs to show improvement in the identification of high risk individuals (Romanos et al., 2009, Abraham et al., 2015, Coleman et al., 2015). However, whilst correlation of PRS with overall disease risk has been well explored, the association of the genetic burden with CD clinical phenotype has not been investigated. Indeed there is little evidence as to the effects of genotype on clinical phenotype in CD. The studies available have been limited in terms of size, and have mainly focused on the HLA predisposing alleles or in a small subset of non-HLA genes with inconclusive results (Greco et al., 1998, Mustalahti et al., 2002, Zubillaga et al., 2002, Gudjonsdottir et al., 2009). Moreover, most of these studies were carried out before the recent genetic findings using fine mapping.

Therefore the aim of this chapter is to study whether we can observe an association of genetic burden as measured using PRS, with CD patients' clinical phenotype. In addition, we analysed whether there were changes in CD patients' PRS over time, given the variability in CD clinical presentation reported in Chapter 1 and 2; less classical presentation over time, increasing median age at diagnosis and less co-

existent autoimmune thyroid disease (ATD), and the reported genotype changes over time in T1DM which is highly associated with CD (Furlanos et al., 2008, Vehik et al., 2008, Steck et al., 2011). We furthermore wished to use this information to identify genetic variants, which may be associated with these clinical changes.

## 4.2 Aims

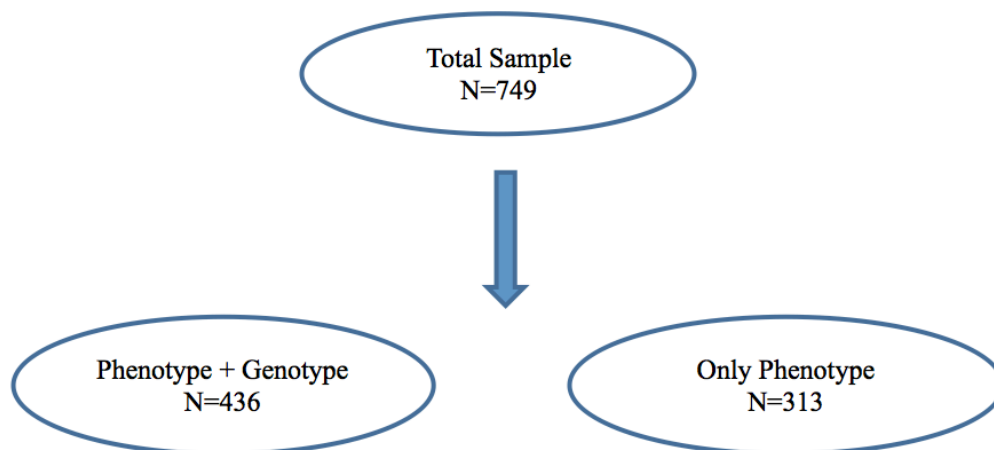
- Correlation of genetic burden and CD clinical phenotype considering the following characteristics; age at CD diagnosis, gender, CD clinical presentation at diagnosis, positive CD family history and presence of co-existent ATD.
- Analysis of the evolution of CD patients PRS over time.
- Investigation of any potential changes in gene variants frequencies that would contribute to potential PRS changes observed over time.

## 4.3 Methods

### 4.3.1 Study design and study population

A retrospective analysis of medical records was performed using a CD patient cohort (n=749) attending five secondary referral centres in Ireland and diagnosed from 1960-2015. Study design was as described in Chapter 2, section 2.2.1. We matched 436 CD patients' phenotypic information to their Immunochip genotype information using PRS (Figure 4.1) (supplementary Table 4, appendix XI) .

**Figure 4.1 Diagram showing total sample and its subsets classified according to availability of genotype information.**



### 4.3.2 Clinical phenotypes

The following information was collected for CD patients; age at CD diagnosis, gender, CD clinical presentation at diagnosis, positive CD family history, year of diagnosis with CD and presence of co-existent ATD (see Chapter 2, section 2.2.4.1 & Chapter 3 section 3.3.2). The clinical presentation at diagnosis was categorised according to the Oslo classification as classical, non-classical and subclinical using the previously recorded manifestations (Ludvigsson et al., 2013a) (see supplementary Table 1, appendix III). A categorical variable was also created for the year or time period in

which the diagnosis was made, as follows; 1)  $\leq 1985$ , 2) 1986-1995, 3) 1996-2005 4) 2006-2010 and 5)  $\geq 2011$  (see Chapter 1, section 2.2.4.1).

For the genotype/phenotype analysis, the variable 'age at diagnosis with CD' was transformed into a binary variable as follows;  $<40$  years and  $\geq 40$  years based on the median age at diagnosis for the full sample (see Chapter 2, Table 2.1). Similarly, the variable 'clinical presentation at diagnosis' was combined into a binary variable including classical presentation and non-classical or subclinical presentation. The variable for the year of diagnosis was condensed into three time periods ( $\leq 1995$ , 1996-2005 and  $\geq 2006$ ) due to the low number of patients available in the first time period ( $\leq 1985$ ).

### **4.3.3 Genotype/phenotype analysis method**

Two-tailed t-tests conducted on SPSS version 24.0 (SPSS, Inc., Chicago, IL) were carried out to detect potential differences between PRS for (i) all markers (HLA & non-HLA), (ii) HLA only and (iii) non-HLA PRS at different p-levels and 1) age at CD diagnosis, 2) CD clinical presentation at diagnosis, 3) gender, 4) positive family history at CD diagnosis, and 5) co-existence of ATD. Conservative adjustment assuming independence of the 5 tests carried out yielded a Bonferroni-corrected p-value threshold of  $p=0.01$

Analysis of variance (ANOVA) was carried out using SPSS to capture potential changes in genotype over time using PRS calculated for all markers, HLA only and non-HLA, at different p-value levels and the 3 category year of diagnosis variable described above. A p-value of  $p=0.05$  was considered statistically significant in this analysis.

### **4.3.4 PRS calculation**

Genotyping was performed using the Illumina Infinium High-Density array by Dr C. Coleman. Genotypes were called using the Illumina GenomeStudio software using the cluster set of 172,242 autosome/x---chromosome variants from the Trynka *et al.* (2001) study. Quality control was performed using PLINK software by Dr. R Anney and R. McManus. Samples showing poor genotyping rates (poor quality DNA), excessive heterozygosity (sample contamination), and excessive relatedness (2nd degree and closer relatives) with other samples in the cohort were excluded. The population was screened for population outliers (by cross comparison with European, East Asian and African genotypes) as is standard practice. Genetic markers (SNPs) were removed due to low call rates ( $<97\%$ ) or low allele frequency ( $<5\%$ ). Following quality control,

143,074 markers were available for analysis. Part of this analysis has been reported previously (Coleman et al., 2015), and further genotypes were added for this study.

PRS were generated using the --score flag in PLINK by Dr R. Anney. A discovery data set (Trynka et al., 2011) was used to define SNPs reaching genome wide significance (GWS) and the odds ratios (OR) of disease associated alleles. To ensure SNPs were independent of each other, GWS SNPs were pruned for linkage disequilibrium as described below. Individual risk scores for each disease associated allele were generated using the log of the OR for that allele; PRS for each individual were derived by summing the risk scores by the number of alleles for each GWS locus, giving a cumulative polygene risk score (PRS) for that individual.

Thus ten PRSs were generated at p-value association thresholds of  $p < 5 \times 10^{-8}$ ,  $p < 1 \times 10^{-7}$ ,  $p < 1 \times 10^{-6}$ ,  $p < 1 \times 10^{-5}$ ,  $p < 1 \times 10^{-4}$ ,  $p < 1 \times 10^{-3}$ ,  $p < 0.01$ ,  $p < 0.05$ ,  $p < 0.1$  and  $p < 0.5$ . This means that SNPs with p-values lower than these thresholds from the Trynka discovery set were included. PRS for patients for whom clinical data is available were matched to their correspondent clinical phenotype. PRS containing all markers (i.e. both HLA & non-HLA loci), PRS containing HLA only loci and PRS containing non-HLA only loci were considered in the analysis of phenotype/genotype. The HLA region was defined as the region on chromosome 6 between nucleotide positions 24,000,000 and 34,000,000 (hg38 genome assembly).

rs2187668-A is in high linkage disequilibrium ( $r^2=0.97$ ) with the HLA-DQA1\*0501 DQB1\*0201 (DQ2.5) haplotype and is used as a proxy for this haplotype. We used rs21857668 to genotype this haplotype. rs2187668-G represents all other haplotypes.

#### **4.3.4 Analysis of changing alleles frequencies over time**

In order to identify SNPs that contribute to the change of PRSs over the various time periods, we focused on CD patients who were diagnosed before 1995 ( $n=133$ ) or after 2006 ( $n=184$ ) and compared frequencies of risk alleles between the two groups. We then chose SNP sites that were used for the PRS calculation under different p-value cut-offs;  $p < 10^{-1}$  for the HLA region (149 SNPs) and  $p < 10^{-6}$  for non-HLA alleles (126 SNPs). Frequencies of risk alleles at these SNP sites were calculated for each of the two groups separately, which are defined as  $f_{1995}$  and  $f_{2006}$ . Differences in the allele frequencies were given as  $\Delta_f = f_{2006} - f_{1995}$ , so that positive or negative values represent the increase or decrease respectively of risk allele frequencies from 1995 to 2006. The relationship between  $\Delta_f$  and association coefficients is shown in Figure 4.12.



To identify SNPs that show a significant increase or decrease of allele frequency from <1996 to >2006, we simulated data that model changes of allele frequencies over a single generation by taking account of the different sample sizes in the two periods. This analysis was carried out by Dr. S. Nakagome. First, we randomly sampled the time ( $t$ ) when a risk allele appeared in a population from a uniform distribution with the condition of  $0 < t < 4N_0$ , where  $N_0$  is a current population size and assumed as 120,000 (Abecasis et al., 2010).

We then generated the trajectory of the risk allele forward in time under the Wright-Fisher model. If  $t$  survives to the current generation ( $t=0$ ) and one generation before the present ( $t=1$ ), we calculated the allele frequency of  $f_0$  and  $f_1$  by randomly sampling the number of chromosomes carrying the risk allele from a binomial distribution with the parameters of  $n=133$  and  $f_1$  or of  $n=184$  and  $f_0$ . We repeated these steps for 1000 times and generated a distribution of  $\Delta_f$ .

Because common alleles are more likely to have a large variance than rare alleles, we separately simulated the trajectories by stratifying values of  $f_1$  into 10 different classes: 0%-10%, 10%-20%, 20%-30%, 30%-40%, 40%-50%, 50%-60%, 60%-70%, 70%-80%, 80%-90%, and 90%-100%. We tested the deviation of observed  $\Delta_f$  from 95% confidence intervals of the simulated distributions. Violin plots showing the distribution of alleles are shown in Figures 4.13 and 4.14.

## 4.4 Results

Our objective was to correlate genetic inheritance with clinical presentation, and also to identify potential genotype changes over time. We had 436 patients with detailed phenotypic and immunochip genotype information with approximately 313 samples lost to follow up. PRSs were generated and matched to their corresponding patient clinical phenotype. We considered every PRS significance p-value threshold to identify the threshold that would maximise our ability to detect differences in clinical phenotype.

#### 4.4.1 Correlation of HLA & Non-HLA, HLA only and Non-HLA PRS with clinical phenotype.

**Table 4.1 Mean PRS for all markers according to phenotypic characteristics of CD patients.**

Mean PRS for all markers (HLA & non-HLA combined) for phenotypic characteristics (i) age of CD diagnosis divided as <40 years or ≥40 years, (ii) CD clinical presentation at diagnosis divided as classical or non-classical/subclinical, (iii) positive CD family history and (iv) presence of autoimmune thyroid disease at different p-values are presented.

| CHARACTERISTIC        |                                 | TOTAL N | PRS at p=5×10 <sup>-8</sup> | P-Value | PRS at p=1×10 <sup>-7</sup> | P-Value | PRS at p=1×10 <sup>-6</sup> | P-Value | PRS at p=1×10 <sup>-5</sup> | P-Value | PRS at p=1×10 <sup>-4</sup> | P-Value |
|-----------------------|---------------------------------|---------|-----------------------------|---------|-----------------------------|---------|-----------------------------|---------|-----------------------------|---------|-----------------------------|---------|
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |         | 0.20829                     |         | 0.19918                     |         | 0.17646                     |         | 0.16354                     |         | 0.14354                     |         |
|                       | ≥ 40 years of age (249)         | 436     | 0.20737                     | 0.442   | 0.19849                     | 0.556   | 0.17561                     | 0.399   | 0.16271                     | 0.293   | 0.14285                     | 0.275   |
| CLINICAL PRESENTATION | Classical (195)                 |         | 0.20752                     |         | 0.19839                     |         | 0.175595                    |         | 0.16281                     |         | 0.14307                     |         |
|                       | Non-Classical/Subclinical (147) | 342     | 0.20755                     | 0.980   | 0.19895                     | 0.670   | 0.17609                     | 0.663   | 0.16316                     | 0.694   | 0.14310                     | 0.972   |
| GENDER                | Male (133)                      |         | 0.20617                     |         | 0.19725                     |         | 0.17469                     |         | 0.16196                     |         | 0.14224                     |         |
|                       | Female (303)                    | 436     | 0.20846                     | 0.074   | 0.19947                     | 0.078   | 0.176535                    | 0.090   | 0.16356                     | 0.060   | 0.14355                     | 0.052   |
| CD FAMILY HISTORY     | Yes (154)                       |         | 0.21002                     |         | 0.20078                     |         | 0.17787                     |         | 0.16461                     |         | 0.14451                     |         |
|                       | No (282)                        | 436     | 0.20653                     | 0.005   | 0.19770                     | 0.011   | 0.174935                    | 0.005   | 0.16223                     | 0.004   | 0.14241                     | 0.001   |
| THYROID DISEASE       | Yes (92)                        |         | 0.20855                     |         | 0.19945                     |         | 0.17669                     |         | 0.16344                     |         | 0.14363                     |         |
|                       | No (344)                        | 436     | 0.20755                     | 0.492   | 0.19861                     | 0.559   | 0.17578                     | 0.461   | 0.16297                     | 0.627   | 0.14302                     | 0.421   |
| CHARACTERISTIC        |                                 | TOTAL N | PRS at p=1×10 <sup>-3</sup> | P-Value | PRS at p=0.01               | P-Value | PRS at p=0.05               | P-Value | PRS at p=0.1                | P-Value | PRS at p=0.5                | P-Value |
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |         | 0.10632                     |         | 0.07132                     |         | 0.04927                     |         | 0.04145                     |         | 0.02570                     |         |
|                       | ≥ 40 years of age (249)         | 436     | 0.10584                     | 0.212   | 0.07108                     | 0.174   | 0.04920                     | 0.371   | 0.04141                     | 0.370   | 0.02567                     | 0.198   |
| CLINICAL PRESENTATION | Classical (195)                 |         | 0.10598                     |         | 0.07118                     |         | 0.04922                     |         | 0.04141                     |         | 0.025675                    |         |
|                       | Non-Classical/Subclinical (147) | 342     | 0.10606                     | 0.859   | 0.07113                     | 0.826   | 0.04922                     | 0.995   | 0.041425                    | 0.815   | 0.02569                     | 0.600   |
| GENDER                | Male (133)                      |         | 0.10551                     |         | 0.070945                    |         | 0.04913                     |         | 0.04136                     |         | 0.02567                     |         |
|                       | Female (303)                    | 436     | 0.10628                     | 0.063   | 0.07128                     | 0.074   | 0.049275                    | 0.069   | 0.04146                     | 0.064   | 0.02569                     | 0.263   |
| CD FAMILY HISTORY     | Yes (154)                       |         | 0.10686                     |         | 0.07156                     |         | 0.04939                     |         | 0.04153                     |         | 0.02572                     |         |
|                       | No (282)                        | 436     | 0.10561                     | 0.002   | 0.07097                     | 0.001   | 0.049145                    | 0.002   | 0.04137                     | 0.003   | 0.02566                     | 0.007   |
| THYROID DISEASE       | Yes (92)                        |         | 0.10632                     |         | 0.07127                     |         | 0.04927                     |         | 0.04145                     |         | 0.02567                     |         |
|                       | No (344)                        | 436     | 0.105975                    | 0.465   | 0.07116                     | 0.604   | 0.04922                     | 0.627   | 0.04142                     | 0.685   | 0.025685                    | 0.616   |

In Table 4.1 we observe a significant association between genetic burden and patients with a positive CD family history compared to those without, when using PRS for all markers (i.e. HLA & non-HLA) at all p-value thresholds. Those with a family history have a higher PRS indicating a higher genetic load for CD. This finding is significant at the corrected p-value ( $p=0.01$ ) for all PRS p-value cut-off points (bold red). Gender shows a trend towards higher risk scores for females however this is not significant. None of the other clinical characteristics show any association with genetic factors as measured by these PRS.

**Table 4.2 Mean PRS for HLA only markers according to phenotypic characteristics of CD patients.**

Mean PRS for HLA only markers for phenotypic characteristics (i) age of CD diagnosis divided as <40 years or ≥40 years, (ii) CD clinical presentation at diagnosis divided as classical or non-classical/subclinical, (iii) positive CD family history and (iv) presence of autoimmune thyroid disease at different p-values are presented.

| CHARACTERISTIC        |                                 | TOTAL N | PRS at $p=5 \times 10^{-8}$ | P-Value      | PRS at $p=1 \times 10^{-7}$ | P-Value      | PRS at $p=1 \times 10^{-6}$ | P-Value      | PRS at $p=1 \times 10^{-5}$ | P-Value      | PRS at $p=1 \times 10^{-4}$ | P-Value      |
|-----------------------|---------------------------------|---------|-----------------------------|--------------|-----------------------------|--------------|-----------------------------|--------------|-----------------------------|--------------|-----------------------------|--------------|
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |         | 0.29552                     |              | 0.29385                     |              | 0.28247                     |              | 0.26834                     |              | 0.25866                     |              |
|                       | ≥ 40 years of age (249)         | 436     | 0.29442                     | 0.536        | 0.29263                     | 0.490        | 0.28092                     | 0.347        | 0.26734                     | 0.556        | 0.25765                     | 0.524        |
| CLINICAL PRESENTATION | Classical (195)                 |         | 0.29450                     |              | 0.29272                     |              | 0.28118                     |              | 0.26735                     |              | 0.25776                     |              |
|                       | Non-Classical/Subclinical (147) | 342     | 0.29490                     | 0.841        | 0.29313                     | 0.837        | 0.28145                     | 0.885        | 0.26770                     | 0.854        | 0.25795                     | 0.914        |
| GENDER                | Male (133)                      |         | 0.29238                     |              | 0.29083                     |              | 0.27938                     |              | 0.26538                     |              | 0.25589                     |              |
|                       | Female (303)                    | 436     | 0.29600                     | <b>0.058</b> | 0.29418                     | <b>0.078</b> | 0.28255                     | <b>0.074</b> | 0.26882                     | <b>0.059</b> | 0.25904                     | <b>0.063</b> |
| CD FAMILY HISTORY     | Yes (154)                       |         | 0.29790                     |              | 0.29606                     |              | 0.28459                     |              | 0.270665                    |              | 0.26072                     |              |
|                       | No (282)                        | 436     | 0.29325                     | <b>0.011</b> | 0.29157                     | <b>0.014</b> | 0.27994                     | <b>0.006</b> | 0.26619                     | <b>0.011</b> | 0.25664                     | <b>0.013</b> |
| THYROID DISEASE       | Yes (92)                        |         | 0.29609                     |              | 0.29420                     |              | 0.28263                     |              | 0.26887                     |              | 0.25919                     |              |
|                       | No (344)                        | 436     | 0.29458                     | 0.484        | 0.29288                     | 0.536        | 0.28130                     | 0.600        | 0.26747                     | 0.497        | 0.25779                     | 0.463        |
| CHARACTERISTIC        |                                 | TOTAL N | PRS at $p=1 \times 10^{-3}$ | P-Value      | PRS at $p=0.01$             | P-Value      | PRS at $p=0.05$             | P-Value      | PRS at $p=0.1$              | P-Value      | PRS at $p=0.5$              | P-Value      |
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |         | 0.25297                     |              | 0.24445                     |              | 0.24323                     |              | 0.24327                     |              | 0.23083                     |              |
|                       | ≥ 40 years of age (249)         | 436     | 0.25173                     | 0.407        | 0.24347                     | 0.509        | 0.24208                     | 0.411        | 0.24176                     | 0.235        | 0.22955                     | 0.302        |
| CLINICAL PRESENTATION | Classical (195)                 |         | 0.25189                     |              | 0.24351                     |              | 0.24231                     |              | 0.242185                    |              | 0.22989                     |              |
|                       | Non-Classical/Subclinical (147) | 342     | 0.25203                     | 0.934        | 0.24380                     | 0.862        | 0.24246                     | 0.924        | 0.24205                     | 0.923        | 0.22982                     | 0.956        |
| GENDER                | Male (133)                      |         | 0.25051                     |              | 0.24199                     |              | 0.24069                     |              | 0.24079                     |              | 0.22847                     |              |
|                       | Female (303)                    | 436     | 0.25303                     | <b>0.116</b> | 0.24473                     | <b>0.085</b> | 0.24340                     | <b>0.070</b> | 0.24312                     | <b>0.088</b> | 0.23082                     | <b>0.076</b> |
| CD FAMILY HISTORY     | Yes (154)                       |         | 0.25491                     |              | 0.24651                     |              | 0.24491                     |              | 0.24481                     |              | 0.23240                     |              |
|                       | No (282)                        | 436     | 0.25081                     | <b>0.008</b> | 0.24246                     | <b>0.008</b> | 0.24130                     | <b>0.012</b> | 0.24109                     | <b>0.005</b> | 0.22884                     | <b>0.005</b> |
| THYROID DISEASE       | Yes (92)                        |         | 0.25291                     |              | 0.24474                     |              | 0.24337                     |              | 0.24296                     |              | 0.23072                     |              |
|                       | No (344)                        | 436     | 0.252805                    | 0.649        | 0.243665                    | 0.551        | 0.24236                     | 0.552        | 0.24226                     | 0.648        | 0.22994                     | 0.601        |

In Table 4.2 we observe that HLA-only markers also show a significant association ( $p=0.01$ ) for the patient subgroup with a positive family history. This is significant at all of the PRS p-value cut-off points (bold red) (rounded p-value considered). There is again a non-significant trend indicating that females (reaching a p-value of 0.058 at  $p=1\times 10^{-8}$ ) have a higher PRS than males. None of the other phenotypic comparisons were significant.

Given that we see a higher genetic burden in individuals with a family history of CD, we asked whether this translated into more severe clinical presentation. However these higher PRS do not seem to be related to disease severity as measured by comparing classical malabsorptive presentation between the subcohorts with and without a family history (55.6% and 57.9% respectively,  $p=0.73$ ), and age at diagnosis (42 and 44 years respectively,  $p=0.083$ ) (data not shown).

**Table 4.3 Mean PRS for non-HLA only markers according to phenotypic characteristics of CD patients.**

Mean PRS for non-HLA only markers for phenotypic characteristics (i) age of CD diagnosis divided as <40 years or ≥40 years, (ii) CD clinical presentation at diagnosis divided as classical or non-classical/subclinical, (iii) positive CD family history and (iv) presence of autoimmune thyroid disease at different p-values are presented.

| CHARACTERISTIC        | TOTAL N                         | PRS at $p=5 \times 10^{-8}$ | P-Value      | PRS at $p=1 \times 10^{-7}$ | P-Value | PRS at $p=1 \times 10^{-6}$ | P-Value | PRS at $p=1 \times 10^{-5}$ | P-Value | PRS at $p=1 \times 10^{-4}$ | P-Value      |
|-----------------------|---------------------------------|-----------------------------|--------------|-----------------------------|---------|-----------------------------|---------|-----------------------------|---------|-----------------------------|--------------|
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |                             |              | 0.07962                     |         | 0.079025                    |         | 0.068445                    |         | 0.06663                     |              |
|                       | ≥ 40 years of age (249)         | 436                         | 0.911        | 0.07971                     | 0.874   | 0.06806                     | 0.537   | 0.06628                     | 0.477   | 0.06792                     | 0.459        |
| CLINICAL PRESENTATION | Classical (195)                 |                             |              | 0.08007                     |         | 0.07918                     |         | 0.06835                     |         | 0.06648                     |              |
|                       | Non-Classical/Subclinical (147) | 342                         | 0.523        | 0.07945                     | 0.712   | 0.06825                     | 0.889   | 0.066525                    | 0.935   | 0.06806                     | 0.992        |
| GENDER                | Male (133)                      |                             |              | 0.07919                     |         | 0.07860                     |         | 0.06794                     |         | 0.06612                     |              |
|                       | Female (303)                    | 436                         | 0.449        | 0.07988                     | 0.537   | 0.06835                     | 0.539   | 0.06657                     | 0.388   | 0.06814                     | 0.308        |
| CD FAMILY HISTORY     | Yes (154)                       |                             |              | 0.08010                     |         | 0.07946                     |         | 0.06883                     |         | 0.06696                     |              |
|                       | No (282)                        | 436                         | 0.452        | 0.07944                     | 0.331   | 0.06790                     | 0.145   | 0.06615                     | 0.107   | 0.06775                     | <b>0.027</b> |
| THYROID DISEASE       | Yes (92)                        |                             |              | 0.08042                     |         | 0.07938                     |         | 0.06859                     |         | 0.06675                     |              |
|                       | No (344)                        | 436                         | 0.357        | 0.07947                     | 0.570   | 0.06813                     | 0.544   | 0.06635                     | 0.493   | 0.06799                     | 0.696        |
| CHARACTERISTIC        | TOTAL N                         | PRS at $p=1 \times 10^{-3}$ | P-Value      | PRS at $p=0.01$             | P-Value | PRS at $p=0.05$             | P-Value | PRS at $p=0.1$              | P-Value | PRS at $p=0.5$              | P-Value      |
| AGE OF DIAGNOSIS      | < 40 years of age (187)         |                             |              | 0.06148                     |         | 0.0489                      |         | 0.03989                     |         | 0.03570                     |              |
|                       | ≥ 40 years of age (249)         | 436                         | 0.173        | 0.06121                     | 0.256   | 0.039885                    | 0.897   | 0.035685                    | 0.605   | 0.02352                     | 0.403        |
| CLINICAL PRESENTATION | Classical (195)                 |                             |              | 0.06135                     |         | 0.04885                     |         | 0.03989                     |         | 0.03569                     |              |
|                       | Non-Classical/Subclinical (147) | 342                         | 0.928        | 0.06137                     | 0.813   | 0.03989                     | 0.911   | 0.03569                     | 0.852   | 0.02353                     | 0.711        |
| GENDER                | Male (133)                      |                             |              | 0.06116                     |         | 0.048785                    |         | 0.03985                     |         | 0.03566                     |              |
|                       | Female (303)                    | 436                         | 0.268        | 0.061395                    | 0.394   | 0.03990                     | 0.276   | 0.035705                    | 0.194   | 0.02352                     | 0.840        |
| CD FAMILY HISTORY     | Yes (154)                       |                             |              | 0.06163                     |         | 0.048955                    |         | 0.039935                    |         | 0.03572                     |              |
|                       | No (282)                        | 436                         | <b>0.020</b> | 0.06116                     | 0.068   | 0.03986                     | 0.100   | 0.035675                    | 0.134   | 0.02352                     | 0.283        |
| THYROID DISEASE       | Yes (92)                        |                             |              | 0.06143                     |         | 0.04887                     |         | 0.03991                     |         | 0.03570                     |              |
|                       | No (344)                        | 436                         | 0.559        | 0.06130                     | 0.735   | 0.03988                     | 0.538   | 0.03569                     | 0.861   | 0.02353                     | 0.236        |

Table 4.3 shows that no significant associations were found when using the non-HLA only PRS at the different p-values. However, we again observe, as in the case of HLA & Non-HLA, and HLA only PRS, a higher non-HLA genetic burden in individuals who reported a family history of CD. This reaches borderline significance at the corrected p-value 0.01, at PRS p-value levels  $p=1\times10^{-4}$  and  $p=1\times10^{-3}$  ( $p=0.027$  and  $p=0.020$  respectively).

We conclude from these PRS analyses that the only clinical variable that can be predicted by our PRS is a positive family history of CD. Thus, individuals with a family history of CD have a significantly higher HLA & non-HLA, and HLA only genetic burden, with the non-HLA only also showing also a tendency to be higher in this group, but not reaching statistical significance.

In the next Table, we show our results when the same clinical phenotype variables as in Tables 4.1, 4.2 and 4.3 were analysed, segregating the sample into those with and without a family history of CD. The rationale was; given that this group seems to have a higher genetic burden, we may be able to better predict their clinical phenotype by genetic means, i.e. by using the PRS. This better prediction of PRS using selected subcohorts of cases, has previously been reported in psychiatric disorders, which, similarly to CD, are highly heterogeneous in their presentation (So and Sham, 2017). Presented in Table 4.4 are results for HLA & non-HLA PRS at selected p-value levels.



**Table 4.4 Correlation of clinical phenotype with mean PRS for all markers (HLA & non-HLA) segregated into patients with a family history of CD (n=154; Table A) and those without, a family history of CD (n=282;Table B).**

| Table A               |                                |         |                             |              |                             |              |                 |              |                 |         |
|-----------------------|--------------------------------|---------|-----------------------------|--------------|-----------------------------|--------------|-----------------|--------------|-----------------|---------|
| CHARACTERISTIC        |                                | TOTAL N | PRS at $p=5 \times 10^{-8}$ | P-Value      | PRS at $p=1 \times 10^{-6}$ | P-Value      | PRS at $p=0.01$ | P-Value      | PRS at $p=0.05$ | P-Value |
| AGE OF DIAGNOSIS      | < 40 years of age (72)         | 154     | 0.21241                     | <b>0.018</b> | 0.17990                     | <b>0.016</b> | 0.07197         | <b>0.005</b> | 0.04952         | 0.037   |
|                       | ≥ 40 years of age (82)         |         | 0.20791                     |              | 0.17609                     |              | 0.07120         |              | 0.04927         |         |
| CLINICAL PRESENTATION | Classical (70)                 | 126     | 0.20919                     | 0.729        | 0.17690                     | 0.475        | 0.071485        | 0.795        | 0.049375        | 0.913   |
|                       | Non-Classical/Subclinical (56) |         | 0.20990                     |              | 0.17812                     |              | 0.07156         |              | 0.04936         |         |
| GENDER                | Male (42)                      | 154     | 0.209595                    | 0.787        | 0.177405                    | 0.720        | 0.07144         | 0.577        | 0.04931         | 0.437   |
|                       | Female (112)                   |         | 0.210175                    |              | 0.17805                     |              | 0.07161         |              | 0.049415        |         |
| THYROID DISEASE       | Yes (92)                       | 436     | 0.20855                     | 0.492        | 0.17669                     | 0.461        | 0.07127         | 0.604        | 0.04927         | 0.627   |
|                       | No (344)                       |         | 0.20755                     |              | 0.17578                     |              | 0.07116         |              | 0.04922         |         |

| Table B               |                                |         |                             |              |                             |              |                 |              |                 |              |
|-----------------------|--------------------------------|---------|-----------------------------|--------------|-----------------------------|--------------|-----------------|--------------|-----------------|--------------|
| CHARACTERISTIC        |                                | TOTAL N | PRS at $p=5 \times 10^{-8}$ | P-Value      | PRS at $p=1 \times 10^{-6}$ | P-Value      | PRS at $p=0.01$ | P-Value      | PRS at $p=0.05$ | P-Value      |
| AGE OF DIAGNOSIS      | < 40 years of age (115)        | 282     | 0.20570                     | 0.358        | 0.17431                     | 0.413        | 0.07091         | 0.632        | 0.04911         | 0.566        |
|                       | ≥ 40 years of age (167)        |         | 0.20710                     |              | 0.17537                     |              | 0.07102         |              | 0.04917         |              |
| CLINICAL PRESENTATION | Classical (125)                | 216     | 0.20658                     | 0.788        | 0.17486                     | 0.989        | 0.07100         | 0.606        | 0.04914         | 0.995        |
|                       | Non-Classical/Subclinical (91) |         | 0.20611                     |              | 0.17484                     |              | 0.07087         |              | 0.049135        |              |
| GENDER                | Male (115)                     | 282     | 0.204585                    | <b>0.071</b> | 0.17344                     | <b>0.103</b> | 0.07072         | <b>0.114</b> | 0.04904         | <b>0.137</b> |
|                       | Female (167)                   |         | 0.20746                     |              | 0.17565                     |              | 0.07109         |              | 0.04919         |              |
| THYROID DISEASE       | Yes (56)                       | 282     | 0.20811                     | 0.289        | 0.17637                     | 0.259        | 0.07120         | 0.310        | 0.049175        | 0.753        |
|                       | No (226)                       |         | 0.206135                    |              | 0.17458                     |              | 0.07092         |              | 0.04914         |              |

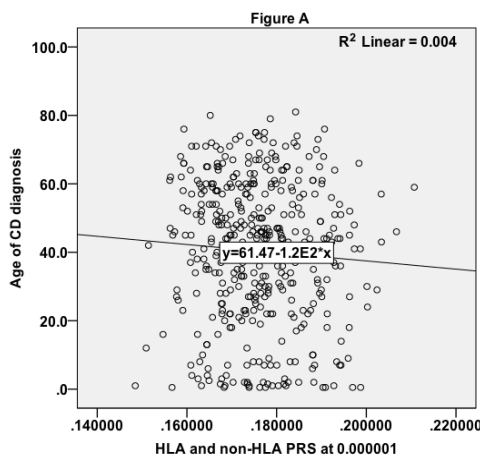
In table 4.4A we observe that HLA & non-HLA PRS are significantly negatively correlated with age of diagnosis  $<40$  and  $\geq 40$  for patients with a positive family history. This is not seen in patients without CD family history (Table 4.4B). The trend observed previously in differences by gender is not seen in the subcohort with a family history, although females consistently have a higher nominal mean PRS (Table 4.4A), but is still evident in the other subcohort (Table 4.4B). The latter observation could plausibly be due to the smaller sample size in Table 4.4A.

#### 4.4.2 Correlation between age of diagnosis with CD and HLA & non-HLA, HLA only and non-HLA only PRS.

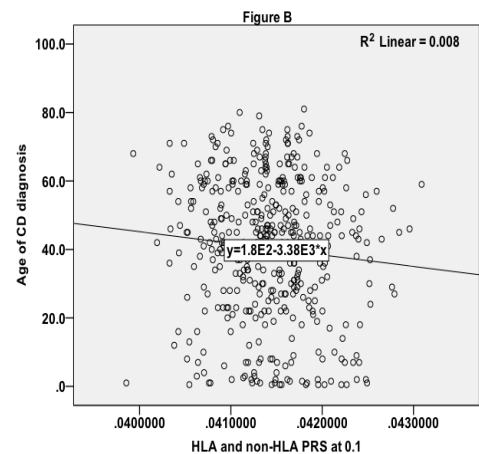
Given the variability in the age at which CD is first diagnosed, we wanted to investigate if this parameter could be detected in differences in PRS. Therefore, we correlated the continuous variable 'age of CD diagnosis' with all the PRS considered at two different p-value levels, to identify any potential associations between them. The following figures show plots representative of the general pattern.

**Figure 4.2 Correlation between PRS for all markers (HLA & non-HLA) and age of CD Diagnosis.**

Scatterplots showing correlation between age of diagnosis with CD and HLA & non-HLA PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B).



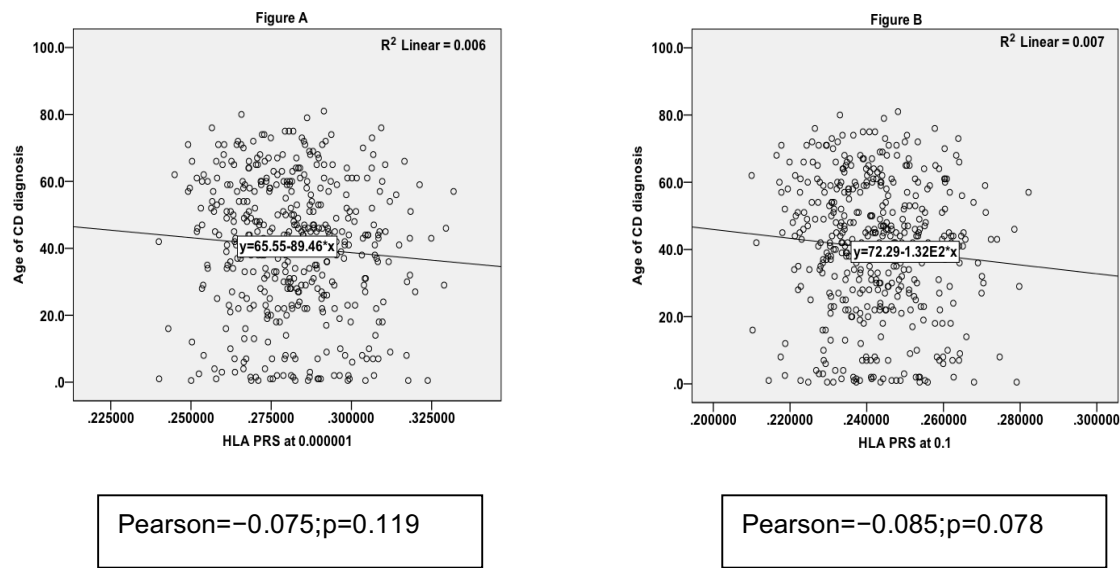
Pearson=-0.061;p=0.201



Pearson=-0.088;p=0.065

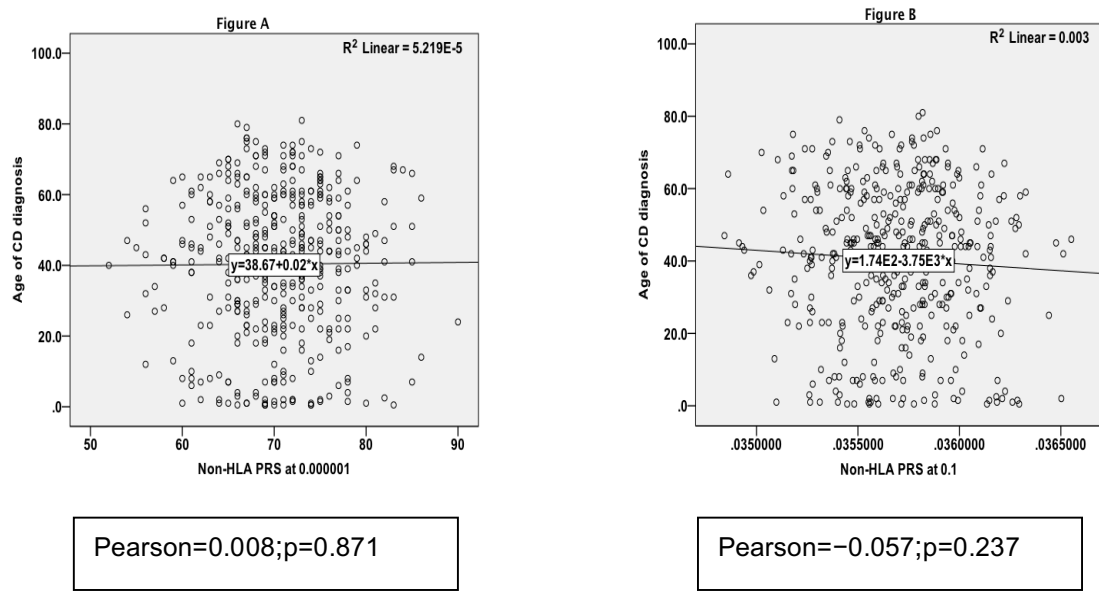
**Figure 4.3 Correlation between HLA only PRS and age of CD Diagnosis.**

Scatterplots showing correlation between age of diagnosis with CD and HLA only PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B).



**Figure 4.4 Correlation between non-HLA only PRS and age of CD diagnosis.**

Scatterplots showing correlation between age of diagnosis with CD and non-HLA only PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B).



No significant correlation was found between any PRS at any p-value threshold and age of diagnosis with CD (Figure 4.2, 4.3 and 4.4). However, we observed a tendency approaching statistical significance ( $p=0.078$ ) towards a negative correlation for the HLA & non-HLA and HLA only PRS at a threshold of  $p<0.1$ .

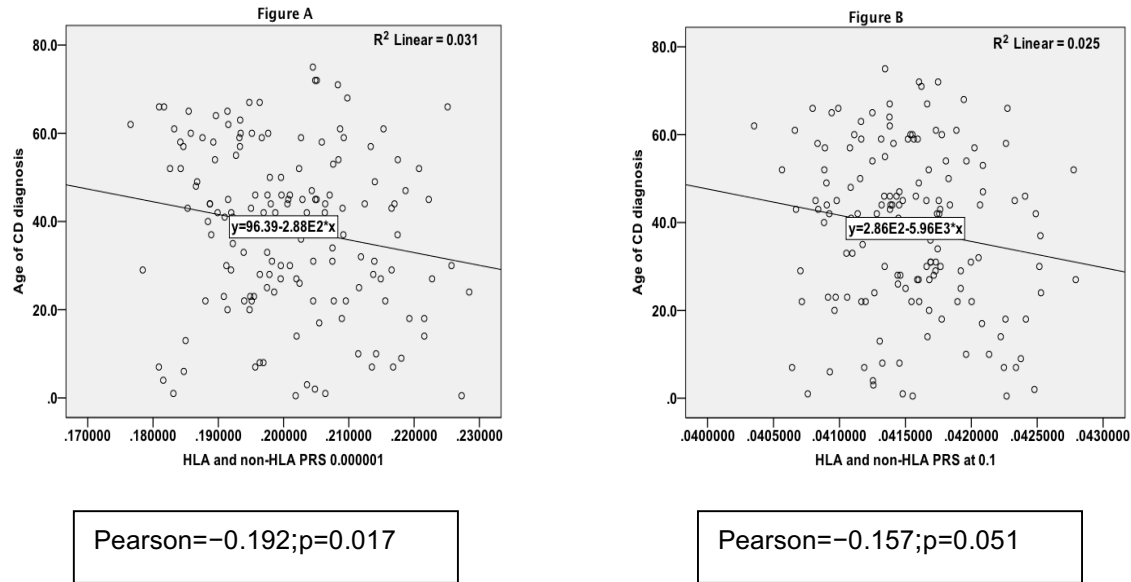
According to our findings in section 4.4.1 above, a positive CD family history seems to have a higher genetic load and thus to be more genetically determined. Therefore, in the next section we analysed the age of CD diagnosis and genetic burden, only in patients with a positive CD family history. The rationale for this approach is as previously outlined (i) this group seem to have a higher genetic burden and by using HLA & non-HLA PRS we can predict patients who were diagnosed  $<40$  and  $\geq 40$  years of age (Table 4.1 and 4.4), (ii) analysing the continuous variable 'age of diagnosis with CD' and correlating it with PRS will allow us to visualize the trend in better detail in this subcohort of patients, and (iii) there is a potentially greater awareness of the disease when other family members are already affected which is likely to translate into a decreased delay between the onset of symptoms and the disease being diagnosed.

#### **4.4.3 Correlation between age of diagnosis with CD and HLA & non-HLA, HLA only and non-HLA only PRS in CD patients with a family history of CD (n=154).**

The analysis in table 4.4 showed a trend towards an association of diagnosis  $<40$  and  $\geq 40$  (as a categorical variable) and genetic burden (HLA & non-HLA PRS) in CD patients with a positive family history. In order to visualise this trend in greater detail we analysed the age of CD diagnosis as a continuous variable correlating it with the different PRS.

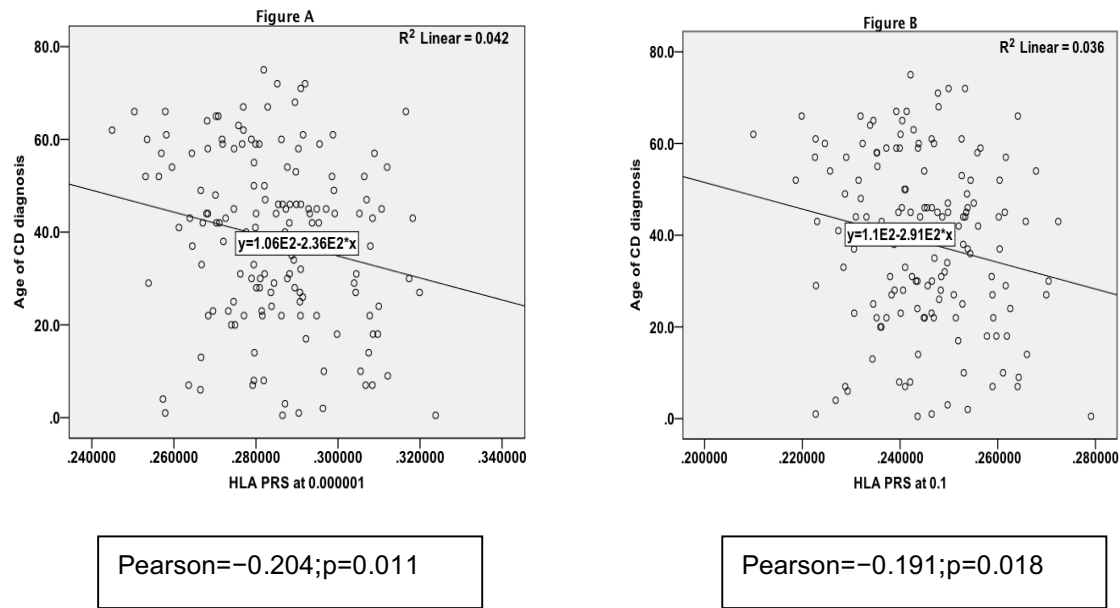
**Figure 4.5 Correlation between PRS for all markers (HLA & non-HLA) and age of CD diagnosis in patients with family history of CD.**

Scatterplots showing the correlation between age of diagnosis with CD and HLA & non-HLA PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B) in CD patients with a family history of CD (n=154).



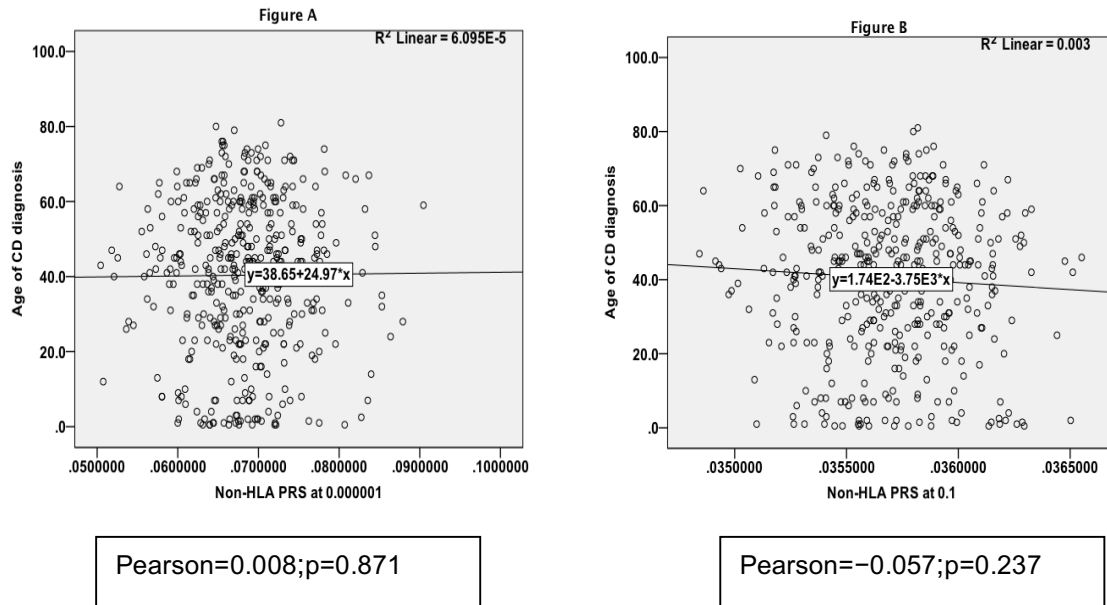
**Figure 4.6 Correlation between HLA only PRS and age of CD diagnosis in patients with family history of CD.**

Scatterplots showing the correlation between age of diagnosis with CD and HLA only PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B) in CD patients with a family history of CD (n=154).



**Figure 4.7 Correlation between non-HLA only PRS and age of CD diagnosis in patients with family history of CD.**

Scatterplots showing the correlation between age of diagnosis with CD and non-HLA only PRS at p-values,  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B) in CD patients with a family history of CD (n=154).



Therefore PRS for all markers (HLA & non-HLA) and HLA only are negatively correlated with age of diagnosis in the subcohort of patients with a family history of CD. This implies that those developing CD at a younger age carry a greater burden of genetic risk (i.e. either more susceptibility genes or an array of susceptibility genes conferring higher risk of disease). The correlation coefficient and level of statistical significance improves slightly with the HLA only score.

#### 4.4.4 Analysis of HLA & non-HLA, HLA only and non-HLA only PRS changes over time.

Given our observations of changes in CD clinical presentation described in Chapter 2 and 3, in this section we investigated potential CD genotype changes as reflected in mean PRS scores over three time periods;  $\leq 1995$ , 1996-2005 and  $\geq 2006$ .

Table 4.5 Analysis of HLA & non-HLA PRS at different p-value levels over time.

| HLA & non-HLA PRS    |                          |               |                   |                    |
|----------------------|--------------------------|---------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value              | ANOVA p-value | Beta coefficient  | Regression p-value |
|                      | <b>5×10<sup>-8</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.20824                  |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.20857                  |               | 0.012             | 0.834              |
| ≥2006 n=184          | 0.20689                  | 0.443         | - 0.054           | 0.336              |
|                      | <b>1×10<sup>-7</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.19915                  |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.19934                  |               | 0.007             | 0.903              |
| ≥2006 n=184          | 0.19817                  | 0.655         | - 0.040           | 0.476              |
|                      | <b>1×10<sup>-6</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.17629                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.17642                  |               | 0.006             | 0.922              |
| ≥2006 n=184          | 0.17545                  | 0.673         | - 0.040           | 0.483              |
|                      | <b>1×10<sup>-5</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.16345                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.16314                  |               | - 0.017           | 0.766              |
| ≥2006 n=184          | 0.16275                  | 0.750         | - 0.042           | 0.453              |
|                      | <b>1×10<sup>-4</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.14344                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.14340                  |               | - 0.003           | 0.957              |
| ≥2006 n=184          | 0.14277                  | 0.582         | - 0.052           | 0.361              |
|                      | <b>1×10<sup>-3</sup></b> |               |                   |                    |
| ≤1995 n=133          | 0.10622                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.106175                 |               | - 0.005           | 0.929              |
| ≥2006 n=184          | 0.10584                  | 0.652         | - 0.047           | 0.406              |
|                      | <b>0.01</b>              |               |                   |                    |
| ≤1995 n=133          | 0.07124                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.07124                  |               | 0.001             | 0.988              |
| ≥2006 n=184          | 0.07110                  | 0.709         | - 0.039           | 0.485              |
|                      | <b>0.05</b>              |               |                   |                    |
| ≤1995 n=133          | 0.04925                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.04928                  |               | 0.015             | 0.788              |
| ≥2006 n=184          | 0.04918                  | 0.544         | - 0.043           | 0.441              |
|                      | <b>0.1</b>               |               |                   |                    |
| ≤1995 n=133          | 0.041445                 |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.04147                  |               | 0.022             | 0.701              |
| ≥2006 n=184          | 0.041385                 | 0.354         | - 0.055           | 0.325              |
|                      | <b>0.5</b>               |               |                   |                    |
| ≤1995 n=133          | 0.02569                  |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.02570                  |               | 0.019             | 0.730              |
| ≥2006 n=184          | 0.02566                  | 0.215         | - 0.072           | 0.199              |

1.0\$= Reference group

Table 4.6 Analysis of HLA only PRS at different p-value levels over time.

| HLA only PRS         |                                      |               |                   |                    |
|----------------------|--------------------------------------|---------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value                          | ANOVA p-value | Beta coefficient  | Regression p-value |
|                      | <b><math>5 \times 10^{-8}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.29600                              |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.29603                              |               | 0.001             | 0.990              |
| ≥2006 n=184          | 0.29336                              | 0.328         | − 0.071           | 0.206              |
|                      | <b><math>1 \times 10^{-7}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.29427                              |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.29421                              |               | − 0.002           | 0.978              |
| ≥2006 n=184          | 0.29167                              | 0.348         | − 0.070           | 0.211              |
|                      | <b><math>1 \times 10^{-6}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.28277                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.28280                              |               | 0.001             | 0.989              |
| ≥2006 n=184          | 0.27995                              | 0.231         | − 0.082           | 0.147              |
|                      | <b><math>1 \times 10^{-5}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.26887                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.26884                              |               | − 0.001           | 0.989              |
| ≥2006 n=184          | 0.266275                             | 0.315         | − 0.073           | 0.193              |
|                      | <b><math>1 \times 10^{-4}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.25904                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.259135                             |               | 0.003             | 0.963              |
| ≥2006 n=184          | 0.25671                              | 0.325         | − 0.071           | 0.211              |
|                      | <b><math>1 \times 10^{-3}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.25325                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.25305                              |               | − 0.006           | 0.920              |
| ≥2006 n=184          | 0.25103                              | 0.364         | − 0.071           | 0.208              |
|                      | <b>0.01</b>                          |               |                   |                    |
| ≤1995 n=133          | 0.24481                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.24480                              |               | − 0.001           | 0.992              |
| ≥2006 n=184          | 0.24264                              | 0.345         | − 0.070           | 0.212              |
|                      | <b>0.05</b>                          |               |                   |                    |
| ≤1995 n=133          | 0.24351                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.24352                              |               | 0.000             | 0.996              |
| ≥2006 n=184          | 0.24129                              | 0.279         | − 0.077           | 0.174              |
|                      | <b>0.1</b>                           |               |                   |                    |
| ≤1995 n=133          | 0.24345                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.24335                              |               | − 0.003           | 0.955              |
| ≥2006 n=184          | 0.24104                              | 0.178         | − 0.091           | 0.107              |
|                      | <b>0.5</b>                           |               |                   |                    |
| ≤1995 n=133          | 0.23099                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.23095                              |               | − 0.002           | 0.977              |
| ≥2006 n=184          | 0.22890                              | 0.245         | − 0.081           | 0.149              |

1.0<sup>\$</sup>= Reference group



**Table 4.7 Analysis of non-HLA only PRS at different p-value levels over time.**

| Non-HLA only PRS     |                                      |               |                   |                    |
|----------------------|--------------------------------------|---------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value                          | ANOVA p-value | Beta coefficient  | Regression p-value |
|                      | <b><math>5 \times 10^{-8}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.07879                              |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.07956                              |               | 0.039             | 0.491              |
| ≥2006 n=184          | 0.08038                              | 0.282         | 0.089             | 0.114              |
|                      | <b><math>1 \times 10^{-7}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.07826                              |               | 1.0 <sup>\$</sup> | 1.0 <sup>\$</sup>  |
| 1996-2005 n=119      | 0.07861                              |               | 0.020             | 0.728              |
| ≥2006 n=184          | 0.079685                             | 0.253         | 0.088             | 0.118              |
|                      | <b><math>1 \times 10^{-6}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.06760                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.06772                              |               | 0.009             | 0.878              |
| ≥2006 n=184          | 0.06901                              | 0.092         | 0.109             | 0.053              |
|                      | <b><math>1 \times 10^{-5}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.06610                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.06591                              |               | - 0.017           | 0.767              |
| ≥2006 n=184          | 0.06701                              | 0.117         | 0.090             | 0.111              |
|                      | <b><math>1 \times 10^{-4}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.067845                             |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.06774                              |               | - 0.014           | 0.807              |
| ≥2006 n=184          | 0.06834                              | 0.265         | 0.070             | 0.213              |
|                      | <b><math>1 \times 10^{-3}</math></b> |               |                   |                    |
| ≤1995 n=133          | 0.06123                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.06115                              |               | - 0.018           | 0.744              |
| ≥2006 n=184          | 0.06151                              | 0.265         | 0.067             | 0.234              |
|                      | <b>0.01</b>                          |               |                   |                    |
| ≤1995 n=133          | 0.04881                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.04880                              |               | - 0.002           | 0.970              |
| ≥2006 n=184          | 0.04897                              | 0.606         | 0.047             | 0.405              |
|                      | <b>0.05</b>                          |               |                   |                    |
| ≤1995 n=133          | 0.039865                             |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.03990                              |               | 0.037             | 0.508              |
| ≥2006 n=184          | 0.03989                              | 0.781         | 0.031             | 0.583              |
|                      | <b>0.1</b>                           |               |                   |                    |
| ≤1995 n=133          | 0.03569                              |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.03571                              |               | 0.031             | 0.586              |
| ≥2006 n=184          | 0.035685                             | 0.807         | - 0.01            | 0.98               |
|                      | <b>0.5</b>                           |               |                   |                    |
| ≤1995 n=133          | 0.023525                             |               | 1.0 <sup>\$</sup> |                    |
| 1996-2005 n=119      | 0.02353                              |               | 0.031             | 0.580              |
| ≥2006 n=184          | 0.02352                              | 0.601         | - 0.024           | 0.668              |

1.0<sup>\$</sup>= Reference group

In Tables 4.5, 4.6, and 4.7, no significant differences were found in any of the mean PRS over the three time periods considered. However, both HLA and Non-HLA PRS, show a tendency towards a lower PRS in the period ≥2006 (Table 4.5). For the HLA PRS, this reaches an ANOVA p-value of 0.178 at a p-value of 0.1 however there are consistently lower beta values for this time period at all p-value thresholds (Table 4.6).

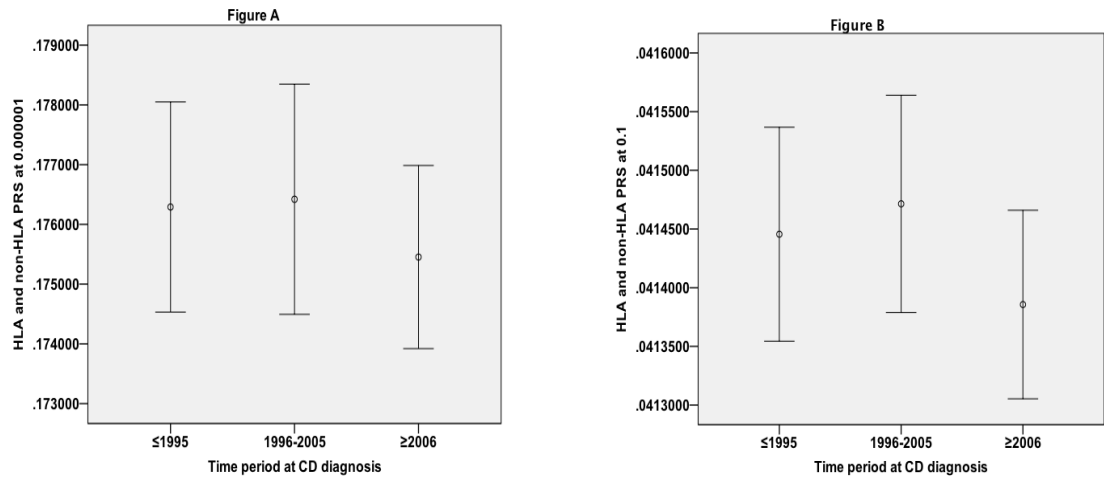
This may indicate that including more SNPs in the case of HLA PRS leads to better detection of changes in genetic burden.

The opposite trend is seen for the non-HLA PRS, where we observe a tendency to higher mean PRS values since 2006, which reaches a p-value of 0.092 for PRS p-value cut-off  $1 \times 10^{-6}$ . However, the previous non-HLA tendency is reversed at less significant p-values (0.05, 0.1 and 0.5; Table 4.7 and Figure 4.10). Therefore, including more SNPs with lower evidence of disease association in the non-HLA score leads to a reversal in the trend. This could be construed as adding more ‘noise’ to the data and potentially drowning out the signal. The beta coefficient shows that in the majority of the cases for all PRS, the period having the biggest impact in the linear regression is the last period  $\geq 2006$  (Tables 4.5, 4.6 and 4.7); thus the greatest change in allele frequencies seems to have occurred in this period.

The following error bar graphs are provided as a visual aid to observe these trends in HLA & non-HLA, HLA only and non-HLA only PRS (Tables 4.5, 4.6 and 4.7). We show two different, representative, p-value levels over the three time periods considered.

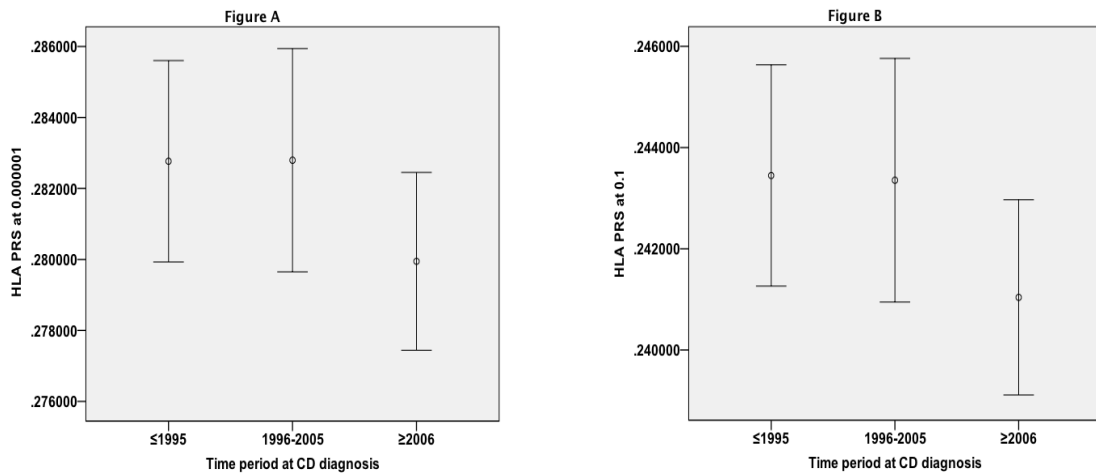
**Figure 4.8 Mean PRS for All SNPs According to Time Period.**

Graphs showing mean HLA & non-HLA PRS of according to time period of CD diagnosis, with standard deviation error bars at representative PRS p-values of  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B).



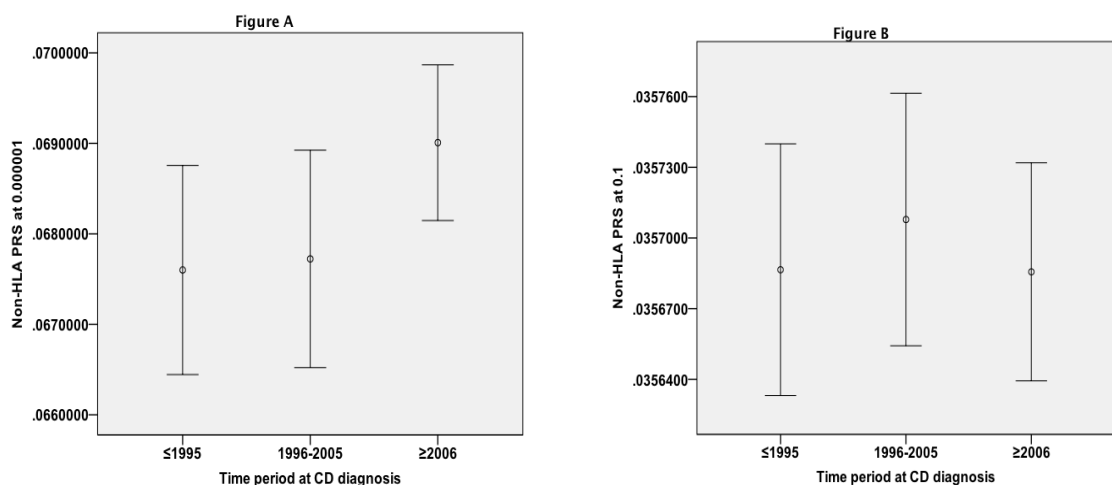
**Figure 4.9 Mean HLA only PRS According to Time Period.**

Graphs of mean HLA only PRS of CD sample with standard deviation error bars at PRS p-values  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B) are shown.



**Figure 4.10 Mean non-HLA PRS According to Time Period.**

Graphs of mean non-HLA PRS of CD sample with standard deviation error bars at PRS  $1 \times 10^{-6}$  (Figure A) and 0.1 (Figure B) are shown.



In the figures shown above, it is apparent that the mean PRS of the first two time periods are closer to each other compared to the latter period and this pattern is more or less consistent across all comparisons. It could be said that for the HLA & Non-HLA and HLA only PRS, the first two periods nearly form a plateau, which plummets in the latter period (Figures 4.8 & 4.9). This is not so clear for the non-HLA PRS, where more variability is observed depending on the PRS p-value level used, however at a non-HLA p-value of  $p=1 \times 10^{-6}$  which provided the lowest ANOVA p-value (Table 4.7), we observe a plateau in the first two time periods followed by a pronounced increase in the

latter period (Figure 4.10A). Seeking to investigate this further, based on the observation that the first two time periods have similar PRS while the latter period appears different (Figures 4.8, 4.9 and 4.10), we combined the earlier periods and carried out an analysis over time with a binary time variable  $\leq 2005$  and  $\geq 2006$  to see if these trends are sustained or become more statistically significant. Indeed, by doing so we observed a trend in reduction of p-values for all PRS, however only non-HLA PRS at cut-off p-value  $1 \times 10^{-6}$  and  $1 \times 10^{-5}$  produce a statistically significant result ( $p=0.029$  and  $p=0.04$  respectively) (Table 4.8 and supplementary Tables A and B, appendix XII).

We conclude that the preceding analyses point to a general trend, whereby the genetic load in CD has decreased over the time frame of this study. In particular, the genetic load contributed by the HLA appears to have reduced, while on the contrary, that contributed by non-HLA genes seems to have increased.

**Table 4.8 Analysis of non-HLA only PRS at different p-value levels over time using binary time variable  $\leq 2005$  and  $\geq 2006$ .**

| Non-HLA PRS          |                                      |                |                   |                    |
|----------------------|--------------------------------------|----------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value                          | t-test p-value | Beta coefficient  | Regression p-value |
|                      | <b><math>5 \times 10^{-8}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.07915                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.08038                              | 0.151          | 0.069             | 0.151              |
|                      | <b><math>1 \times 10^{-7}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.07842                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.079685                             | 0.105          | 0.078             | 0.105              |
|                      | <b><math>1 \times 10^{-6}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.06766                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.06901                              | <b>0.029</b>   | 0.104             | 0.029              |
|                      | <b><math>1 \times 10^{-5}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.06601                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.06701                              | <b>0.040</b>   | 0.098             | 0.040              |
|                      | <b><math>1 \times 10^{-4}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.067795                             |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.06834                              | 0.107          | 0.077             | 0.107              |
|                      | <b><math>1 \times 10^{-3}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.06119                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.06151                              | 0.110          | 0.077             | 0.110              |
|                      | <b>0.01</b>                          |                |                   |                    |
| $\leq 2005$ n=252    | 0.048805                             |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.04897                              | 0.317          | 0.048             | 0.317              |
|                      | <b>0.05</b>                          |                |                   |                    |
| $\leq 2005$ n=252    | 0.03988                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.03989                              | 0.812          | 0.011             | 0.812              |
|                      | <b>0.1</b>                           |                |                   |                    |
| $\leq 2005$ n=252    | 0.03570                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.035685                             | 0.715          | - 0.018           | 0.715              |
|                      | <b>0.5</b>                           |                |                   |                    |
| $\leq 2005$ n=252    | 0.02353                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.02352                              | 0.399          | - 0.041           | 0.399              |

1.0<sup>\$</sup>= Reference group

Statistical significance is reached for non-HLA PRS at a p-level of  $1 \times 10^{-6}$  and  $1 \times 10^{-5}$  (Table 4.8) (bold black).

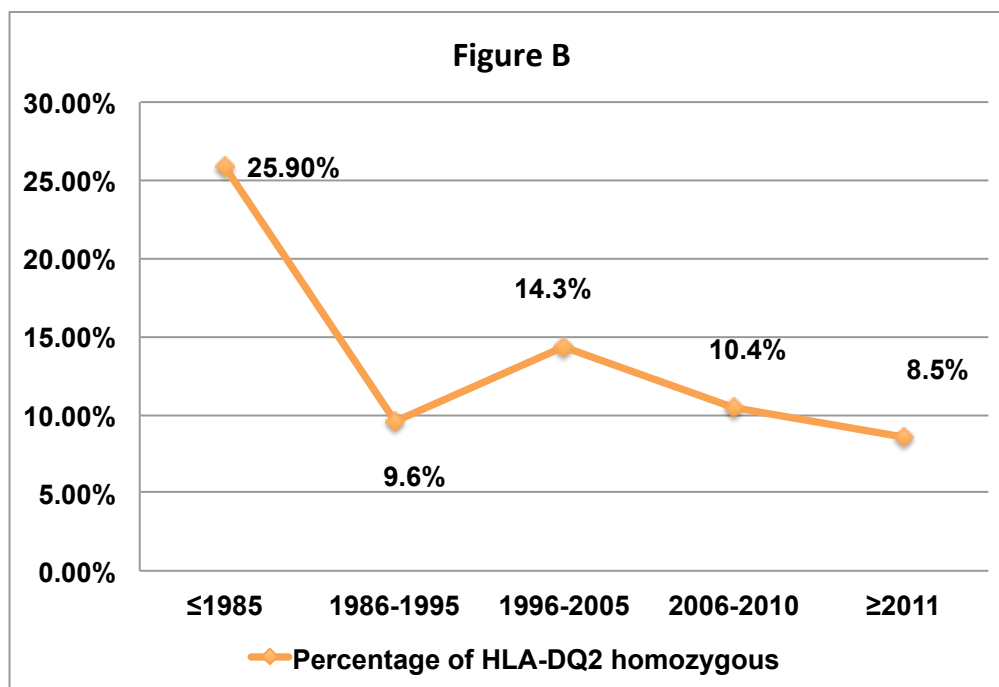
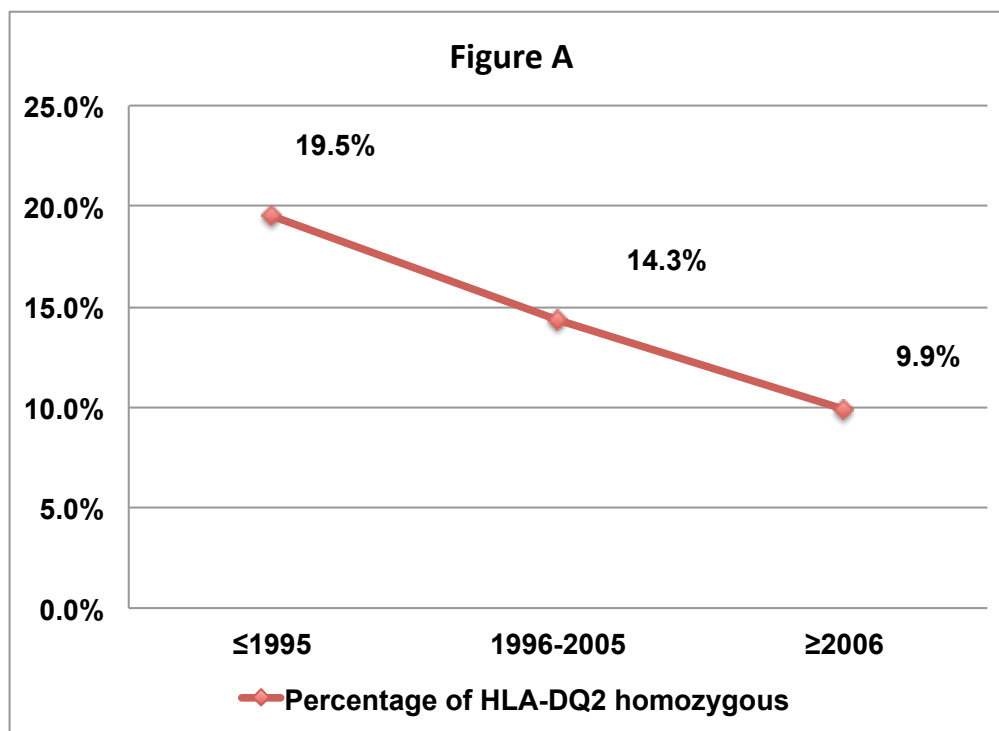
#### 4.4.5 Investigation of whether allele frequencies contributing to PRS change over time.

PRS are composite scores which calculate the overall genetic burden of an individual. They are composed of alleles of disease susceptibility loci weighted according to their attributable risk of disease. We have seen indicators that the PRS have changed over time. In this section, we wished thus to (i) explore potential changes in HLA-DQ2 homozygosity, and (ii) identify which other susceptibility locus alleles have changed

frequency over time that could explain the differences in PRS observed in section 4.4.3 and 4.4.4.

Identification of which susceptibility genes were changing frequency could help us understand the potential biological mechanisms underlying the clinical changes reported in Chapter 1 and 2.

**Figure 4.11 Percentage of HLA-DQ2 homozygous patients over three time periods (Figure A) and over five time periods (Figure B).**



The percentage of CD patients homozygous for HLA-DQ2.5 decreases significantly over the three time periods ( $p=0.051$ ). The SNP rs2187668-A was used as a proxy for the DQ2.5 haplotype as explained in the methods section. Regression analysis was carried out using the three time periods as our independent variable, and being homozygous for HLA-DQ2.5 as the outcome of interest. This analysis resulted in a borderline significant result indicating that it was more than two fold more likely to be homozygous for HLA-DQ2.5  $<1995$  compared to  $\geq 2006$  (OR 2.2; 95%CI 1.2-4.2) ( $p=0.056$ ).

Interestingly, we also carried out regression analysis using our original five time period variable ( $\leq 1985$ , 1986-1995, 1996-2005, 2006-2010 and  $\geq 2011$ ) and this resulted in a significant result indicating that in the period  $\leq 1985$  it was nearly 4 fold more likely to be homozygous for HLA-DQ2 (OR 3.8; 1.2-11.7) ( $p=0.016$ ) (data not shown).



Figure 4.12 Scatterplot representing changes in gene variants frequencies between the first time period  $\leq 1995$  and the last time period  $\geq 2006$ .

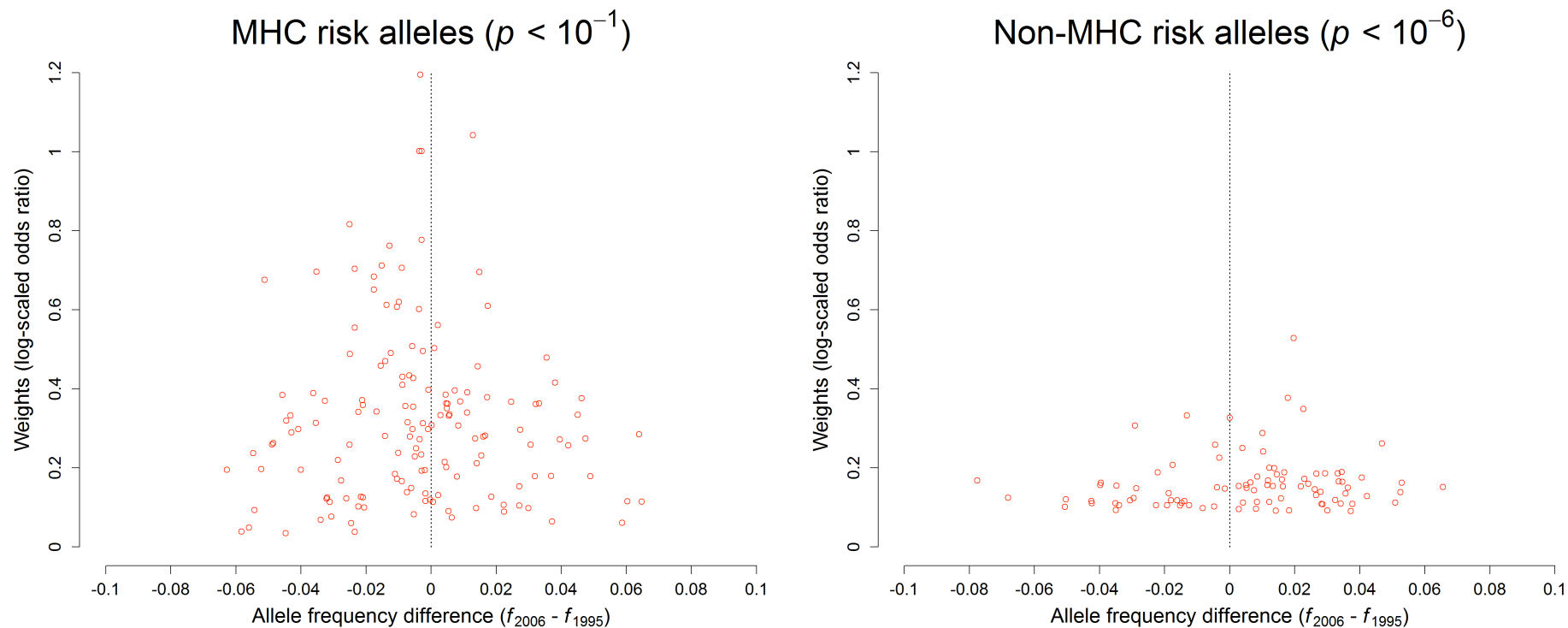
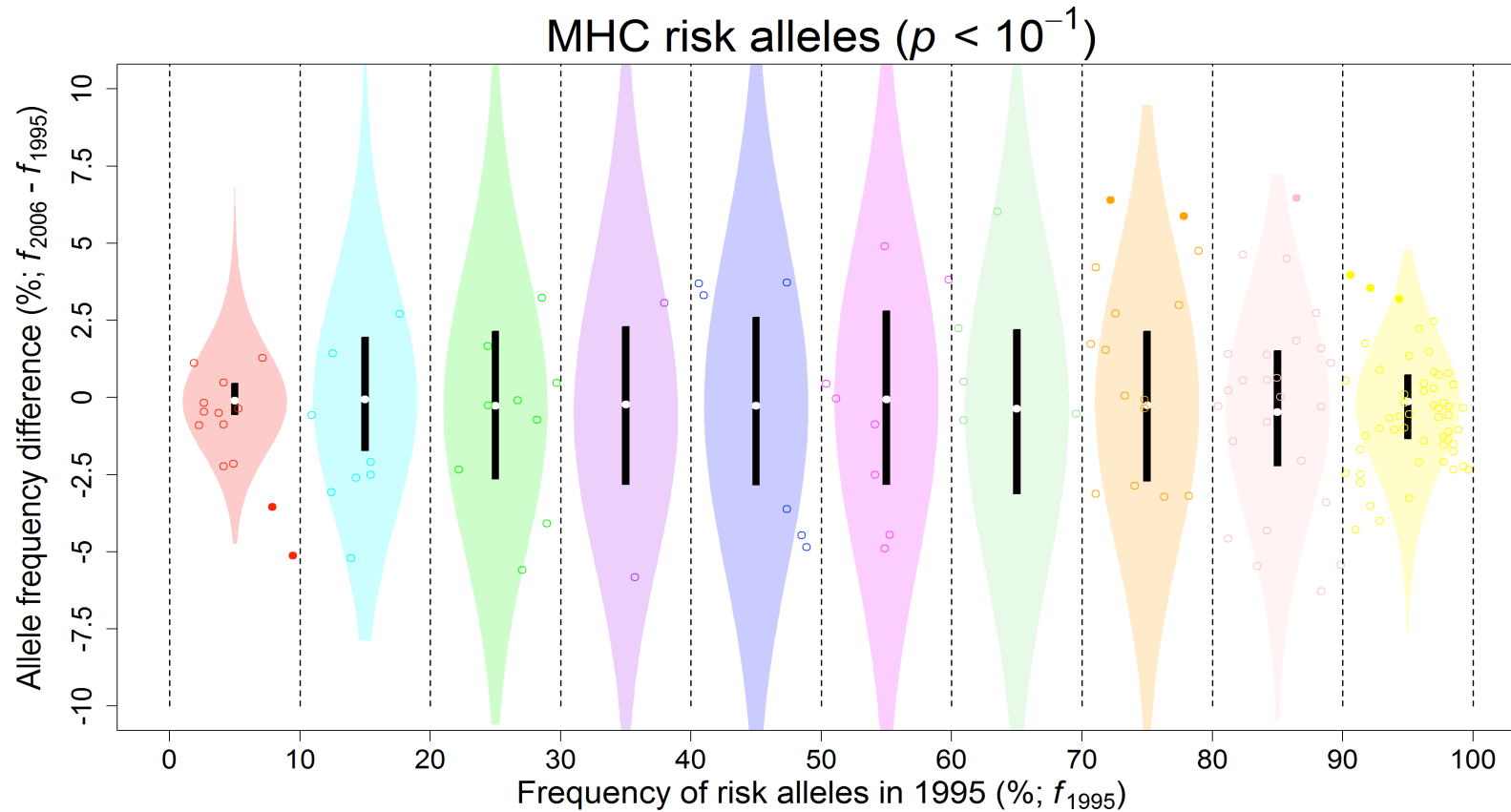


Figure 4.12 shows the distribution of alleles which form the PRS for the MHC only PRS at p-value  $< 10^{-1}$  and non-MHC alleles at  $p < 10^{-6}$ . Negative values indicate that the allele frequency has fallen from 1995 to 2006, while positive values indicate an increase in frequency. The Y axis shows the PRS weight of each allele as a log-transformed odds ratio.

**Figure 4.13 HLA alleles simulation distribution plot based on 1000 simulation data.**

The following figure represents a study of changes in HLA allele variance between the first time period  $\leq 1995$  and the later time period  $\geq 2006$  to identify any significant tendencies. Filled dots represent statistically significant changes (falling outside the 95% CI) in major histocompatibility complex (MHC) allele frequencies. The black bar represents the range between the 25th and 75th percentiles. The white dot in the black bar represents the median value.



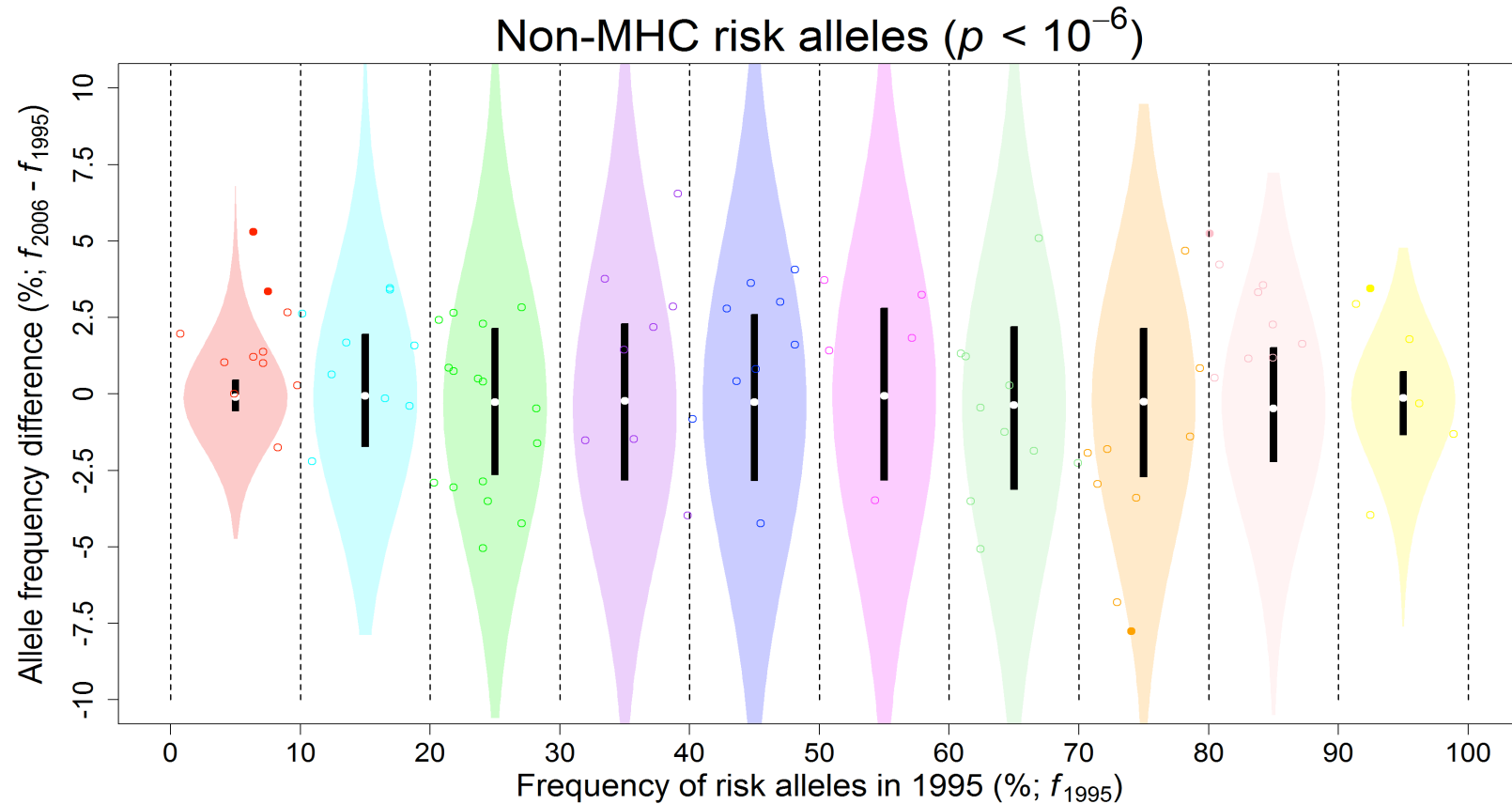
In Figure 4.13 above it can be observed a significant change in some HLA alleles as follows; (i) 0% to 10% interval rs241420 and rs116636762 (decrease from 1995 to 2006), (ii) 70% to 80% interval rs477515 and rs9380351 (increase from 1995 to 2006), (iii) 80% to 90% interval rs6916301 (increase from 1995 to 2006) (iv) 90% to 100% interval, rs6917517, rs114992171 and rs411337 (increase from 1995 to 2006).

These markers are listed in Table 4.9. In an effort to understand whether these SNPs could be having a functional effect on disease pathogenesis, we have included evidence from a number of publicly available data sources on the characteristics of these SNP markers from a functional standpoint. This includes data on whether these SNPs (which mostly do not fall within the coding regions of genes) are found in genomic regions with epigenetic marks indicative of permissive or accessible chromatin; and/or whether these sites represent cognate sites for, or are known to be bound by, DNA binding proteins such as transcription factors or RNA polymerases. Any evidence that the markers are associated with altered regulation of known genes (i.e. they act as expression quantitative trait loci, eQTLs) is also included.

Since many, but not all, of the markers in these analyses show high levels of linkage disequilibrium (LD) with other genetic variants in their local region, some of these are included where they may represent a good candidate as a causatively associated SNP.

**Figure 4.14 Non-HLA alleles simulation distribution plot based on 1000 simulation data.**

The following figure represents a study of changes in non-HLA genes allele variance between the first time period  $\leq 1995$  and the later time period  $\geq 2006$  to identify any significant tendencies. Filled dots represent statistically significant changes (falling outside the 95% CI) in non-MHC allele frequencies. The black bar represents the range between the 25<sup>th</sup> and 75<sup>th</sup> percentiles. The white dot in the black bar represents the median value.



In Figure 4.14 above a significant change in several non-HLA alleles can be seen. These are as follows; (i) 0% to 10% interval rs77589207 and rs6806528 (increase from 1995 to 2006), (ii) 70% to 80% interval rs1353248 (decrease from 1995 to 2006), (iii) 80% to 90% interval rs11801183 (increase from 1995 to 2006), (iv) 90% to 100% interval rs76515784 (increase from 1995 to 2006).

As with the HLA markers, the non-HLA markers above are annotated and tabulated in Figure 4.10.

**Table 4.9 HLA markers showing significant changes in frequency between 1995-2006.**

| Marker              | Freq Diff | PRS weight | r2   | Position   | Regulome | Ref Allele | Alt Allele | Location                      | Expression     | Reg motif                    | Proteins bound | eQTL                                                          | Alternate Allele Freq | Conserved |
|---------------------|-----------|------------|------|------------|----------|------------|------------|-------------------------------|----------------|------------------------------|----------------|---------------------------------------------------------------|-----------------------|-----------|
| <b>rs6916301</b>    | 0.065     | 0.1145     |      | 27,074,479 | 5        | C          | T          | 35kb 3' of U2                 | -              | 5 altered motifs             | -              | HIST1H1C, BTN2A1, PRSS16                                      | 0.135                 |           |
| <b>rs9357031</b>    |           |            | 1    | 27,078,975 | 2b       | C          | T          | 31kb 3' of U2                 | 21 tissues     | Eomes,Pax-4                  | 7              | HIST1H1C, BTN2A1,                                             | 0.114                 |           |
| <b>rs477515</b>     | 0.064     | 0.2853     |      | 32,601,914 | 6        | G          | A          | 12kb 5' of HLA-DRB1           | BLD            | HDAC2                        | -              | DQA1, DQB1, DQB1-AS1, DQA2, DQB2, DRB1, DRB6, 250 other       | 0.278                 |           |
| <b>rs556025</b>     |           |            | 0.82 | 32,603,103 | 6        | C          | T          | 13kb 5' of HLA-DRB1           | BLD            | BCL,Ets,Sin3Ak-20            | -              | DQA1, DQB1, DQB1-AS1, DQA2, DQB2, DRB1, DRB6, 180 Other       | 0.325                 |           |
| <b>rs9380351</b>    | 0.059     | 0.0608     |      | 33,161,969 | 2b       | G          | A          | 711bp 3' of COL11A2           | BLD, GI, other | GATA, Elf3, PU.1, P300, AP-1 | -              | B3GALT4, HLA-DOA, HLA-DPB1, HSD17B8, RING1                    | 0.222                 |           |
| <b>rs6917517</b>    | 0.040     | 0.2714     |      | 31,114,584 | 4        | C          | T          | 165bp 5' of PSORS1C1          | 19 tissues     | Irf,NF-I                     | 7              | 70 hits VARS1, HCG27, VARS2, HLA-C, HLA-B, PSORS1C1, PSORS1C2 | 0.094                 |           |
| <b>rs78000963 §</b> | 0.035     | 0.4797     |      | 30,535,025 | 6        | G          | A          | 6.4kb 3' of GNL1              | BLD, GI        | STAT, ELF3, CPD, FOXA        | -              | VARS2                                                         | 0.079                 |           |
| <b>rs411337</b>     | 0.032     | 0.1788     |      | 32,109,603 | 5        | C          | T          | 228bp 5' of TNXB              | BLD, GI, other | 6 altered motifs             | -              | SKIV2L, NELFE                                                 | 0.056                 |           |
| <b>rs169494</b>     |           |            | 1    | 32,130,099 | 3a       | G          | A          | 5' UTR FKBPL                  | BLD, GI, other | CDP,CTCF,HNF1                | 16             | SKIV2L, NELFE                                                 | 0.081                 |           |
| <b>rs241420</b>     | -0.051    | 0.6760     |      | 32,855,616 | 3a       | C          | G,T        | PSMB9 within gene             | BLD, GI, other | -                            | -              | SKIV2L                                                        | 0.043                 |           |
| <b>rs241419</b>     |           |            | 1    | 32,856,171 | 5        | G          | A,C        | PSMB9 missense                | BLD, GI, other | -                            | -              | SKIV2L                                                        | 0.029                 | Yes       |
| <b>rs2534826 §</b>  | -0.035    | 0.3141     |      | 30,527,640 | 5        | G          | A          | 13.7/8kb 3' of GNL1/LINC02569 | BLD, GI, other | POU3                         |                | HCG18, HLA-L, PSORS1C2, HLA-A, TRIM26                         | 0.079                 |           |
| <b>rs28360036</b>   |           |            | 1    | 30,571,044 | 2b       | C          | T          | 331bp 5' of ABCF1             | BLD, GI, other | 16 altered motifs            | INI1, POL2     | HCG17, HLA-L, PSORS1C2, HLA-A, TRIM26, TCF19                  | 0.043                 |           |

BLD = Blood; GI= Gastrointestinal

§ rs116636762 replaced by rs2534826

§ rs114992171 replaced by rs78000963

PSMB9 = proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2) [Source:HGNC Symbol;Acc:9546]

GNL1= guanine nucleotide binding protein-like 1 [Source:HGNC Symbol;Acc:4413]

ABCF1= ATP-binding cassette, sub-family F (GCN20), member 1 [Source:HGNC Symbol;Acc:70]

FKBPL=FK506 binding protein like [Source:HGNC Symbol;Acc:13949]

**Table 4.10 Non-HLA markers showing significant changes in frequency between 1995-2006.**

| Marker                     | Freq Diff | PRS Weight | r2   | Chromosome | Position    | Regulome | Ref Allele | Alt Allele | Location               | Expression     | Reg motif         | Proteins bound    | eQTL (sig) | Alternate Allele Freq | Conserved | Disease Interval or SNP                      |
|----------------------------|-----------|------------|------|------------|-------------|----------|------------|------------|------------------------|----------------|-------------------|-------------------|------------|-----------------------|-----------|----------------------------------------------|
| <a href="#">rs77589207</a> | 0.053     | 0.1621     |      | chr2       | 181,231,598 | 6        | A          | G          | AC104820.2 within gene | BLD, other     | 7 altered motifs  | -                 | -          | 0.117                 | -         | CD, AS                                       |
| <a href="#">rs11801183</a> | 0.053     | 0.1383     |      | chr1       | 172,893,844 | 4        | C          | T          | 3' 147374 TNFSF18      | BLD, GI, Other | AP1, GATA, PPAR3  | -                 | FASLG      | 0.147                 | -         | CD, CRO                                      |
| <a href="#">rs12142280</a> |           |            | 0.99 | chr1       | 172,895,512 | 4        | T          | A          | 3' 145706 TNFSF18      | BLD, GI, Other | Foxp1,RREB-1      | -                 | -          | 0.182                 | -         | CD (OR 0.87 8e-9)                            |
| <a href="#">rs6684112</a>  |           |            | 0.87 | chr1       | 172,897,756 | 2b       | A          | G          | 3' 143462 TNFSF18      | BLD, GI, Other | 4 altered motifs  | POL2, CFOS, GATA2 | -          | 0.170                 | -         |                                              |
| <a href="#">rs76515784</a> | 0.034     | 0.1897     |      | chr6       | 137,680,575 | 6        | C          | T          | 5' 186180 OLIG3        | LNG, BLD, CRVX | Foxa,GATA,SRF     | -                 | -          | 0.041                 | -         | CD, UC, MS, PBC, RA, SJO, T1DM               |
| <a href="#">rs6806528</a>  | 0.034     | 0.1655     |      | chr3       | 69,203,748  | 4        | C          | T          | FRMD4 Within gene      | BLD, GI, other | TCF12             | -                 | -          | 0.109                 | -         | CD (2e-7)                                    |
| <a href="#">rs6776027</a>  |           |            | 0.99 | chr3       | 69,204,304  | 2b       | T          | C          | FRMD4 Within gene      | BLD, other     | Pou1f1,RXRA       | IRF4,NFKB         | -          | 0.105                 | -         |                                              |
| <a href="#">rs6784841</a>  |           |            | 0.99 | chr3       | 69,204,481  | 2a       | G          | A          | FRMD4 Within gene      | 10 tissues     | 22 altered motifs | 8 proteins        | -          | 0.106                 | Yes       |                                              |
| <a href="#">rs1353248</a>  | -0.078    | 0.1679     |      | chr3       | 159,905,770 | 7        | C          | T          | SCHIP1 (8.4kb 3')      | BLD, Skin, BRN | CEBPB,HLF,Hsf     | -                 | -          | 0.337                 | -         | CD (OR 0.88 1e-10)<br>MS, PBC, SJO, SSC, SLE |

BLD = Blood; GI= Gastrointestinal; LNG= Lung; CRVX= Cervix; BRN= Brain

CRO= Crohn's; UC= Ulcerative colitis; MS= Multiple Sclerosis; PBC= Primary biliary cirrhosis; SLE=Systemic Lupus Erythematosus; RA=rheumatoid arthritis; SJO=Sjogren's syndrome; SSC=systemic sclerosis

Most of the MHC SNPs shown in Table 4.9 show extensive evidence of potential functionality. The majority are associated with epigenetic markers indicative of accessible chromatin or histone modifications which are associated with poised or active promoters or enhancer regions ('Expression'). Where these specifically incorporate blood tissues (BLD) or tissues associated with the gastrointestinal (GI) tract this is shown. Many appear to reside in active domains in a wide variety of tissues. Most of the sites either overlap regulatory motifs, which represent binding sites for DNA binding proteins, or are known to be bound by proteins in specific tissues as a result of chromatin immunoprecipitation assays.

Finally, it is notable that all are associated with variable expression of linked genes (eQTLs). rs477515 shows one of the highest frequency changes over the period 1995-2006 (increasing 6.4% to 78.8% frequency for the risk allele in 2006) and has been associated with the regulation of over 250 genes in a study of gene expression human intestinal tissue (Kabakchiev and Silverberg, 2013). Amongst these are several HLA class II genes including HLA-DQA1, -DQB1 and DRB1. rs6916301 also increases frequency dramatically to near fixation in the population by 2006 (up 6.5% to 92.9% frequency for the risk allele). The risk allele at this locus is associated with decreased expression of BTN2A1, which is a member of the butyrophilin gene family. Butyrophilins are believed to be co-stimulatory molecules involved in the regulation of CD4+ T cell function among others, and show similarities to the B7 family of co-stimulators (including CD80 and CD86). Looking at the extended table of genes showing highest derangement of gene frequency (supplementary Table 5, appendix XIII), another member of the butyrophilin family, BTN3A2, features as regulated by several markers showing nominal changes in allele frequency.

Another notable gene which is associated with two LD independent markers (rs411337 +3.2% and rs241420 -5.1%) experiencing significant changes in frequency is SKIV2L, which is an RNA helicase with antiviral effects. Both of these changes in allele frequency are associated with reduced SKIV2L expression (data from GTEx Portal, [gtexportal.org](http://gtexportal.org)). The eQTL effect sizes associated with rs411337 are shown in Figure 4.15, and it can be seen that this site is associated which leads to substantial and significant changes in SKIV2L gene expression in almost all tissues surveyed.

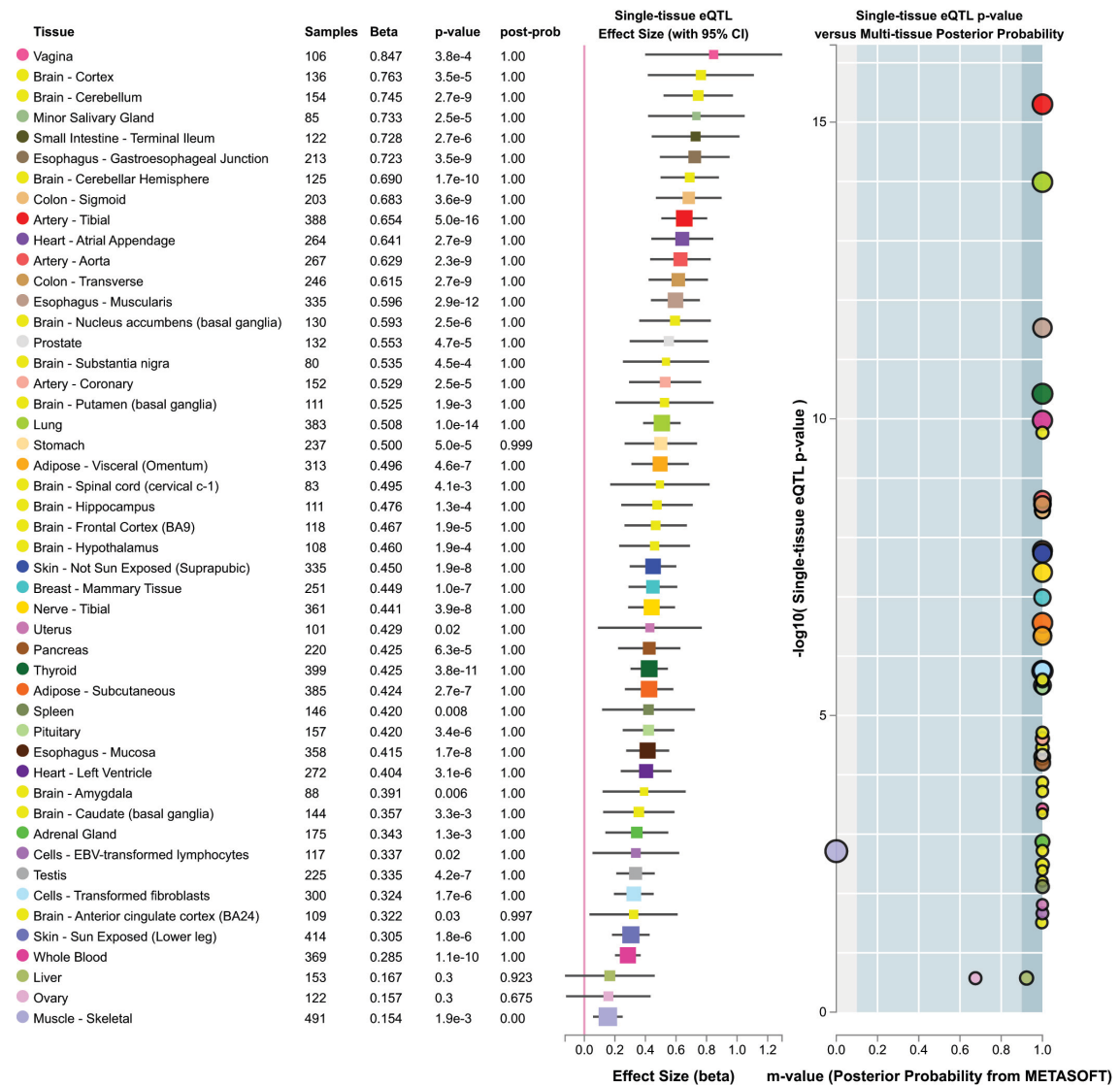
Table 4.10 lists the five non-HLA markers which show significant changes in allele frequencies between 1995 and 2006. Again all are situated in genomic regions showing evidence of having a regulatory function in blood and GI tissues predominantly. All markers are located in known CD genomic susceptibility regions. rs11801183 (+5.3%) is the only known eQTL in this group. It is very significantly



associated with Fas ligand (FASLG) expression in whole blood (Westra et al., 2013) in a region also encoding TNFSF18 and TNFSF4. FASLG is important in regulating T cell survival and is used by cytotoxic T cells to kill target cells, while TNFSF18 acts to regulate survival of T cells in the periphery.

Of course, it should be noted that these are aggregated data, and regulatory motifs, while present in the DNA, may sometimes be active only in specific tissues or cell types while other sites are active across a range of tissues. Evidence of DNA binding (derived from the Haploreg 4.1 database) does not specify the cell type. Because of the extensive LD particularly in the MHC/HLA region, it may not be possible to isolate the SNP responsible for these changes by genetic means alone. However these observations do provide compelling evidence that these markers are located in sites with the potential to be active in some tissues, and thus possibly in tissues important in CD. Finally all are located in regions genetically associated with CD supporting the view that changes in their frequencies may be causally associated with disease expression rather than being random observations.

**Figure 4.15 Changes in gene expression of SKIV2L associated eQTL locus rs411337 in various tissues.**



## 4.5 Discussion

In our cohort of 436 genotyped CD patients we found increased genetic burden, considering HLA & non-HLA and HLA only PRS, in the subcohort of patients who were found to have a positive CD family history. Other prominent findings are that PRS can be used to predict age of diagnosis in CD patients with a family history, and a significant decrease in HLA-DQ 2.5 homozygosity over time. We also observe significant changes in allele frequency of several SNPs throughout the genome during this time period. This higher genetic load in these patients was not linked to a more severe presentation measured as a greater malabsorptive disease at presentation or earlier age of diagnosis (section 4.4.1). To the extent of our knowledge this is the first study to report this finding in CD. However, interestingly enough, in a similar study in IBD patients, a higher genetic load was more common in ulcerative colitis patients who had a first degree relative affected by the condition. Similarly to our finding, they did not find that this increased genetic burden in IBD patients with a positive family history was linked to disease severity (Ananthakrishnan et al., 2014).

### 4.5.1 Genetic burden and clinical phenotype

The larger PRS in our individuals with CD relatives suggests that patients with a family history of CD have a more genetically driven disease. This is mainly driven by a larger HLA PRS. The non-HLA PRS in this group are also nominally larger, however the differences are not significant (Table 4.3). However, it is interesting how the sensitivity to detect differences between both groups is increased when considering HLA & non-HLA PRS, with more significant p-values occurring at all PRS p-value levels (Table 4.1). Therefore, this may indicate that patients with a history of CD in relatives are more predisposed to the condition both by a higher HLA and non-HLA gene load. However, we cannot detect significant differences in the latter due to our small sample size.

As far as we are aware few studies have analysed the genetic predictors for the age of CD diagnosis. In 2009 a study by Gudjonsdottir *et al.* found no association between these variables (Gudjonsdottir et al., 2009). However, this study took place before the large CD GWAS studies were published and therefore did not account for all the newly discovered non-HLA loci. More recently, in paediatric CD, homozygosity for HLA-DQ2 was associated to earlier disease onset (Liu et al., 2014). Furthermore, homozygotes for the HLA haplotype DR3-DQ2 (DQ2.5) were found to be nearly 6 fold more likely to develop CD than individuals with the lowest risk HLA haplotypes (HLA-DR4-DQ8).

In our study, we also see an effect of genetic burden predicting age of diagnosis. We show that the percentage of HLA-DQ2.5 homozygotes ( $n=61$ ) decreases significantly over the time period studied, with a large proportion of the homozygotes concentrating in the periods  $\leq 1985$  and  $\leq 1995$  ( $n=21$  &  $n=26$ ;  $p=0.056$  &  $p=0.016$  respectively) (Figure 4.11). The median age of the patients in these two time periods is lower than in later periods with a high proportion in the period  $\leq 1985$  diagnosed in infancy (median age of diagnosis 7 years). However, patients diagnosed in adulthood  $\leq 1985$  also have a higher tendency to be DQ2.5 homozygotes than patients diagnosed later (median age of diagnosis  $\leq 1985 = 34$  years; bivariate analysis  $p=0.056$ ). This seems to confirm findings in other studies that HLA-DQ2.5 homozygosity is linked to earlier disease initiation.

PRS specifically have not been used to explore genetic load and age of diagnosis in CD. However, it has been used effectively to characterise clinical subphenotypes in other diseases. For example, in Crohn's disease, a higher genetic burden has recently been associated with earlier diagnosis, and interestingly, this effect was stronger among non-smokers suggesting the interplay of genes and environmental factors on disease initiation (Ananthakrishnan et al., 2014, Cleynen et al., 2016).

Therefore, we explored the use of PRS to show any potential associations between the age of diagnosis and genetic burden. There was no association when we considered the entire patient sample (Tables 4.1, 4.2 and 4.3 & section 4.4.2). However, when we analysed the subcohort of patients with a family history of CD, both the HLA & non-HLA, and the HLA only PRS could effectively predict age at diagnosis  $<40$  or  $\geq 40$  years (Table 4.4); both PRS show a weak negative correlation which is statistically significant (section 4.4.3). This indicates that, similar to IBD, patients with a higher genetic load are diagnosed at an earlier age than patients with lower genetic predisposition.

Potential explanations for this finding could include; i) that there are more homozygotes for the HLA-DQ2.5 haplotype in this subcohort, which would correlate with the finding of a higher HLA PRS (Table 4.2) and earlier age of diagnosis (Figure 4.11); ii) higher awareness of the condition in families with more than one case may lead to a reduced delay between the onset of symptoms and CD diagnosis (Fuchs et al., 2014).

The percentage of HLA-DQ2 homozygotes is similar between individuals with and without family history (15.0% and 13.5% respectively) ( $p=0.67$ ). However, we have shown that HLA PRS are different between the family history groups implying that genetic elements other than HLA-DQ2 are contributing significantly to the risk score. The hypothesis of higher awareness may also be important in our sample, and thus it

could mean that, if delayed diagnosis was removed from the equation, HLA & non-HLA PRS would be good predictors of 'age of diagnosis with CD' for every individual in our sample.

### 4.5.2 Genetic burden over time

In section 4.4.4 we observed a tendency towards lower overall PRS in the latter time period (Figure 4.8). This seems to be driven by a fall in HLA PRS (Figure 4.9). Interestingly, the non-HLA PRS tends to increase at this time (Figure 4.10). It is intriguing that this observed tendency in lower PRS coincides with a change in the clinical phenotype of the condition as reported in Chapter 2 (Table 2.8). In order to better understand the biological mechanisms driving these changes in the PRS, we thus attempted to identify genetic markers that underwent changes in allele frequency between the time periods under consideration, i.e. from 1995 to 2006 (Figure 4.12). We modelled this using genetic analyses mimicking changes in gene frequency from one generation to the next (see section 4.3.4).

Under the null hypothesis, we would expect random oscillation of gene frequencies, however this analysis detected changes in gene frequency greater than would be expected by chance using a permuted empirical distribution. We identified eight HLA and five non-HLA genetic markers that had either increased or decreased significantly over this time period. All of these markers are located in regions that are genetically associated with CD. We infer that in some part, the phenotypic changes we have observed in clinical presentation of CD over the years are associated with changes of expression of genes in the vicinity of these markers.

We have highlighted in the results section these markers, and the associated evidence of potential functional effects by which these SNPs may contribute to CD pathogenesis (Table 4.9 & 4.10). However, our analysis is impeded by small sample size and these observations are at best a limited part of the story. Furthermore, it is important to consider that by generating PRS, we have removed a lot of information from the data set by eliminating all markers linked to the index marker at an  $r^2 > 0.1$ . This represents the standard approach but is very severe in terms of pruning and thus a lot of interesting and potentially functional genetic markers are removed from consideration. This may be particularly important in the HLA region as for technical reasons the HLA PRS may be underestimating the actual HLA risk. In any event, we are working with a very reduced set of genetic markers.

Nonetheless, in the HLA/MHC we observe 6 of our markers clustering in a 2.6Mb region between positions 30.5 and 33.1Mb, in the core of the MHC, which includes the

HLA class II, class III and part of the class I regions. The other HLA/MHC association is near position 27.1Mb near the histone gene cluster and about 500Kb from the butyrophilin gene cluster (26.6Mb). The locus showing the largest PRS weight (a log transformed measure of the OR) 0.6760, lies within the former region (MHC class III region) and is linked to expression of the anti-viral gene SKIV2L as mentioned in the results. So too does rs477515, linked in a study of prepouch ileum tissue from inflammatory bowel disease patients to the expression of over 250 genes including the HLA-DQ genes (Kabakchiev and Silverberg, 2013). rs477515 itself does not show strong evidence of being a regulatory variant, but may be a proxy through linkage disequilibrium for another SNP that directly acts in a regulatory region. While many regions associated with disease traits are enriched for eQTLs (Nicolae et al., 2010) and many SNPs associated with CD are demonstrated eQTLs (Dubois et al., 2010) the mechanisms underlying the associations of many genetic markers with disease are currently not known. However, the association of rs477515 with DQ expression emphasises that changes in the level of expression of the HLA-DQ genes may be an additional susceptibility factor over and above their dosage.

The association of a gene with RNA virus inhibiting activity (SKIV2L) (Aly et al., 2016) is interesting in view of recent findings that reovirus infections may be a triggering event in CD (Bouziat et al., 2017). Rotaviruses have also been implicated in risk of CD (Stene et al., 2006, Kemppainen et al., 2017). Our analysis of changing gene frequencies suggest patients in later time periods express less SKIV2L. While we have no direct evidence that SKIV2L acts against reoviruses, they are also double stranded RNA viruses similar to HBV the virus described by Aly et al. Bouziat et al. showed that coeliac patients had a tendency towards higher anti-reovirus antibody titres and had significantly more individuals with very high levels, which might be expected if these patients had lower levels of SKIV2L. This observation is interesting given that viral infection is an environmental factor that could therefore be associated with changes in presentation or prevalence.

In terms of the non-HLA genes we have already discussed the changes in gene frequency of eQTLs linked to altered FASLG expression. Looking at the table of markers which show large but not necessarily significant changes in allele frequency over time (supplementary Table 6, appendix XIV), we observe that almost all are in regions associated through GWAS with CD. Many are associated with altered expression of CD candidate genes, such as ICOSLG, UBE2E3, THEMIS, RGS1 and SH2B3/ATXN2.

As we have noted this study is underpowered to analyse some aspects of the CD phenotype and these findings will benefit from an increased population size. However, it is an exploratory study showing some evidence that elements of the clinical phenotype are related to genetic burden. The most notable case is the association of PRS with family history. We also have implemented a novel approach to identifying changing allele frequencies over time and have identified several SNPs that show significant derangement in frequency. We also have identified a significant shift in HLA-DQ2.5 frequency over time, mirroring changes in phenotype, principally age at diagnosis and changes in classical versus non-classical disease. We believe that the current methods of selecting PRS may underestimate the magnitude of the HLA PRS and this is a matter that is currently being pursued in the context of this study.

## **Chapter 5 General Discussion**



This thesis has been a co-operative venture between clinicians, geneticists and statisticians to try to produce an updated report on the clinical phenotype of coeliac disease (CD) in Ireland. We have also tried to link the clinical phenotype to genotypic data in an effort to better understand the varied presentation of this condition, as well as its changes over time. Due to the difficulty finding and collecting CD patient phenotypes, this study is also the result of the collaboration of several hospitals in Ireland.

Our findings in Chapter 2 confirm many well known facts about CD, such as a female gender bias and a high prevalence of co-existent immune-mediated conditions. However, they also report for the first time in Ireland a radical change in CD clinical presentation over the past few decades. Thus, new patients profiles that would have not been suspected to have CD a few decades ago emerge. For example, it is interesting to see how the percentage of patients presenting overweight or obese has increased sharply in the past two decades, now representing approximately half of the patients at diagnosis (Chapter 2, Table 2.8). We have also demonstrated that CD patients have followed worldwide BMI trends when compared to the general population despite their lower absorption of nutrients in active disease (Figure 2.6).

It is also curious how the clinical profile of patients presenting overweight or obese seems to differ from the remainder, presenting mainly non-classically or subclinically and with a milder picture with less damage to the small bowel and osteoporosis (Chapter 2, Table 2.6). Therefore, this finding is important as higher suspicion and awareness is needed to diagnose CD patients presenting with these less severe clinical manifestations. Moreover, patients in this group seem to require more support once diagnosed with CD both to avoid weight gain and promote weight loss, but also potentially, according to our results, to promote compliance with the gluten free diet (GFD) (Figure 2.8).

Overweight and obesity, like CD, are thought to be associated to a greater or lesser extent to a dysbiosis of the gut microbiome (Harley and Karp, 2012) and are clearly on the increase. This dysbiosis in CD has been associated to, among other things, a low grade chronic inflammation and higher permeability of the gut mucosa which could increase transit of gliadin (Verdu et al., 2015, De Re et al., 2017). It is not clear whether the association between the microbiome and CD development is causal. However, the fact that patients with higher BMI present with different clinical characteristics at diagnosis (Table 2.6) may support the theory that differences in the microbiome in higher BMI patients could promote CD development in individuals with an otherwise lower risk for the condition through higher gut permeability and chronic

inflammation. On the other hand, it could also be that higher BMI (which is increasing universally) has a protective effect from the severe manifestations of CD in genetically predisposed patients. In any case, discerning the basis of the relationship between BMI and autoimmune conditions such as CD is highly desirable.

Another type of patient profile that arises from our study are patients diagnosed >65 years of age who are as prevalent as those diagnosed <18 years of age. Again a more subtle clinical presentation, as in the case of the high BMI patients, is observed in this population. Once more, this is an important finding, as delayed diagnoses are common in this group given that CD may not necessarily be the first line of investigation. Therefore, it is clear that currently, CD in Ireland is a condition that presents mainly with a non-classical phenotype and is diagnosed later in life. This trend has been reported internationally both in the paediatric and adult CD population. This has led to a debate on whether it has been caused by better diagnostic techniques and awareness of the condition or, on the contrary, it represents a true rise in the amount of CD cases due to changing environmental factors.

The Swedish epidemic of CD was an important inflection point in the history of the disease. The abrupt increase in childhood CD incidence and the subsequent sharp decrease after changes in infant feeding policy, lead not only to confirmation that CD aetiology was complex, but also easily malleable in genetically predisposed infants by the right combination of genetic and environmental factors. Ivarsson (2005) depicted a hypothetical model of CD pathogenesis in infants whereby genotype and dietary gluten are necessary causal elements influenced by additional factors to precipitate disease development (Figure 1.1). Therefore, throughout an individual's lifespan, their genotype interacts with the environment and is exposed to various challenges which change in their presence or absence and level over time.

In recent years, many publications resulting from the multinational TEDDY study following children at high risk for type 1 diabetes (T1DM) and CD have led to a changed understanding of the factors believed to be associated with CD development. Notably, breastfeeding is no longer considered to be protective in itself and the timing of gluten introduction is no longer thought to increase the risk of CD. Nevertheless, the amount of gluten consumed during the first two years of life remains an important independent factor which increases disease risk in the paediatric population. The study by Kempainen *et al.* (2017) increased the complexity of the previously observed associations in the TEDDY cohort showing how the risk of developing CD was greater after gastrointestinal infection in early life, and was modified by other factors. Thus the risk was increased in the cohort of non HLA-DQ2 carriers who were breastfed for less

than 4 months, in those introduced to gluten before 6 months, and in infants born during the winter. The risk was decreased in children who had received rotavirus vaccination. Bouziat et al., (2017) have confirmed the link with viruses at a molecular level, showing that the closely related reoviruses are capable of causing a breakdown of oral tolerance to dietary gluten. Therefore, it may be that we cannot categorically rule out certain factors such as breastfeeding as being protective against CD, and probably most of those discussed above are important. This paints a picture of a complex system, where the impact of various environmental exposures depends on their timing and interaction with other environmental factors, as well as with certain genotypes.

The TEDDY study has several limitations. First of all, longer follow-up is needed as TEDDY only describes events happening in early life. Secondly, the identified environmental factors are based on a cohort of children with high risk genotypes and therefore may not apply to the general population. Finally, and more importantly, these risk factors may vary from those causing CD at older ages.

In fact, little is known about environmental factors leading to loss of gluten tolerance in adulthood despite different studies showing a true increase in incidence in the adult population. The recently reported association with reovirus discussed above may be an important exception. Given this association with reoviral infection, our own finding showing that a gene with antiviral effects (SKIV2L) appears to have altered regulation, potentially leading to reduced expression in recent years is interesting. Large prospective genotype/phenotype studies of the general population which include the collection of data on environmental exposures will be important in helping to shed light into the mechanisms of adult CD.

Another interesting aspect of our study is the changing relationship of CD with autoimmune thyroid disease (ATD). We could not collect environmental exposures due to the retrospective nature of our study. However, ATD has decreased over time which is not explained by older age at data collection or age of diagnosis with CD, but rather by calendar year of diagnosis (Chapter 3, Table 3.8). This seems to support the theory that changing environmental pressures are having an effect on new genotype combinations and potentially leading to the new clinical presentation of CD. The observed decrease in dermatitis herpetiformis (DH) represents an analogous situation. Moreover, the lack of difference seen when we segregate our sample by region of diagnosis implies that environmental factors rather than higher awareness are driving these observed changes in disease presentation.

We have furthermore confirmed the strong link between CD and other IMCs, and have measured the risk of these in CD patients compared to the general population (Table 3.5). Thus strong risk is observed for the development of T1DM and ATD, and a moderate risk is observed for the development of IBD. Perhaps shifting, common, environmental factors could be behind the intriguingly rising incidence of CD, T1DM and IBD. Importantly, in the case of T1DM, in a scenario analogous to that we observe in CD as discussed below, increased incidence has been linked to a higher prevalence of lower risk genotypes in recent decades.

Therefore, we have investigated in chapter 4 whether temporal genotype changes in CD can be linked to the previously observed varying clinical phenotype. As an approach to linking the overall burden imposed by genetic susceptibility loci with phenotype in CD, we used risk scores (PRS) and measured their association with a variety of clinical characteristics. Interestingly, we found that the HLA & non-HLA PRS was significantly higher in the case of patients with a family history of CD (Table 4.1). So when we focus solely on these patients we find that the PRS can predict age of diagnosis; we believe this is because the type of CD these people suffer is more 'genetically determined' and also, they are also more disease aware, reducing both genetic and diagnostic delay-related noise in this group.

The PRS is an aggregate score, and thus something of a "black box", and we wanted to break it out into its components to understand the mechanisms underlying their change over time. By doing this we were able to identify specific SNPs that were contributing to the change in PRS. Unfortunately, we are limited by the size of our population, however we believe we can identify many more by building on this approach. Our finding that age of diagnosis can be distinguished in the group of the patients with a family history shows this could be a useful tool with potential clinical applications, for example high risk patients with inactive CD could be classified according to their priority for follow-up. Nonetheless, our own experience indicates that PRS currently have limitations and may need to be adapted to allow the prioritisation of genetic information known to be functionally relevant in a given disease, instead of the rather random mechanism used to build scores in the most commonly used methods currently.

Finally, our PRS points to a decreasing genetic burden over time which becomes significant when we collapse the time points into a binary variable. Therefore, we have observed a changing clinical picture that is linked to a decrease in genetic burden over the duration of our study. This is particularly pronounced in the case of HLA-DQ 2.5 - heterozygotes have latterly become more common in the patient population, which may

partly explain the increase in frequency of the disease. We also observe other HLA and non-HLA SNPs to be varying in frequency. This implies that patients are being drawn from a changing genetic pool under the influence of mutating environmental factors.

## Future Directions

In this study we have attempted to link epidemiological data on CD clinical presentation with genetic burden. Our results are promising, and for the first time associate the well reported change in CD clinical presentation to genetic changes.

Our study would benefit from increasing the amount of patients with clinical phenotype and genotype information, especially those diagnosed in the period  $\leq 1985$ . This would increase the power of our analysis allowing thus for the identification of more SNPs which have significantly changed in frequency between periods.

In a future study we would also attempt to increase the sample of patients with ATD to see if any genetic differences can be established between CD patients with co-existent ATD and patients with only CD.

It will be worth expanding what the environmental and genotype interactions may be to include: obesity, gluten itself, innate factors, or even allergic factors associated with gluten.

Future studies should consider different approaches to the current PRS as excessive pruning could result in loss of functionally important information, reducing thus their clinical utility.

## Chapter 6 Bibliography

- ABADIE, V., SOLLID, L. M., BARREIRO, L. B. & JABRI, B. 2011. Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis. *In: PAUL, W. E., LITTMAN, D. R. & YOKOYAMA, W. M. (eds.) Annual Review of Immunology, Vol 29.*
- ABECASIS, G. R., ALTSHULER, D., AUTON, A., BROOKS, L. D., DURBIN, R. M., GIBBS, R. A., HURLES, M. E. & MCVEAN, G. A. 2010. A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061-73.
- ABENAVOLI, L., PROIETTI, I., LEGGIO, L., FERRULLI, A., VONGHIA, L., CAPIZZI, R., ROTOLI, M., AMERIO, P. L., GASBARRINI, G. & ADDOLORATO, G. 2006. Cutaneous manifestations in celiac disease. *World J Gastroenterol*, 12, 843-52.
- ABRAHAM, G., ROHMER, A., TYE-DIN, J. A. & INOUE, M. 2015. Genomic prediction of celiac disease targeting HLA-positive individuals. *Genome Med*, 7, 72.
- ABRAHAM, G., TYE-DIN, J. A., BHALALA, O. G., KOWALCZYK, A., ZOBEL, J. & INOUE, M. 2014. Accurate and robust genomic prediction of celiac disease using statistical learning. *PLoS Genet*, 10, e1004137.
- ABSAH, I., RISHI, A. R., GEBRAIL, R., SNYDER, M. R. & MURRAY, J. A. 2017. Lack of Utility of Anti-tTG IgG to Diagnose Celiac Disease When Anti-tTG IgA Is Negative. *J Pediatr Gastroenterol Nutr*, 64, 726-729.
- AINE, L., MAKI, M. & REUNALA, T. 1992. Coeliac-type dental enamel defects in patients with dermatitis herpetiformis. *Acta Derm Venereol*, 72, 25-7.
- AKOBENG, A. K. & THOMAS, A. G. 2008. Systematic review: tolerable amount of gluten for people with coeliac disease. *Aliment Pharmacol Ther*, 27, 1044-52.
- AL-BAWARDY, B., CODIPILLY, D. C., RUBIO-TAPIA, A., BRUINING, D. H., HANSEL, S. L. & MURRAY, J. A. 2017. Celiac disease: a clinical review. *Abdom Radiol (NY)*.
- ALONSO, A. & HERNAN, M. A. 2008. Temporal trends in the incidence of multiple sclerosis: a systematic review. *Neurology*, 71, 129-35.
- ALVES, S. C., FASANO, S. & ISENBERG, D. A. 2016. Autoimmune gastrointestinal complications in patients with systemic lupus erythematosus: case series and literature review. *Lupus*, 25, 1509-1519.
- ALY, H. H., SUZUKI, J., WATASHI, K., CHAYAMA, K., HOSHINO, S., HIJIKATA, M., KATO, T. & WAKITA, T. 2016. RNA Exosome Complex Regulates Stability of the Hepatitis B Virus X-mRNA Transcript in a Non-stop-mediated (NSD) RNA Quality Control Mechanism. *J Biol Chem*, 291, 15958-74.
- ANANTHAKRISHNAN, A. N., HUANG, H., NGUYEN, D. D., SAUK, J., YAJNIK, V. & XAVIER, R. J. 2014. Differential effect of genetic burden on disease phenotypes in Crohn's disease and ulcerative colitis: analysis of a North American cohort. *Am J Gastroenterol*, 109, 395-400.
- ANDREN ARONSSON, C., KURPPA, K. & AGARDH, D. 2016a. Gluten in infants and celiac disease risk. *Expert Rev Gastroenterol Hepatol*, 10, 669-70.
- ANDREN ARONSSON, C., LEE, H. S., KOLETZKO, S., UUSITALO, U., YANG, J., VIRTANEN, S. M., LIU, E., LERNMARK, A., NORRIS, J. M. & AGARDH, D. 2015. Effects of Gluten Intake on Risk of Celiac Disease: a case-control study on a Swedish birth cohort. *Clin Gastroenterol Hepatol*.
- ANDREN ARONSSON, C., LEE, H. S., KOLETZKO, S., UUSITALO, U., YANG, J., VIRTANEN, S. M., LIU, E., LERNMARK, A., NORRIS, J. M. & AGARDH, D. 2016b. Effects of Gluten Intake on Risk of Celiac Disease: A Case-Control Study on a Swedish Birth Cohort. *Clin Gastroenterol Hepatol*, 14, 403-409.e3.
- ANTONI, L., NUDING, S., WEHKAMP, J. & STANGE, E. F. 2014. Intestinal barrier in inflammatory bowel disease. *World J Gastroenterol*, 20, 1165-79.
- ANTVORSKOV, J. C., JOSEFSEN, K., HAUPT-JORGENSEN, M., FUNDOVA, P., FUNDA, D. P. & BUSCHARD, K. 2016. Gluten-Free Diet Only during Pregnancy Efficiently Prevents Diabetes in NOD Mouse Offspring. *J Diabetes Res*, 2016, 3047574.



- ARKEMA, E. V., GRUNEWALD, J., KULLBERG, S., EKLUND, A. & ASKLING, J. 2016. Sarcoidosis incidence and prevalence: a nationwide register-based assessment in Sweden. *Eur Respir J*, 48, 1690-1699.
- ARNOLD, I. C., HITZLER, I. & MULLER, A. 2012. The immunomodulatory properties of *Helicobacter pylori* confer protection against allergic and chronic inflammatory disorders. *Front Cell Infect Microbiol*, 2, 10.
- ARONSSON, C. A., LEE, H. S., LIU, E., UUSITALO, U., HUMMEL, S., YANG, J., HUMMEL, M., REWERS, M., SHE, J. X., SIMELL, O., TOPPARI, J., ZIEGLER, A. G., KRISCHER, J., VIRTANEN, S. M., NORRIS, J. M. & AGARDH, D. 2015. Age at gluten introduction and risk of celiac disease. *Pediatrics*, 135, 239-45.
- BAKKER, S. F., TUSHUIZEN, M. E., VON BLOMBERG, B. M., BONTKES, H. J., MULDER, C. J. & SIMSEK, S. 2016. Screening for coeliac disease in adult patients with type 1 diabetes mellitus: myths, facts and controversy. *Diabetol Metab Syndr*, 8, 51.
- BAKRY, O. A., EL FARARGY, S. M., EL SHAFIEE, M. K. & SOLIMAN, A. 2016. Serum Vitamin D in patients with alopecia areata. *Indian Dermatol Online J*, 7, 371-377.
- BALTACI, D., KUTLUCAN, A., TURKER, Y., YILMAZ, A., KARACAM, S., DELER, H., UCGUN, T. & KARA, I. H. 2013. Association of vitamin B12 with obesity, overweight, insulin resistance and metabolic syndrome, and body fat composition; primary care-based study. *Med Glas (Zenica)*, 10, 203-10.
- BANNISTER, E. G., CAMERON, D. J., NG, J., CHOW, C. W., OLIVER, M. R., ALEX, G., CATTO-SMITH, A. G., HEINE, R. G., WEBB, A., MCGRATH, K., SIMPSON, D. & HARDIKAR, W. 2014. Can celiac serology alone be used as a marker of duodenal mucosal recovery in children with celiac disease on a gluten-free diet? *Am J Gastroenterol*, 109, 1478-83.
- BARONE, M., DELLA VALLE, N., ROSANIA, R., FACCIORUSSO, A., TROTTA, A., CANTATORE, F. P., FALCO, S., PIGNATIELLO, S., VIGGIANI, M. T., AMORUSO, A., DE FILIPPIS, R., DI LEO, A. & FRANCAVILLA, R. 2016. A comparison of the nutritional status between adult celiac patients on a long-term, strictly gluten-free diet and healthy subjects. *Eur J Clin Nutr*, 70, 23-7.
- BAYDOUN, A., MAAKARON, J. E., HALAWI, H., ABOU RAHAL, J. & TAHER, A. T. 2012. Hematological manifestations of celiac disease. *Scand J Gastroenterol*, 47, 1401-11.
- BERMEJO, F. & GARCIA-LOPEZ, S. 2009. A guide to diagnosis of iron deficiency and iron deficiency anemia in digestive diseases. *World J Gastroenterol*, 15, 4638-43.
- BHATIA, B. K., MILLSOP, J. W., DEBBANEH, M., KOO, J., LINOS, E. & LIAO, W. 2014. Diet and psoriasis, part II: celiac disease and role of a gluten-free diet. *J Am Acad Dermatol*, 71, 350-8.
- BIRD, J. K., RONNENBERG, A. G., CHOI, S. W., DU, F., MASON, J. B. & LIU, Z. 2015. Obesity is associated with increased red blood cell folate despite lower dietary intakes and serum concentrations. *J Nutr*, 145, 79-86.
- BODE, S. & GUDMAND-HOYER, E. 1996. Symptoms and haematologic features in consecutive adult coeliac patients. *Scand J Gastroenterol*, 31, 54-60.
- BOLOTIN, D. & PETRONIC-ROSIC, V. 2011. Dermatitis herpetiformis. Part I. Epidemiology, pathogenesis, and clinical presentation. *J Am Acad Dermatol*, 64, 1017-24; quiz 1025-6.
- BONDAVALLI, P., QUADRI, G., PARODI, A. & REBORA, A. 1998. Failure of gluten-free diet in celiac disease-associated alopecia areata. *Acta Derm Venereol*, 78, 319.
- BORCHERS, A. T., UIBO, R. & GERSHWIN, M. E. 2010. The geoepidemiology of type 1 diabetes. *Autoimmun Rev*, 9, A355-65.
- BOSCA-WATTS, M. M., TOSCA, J., ANTON, R., MORA, M., MINGUEZ, M. & MORA, F. 2015. Pathogenesis of Crohn's disease: Bug or no bug. *World J Gastrointest Pathophysiol*, 6, 1-12.
- BOUZIAT, R., HINTERLEITNER, R., BROWN, J. J., STENCEL-BAERENWALD, J. E., IKIZLER, M., MAYASSI, T., MEISEL, M., KIM, S. M., DISCEPOLO, V., PRUIJSSERS, A. J., ERNEST, J. D., ISKARPATYOTI, J. A., COSTES, L. M., LAWRENCE, I., PALANSKI, B. A., VARMA, M.,

- ZURENSKI, M. A., KHOMANDIAK, S., MCALLISTER, N., ARAVAMUDHAN, P., BOEHME, K. W., HU, F., SAMSOM, J. N., REINECKER, H. C., KUPFER, S. S., GUANDALINI, S., SEMRAD, C. E., ABADIE, V., KHOSLA, C., BARREIRO, L. B., XAVIER, R. J., NG, A., DERMODY, T. S. & JABRI, B. 2017. Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science*, 356, 44-50.
- BROWN, J., WIMPENNY, P. & MAUGHAN, H. 2004. Skin problems in people with obesity. *Nurs Stand*, 18, 38-42.
- BULGER, K., GRIFFIN, M., DERVAN, P., LENNON, J. & CROWE, J. 1988. Coeliac disease in association with inflammatory bowel disease. *Postgrad Med J*, 64, 336.
- BURK, K., FARECKI, M. L., LAMPRECHT, G., ROTH, G., DECKER, P., WELLER, M., RAMMENSEE, H. G. & OERTEL, W. 2009. Neurological symptoms in patients with biopsy proven celiac disease. *Mov Disord*, 24, 2358-62.
- BUSH, W. S., SAWCER, S. J., DE JAGER, P. L., OKSENBERG, J. R., MCCAULEY, J. L., PERICAK-VANCE, M. A. & HAINES, J. L. 2010. Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. *Am J Hum Genet*, 86, 621-5.
- BUSHARA, K. O. 2005. Neurologic presentation of celiac disease. *Gastroenterology*, 128, S92-7.
- CAPUANO, M., IAFFALDANO, L., TINTO, N., MONTANARO, D., CAPOBIANCO, V., IZZO, V., TUCCI, F., TRONCONE, G., GRECO, L. & SACCHETTI, L. 2011. MicroRNA-449a overexpression, reduced NOTCH1 signals and scarce goblet cells characterize the small intestine of celiac patients. *PLoS One*, 6, e29094.
- CASELLA, S., ZANINI, B., LANZAROTTO, F., VILLANACCI, V., RICCI, C. & LANZINI, A. 2012. Celiac disease in elderly adults: clinical, serological, and histological characteristics and the effect of a gluten-free diet. *J Am Geriatr Soc*, 60, 1064-9.
- CASTILLO, N. E., THEETHIRA, T. G. & LEFFLER, D. A. 2015. The present and the future in the diagnosis and management of celiac disease. *Gastroenterol Rep (Oxf)*, 3, 3-11.
- CATALDO, F. & MARINO, V. 2003. Increased prevalence of autoimmune diseases in first-degree relatives of patients with celiac disease. *J Pediatr Gastroenterol Nutr*, 36, 470-3.
- CATALDO, F., MARINO, V., VENTURA, A., BOTTARO, G. & CORAZZA, G. R. 1998. Prevalence and clinical features of selective immunoglobulin A deficiency in coeliac disease: an Italian multicentre study. Italian Society of Paediatric Gastroenterology and Hepatology (SIGEP) and "Club del Tenue" Working Groups on Coeliac Disease. *Gut*, 42, 362-5.
- CATASSI, C., DOLORETTA MACIS, M., RATSCH, I. M., DE VIRGILIIS, S. & CUCCA, F. 2001. The distribution of DQ genes in the Saharawi population provides only a partial explanation for the high celiac disease prevalence. *Tissue Antigens*, 58, 402-6.
- CATASSI, C. & FASANO, A. 2010. Celiac disease diagnosis: simple rules are better than complicated algorithms. *Am J Med*, 123, 691-3.
- CATASSI, C., GATTI, S. & FASANO, A. 2014. The new epidemiology of celiac disease. *J Pediatr Gastroenterol Nutr*, 59 Suppl 1, S7-9.
- CATASSI, C., KRYSZAK, D., BHATTI, B., STURGEON, C., HELZLSOUER, K., CLIPP, S. L., GELFOND, D., PUPPA, E., SFERRUZZA, A. & FASANO, A. 2010. Natural history of celiac disease autoimmunity in a USA cohort followed since 1974. *Ann Med*, 42, 530-8.
- CH'NG, C. L., JONES, M. K. & KINGHAM, J. G. 2007. Celiac disease and autoimmune thyroid disease. *Clin Med Res*, 5, 184-92.
- CHARNEY, A. W., RUDERFER, D. M., STAHL, E. A., MORAN, J. L., CHAMBERT, K., BELLIVEAU, R. A., FORTY, L., GORDON-SMITH, K., DI FLORIO, A., LEE, P. H., BROMET, E. J., BUCKLEY, P. F., ESCAMILLA, M. A., FANOUS, A. H., FOCHTMANN, L. J., LEHRER, D. S., MALASPINA, D., MARDER, S. R., MORLEY, C. P., NICOLINI, H., PERKINS, D. O., RAKOFSKY, J. J., RAPAPORT, M. H., MEDEIROS, H., SOBELL, J. L., GREEN, E. K., BACKLUND, L., BERGEN, S. E., JUREUS, A., SCHALLING, M., LICHTENSTEIN, P.,

- ROUSSOS, P., KNOWLES, J. A., JONES, I., JONES, L. A., HULTMAN, C. M., PERLIS, R. H., PURCELL, S. M., MCCARROLL, S. A., PATO, C. N., PATO, M. T., CRADDOCK, N., LANDEN, M., SMOLLER, J. W. & SKLAR, P. 2017. Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. *Transl Psychiatry*, 7, e993.
- CHEN, X., KUJA-HALKOLA, R., RAHMAN, I., ARPEGARD, J., VIKTORIN, A., KARLSSON, R., HAGG, S., SVENSSON, P., PEDERSEN, N. L. & MAGNUSSON, P. K. 2015. Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. *Am J Hum Genet*, 97, 708-14.
- CHENG, J., BRAR, P. S., LEE, A. R. & GREEN, P. H. 2010. Body mass index in celiac disease: beneficial effect of a gluten-free diet. *J Clin Gastroenterol*, 44, 267-71.
- CHIBNIK, L. B., KEENAN, B. T., CUI, J., LIAO, K. P., COSTENBADER, K. H., PLENGE, R. M. & KARLSON, E. W. 2011. Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. *PLoS One*, 6, e24380.
- CHIOVATO, L., LAPI, P., FIORE, E., TONACCHERA, M. & PINCHERA, A. 1993. Thyroid autoimmunity and female gender. *J Endocrinol Invest*, 16, 384-91.
- CHOUNG, R. S., DITAH, I. C., NADEAU, A. M., RUBIO-TAPIA, A., MARIETTA, E. V., BRANTNER, T. L., CAMILLERI, M. J., RAJKUMAR, S. V., LANDGREN, O., EVERHART, J. E. & MURRAY, J. A. 2015. Trends and racial/ethnic disparities in gluten-sensitive problems in the United States: findings from the National Health and Nutrition Examination Surveys from 1988 to 2012. *Am J Gastroenterol*, 110, 455-61.
- CHOUNG RS, MARIETTA EV, VAN DYKE CT, BRANTNER TL, RAJASEKARAN J, PASRICHA PJ, ET AL. 2016 Determination of B-Cell Epitopes in Patients with Celiac Disease: Peptide Microarrays. *PloS one*, 11(1):e0147777.
- CHOW, M. A., LEBWOHL, B., REILLY, N. R. & GREEN, P. H. 2012. Immunoglobulin A deficiency in celiac disease. *J Clin Gastroenterol*, 46, 850-4.
- CLEYNEN, I., BOUCHER, G., JOSTINS, L., SCHUMM, L. P., ZEISSIG, S., AHMAD, T., ANDERSEN, V., ANDREWS, J. M., ANNESE, V., BRAND, S., BRANT, S. R., CHO, J. H., DALY, M. J., DUBINSKY, M., DUERR, R. H., FERGUSON, L. R., FRANKE, A., GEARRY, R. B., GOYETTE, P., HAKONARSON, H., HALFVARSON, J., HOV, J. R., HUANG, H., KENNEDY, N. A., KUPCINSKAS, L., LAWRENCE, I. C., LEE, J. C., SATSANGI, J., SCHREIBER, S., THÉÂTRE, E., VAN DER MEULEN-DE JONG, A. E., WEERSMA, R. K., WILSON, D. C., PARKES, M., VERMEIRE, S., RIOUX, J. D., MANSFIELD, J., SILVERBERG, M. S., RADFORD-SMITH, G., MCGOVERN, D. P. B., BARRETT, J. C. & LEES, C. W. 2016. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*, 387, 156-67.
- CODEX, A. 2008.
- COLEMAN, C., QUINN, E. M., RYAN, A. W., CONROY, J., TRIMBLE, V., MAHMUD, N., KENNEDY, N., CORVIN, A. P., MORRIS, D. W., DONOHOE, G., O'MORAIN, C., MACMATHUNA, P., BYRNES, V., KIAT, C. & TRYNKA, G. 2015. Common polygenic variation in coeliac disease and confirmation of ZNF335 and NIFA as disease susceptibility loci.
- COLLIN, P., HUHTALA, H., VIRTÄ, L., KEKKONEN, L. & REUNALA, T. 2007. Diagnosis of celiac disease in clinical practice: physician's alertness to the condition essential. *J Clin Gastroenterol*, 41, 152-6.
- COLLIN, P., KAUKINEN, K., VALIMAKI, M. & SALMI, J. 2002. Endocrinological disorders and celiac disease. *Endocr Rev*, 23, 464-83.
- COLLIN, P., SALMI, T. T., HERVONEN, K., KAUKINEN, K. & REUNALA, T. 2017. Dermatitis herpetiformis: a cutaneous manifestation of coeliac disease. *Ann Med*, 49, 23-31.
- COLOMBE, B. W., LOU, C. D. & PRICE, V. H. 1999. The genetic basis of alopecia areata: HLA associations with patchy alopecia areata versus alopecia totalis and alopecia universalis. *J Investig Dermatol Symp Proc*, 4, 216-9.

- COLSTON, K. W., MACKAY, A. G., FINLAYSON, C., WU, J. C. & MAXWELL, J. D. 1994. Localisation of vitamin D receptor in normal human duodenum and in patients with coeliac disease. *Gut*, 35, 1219-25.
- COMINO, I., FERNANDEZ-BANARES, F., ESTEVE, M., ORTIGOSA, L., CASTILLEJO, G., FAMBUENA, B., RIBES-KONINCKX, C., SIERRA, C., RODRIGUEZ-HERRERA, A., SALAZAR, J. C., CAUNEDO, A., MARUGAN-MIGUELSANZ, J. M., GARROTE, J. A., VIVAS, S., LO IACONO, O., NUNEZ, A., VAQUERO, L., VEGAS, A. M., CRESPO, L., FERNANDEZ-SALAZAR, L., ARRANZ, E., JIMENEZ-GARCIA, V. A., ANTONIO MONTES-CANO, M., ESPIN, B., GALERA, A., VALVERDE, J., GIRON, F. J., BOLONIO, M., MILLAN, A., CEREZO, F. M., GUAJARDO, C., ALBERTO, J. R., ROSINACH, M., SEGURA, V., LEON, F., MARINICH, J., MUNOZ-SUANO, A., ROMERO-GOMEZ, M., CEBOLLA, A. & SOUSA, C. 2016. Fecal Gluten Peptides Reveal Limitations of Serological Tests and Food Questionnaires for Monitoring Gluten-Free Diet in Celiac Disease Patients. *Am J Gastroenterol*, 111, 1456-1465.
- COOPER, G. S., BYNUM, M. L. & SOMERS, E. C. 2009. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun*, 33, 197-207.
- CORAZZA, G. R., ANDREANI, M. L., VENTURO, N., BERNARDI, M., TOSTI, A. & GASBARRINI, G. 1995. Celiac disease and alopecia areata: report of a new association. *Gastroenterology*, 109, 1333-7.
- CORAZZA, G. R. & VILLANACCI, V. 2005. Coeliac disease. *J Clin Pathol*, 58, 573-4.
- CORTES, A. & BROWN, M. A. 2011. Promise and pitfalls of the Immunochip. *Arthritis Res Ther*, 13, 101.
- COSNES, J., CELLIER, C., VIOLA, S., COLOMBEL, J. F., MICHAUD, L., SARLES, J., HUGOT, J. P., GINIES, J. L., DABADIE, A., MOUTERDE, O., ALLEZ, M. & NION-LARMURIER, I. 2008. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *Clin Gastroenterol Hepatol*, 6, 753-8.
- CRANNEY, A., ZARKADAS, M., GRAHAM, I. D., BUTZNER, J. D., RASHID, M., WARREN, R., MOLLOY, M., CASE, S., BURROWS, V. & SWITZER, C. 2007. The Canadian Celiac Health Survey. *Dig Dis Sci*, 52, 1087-95.
- CRISCOV, G. I., STANA, A. B., IONIUE, I. K., ALEXOAE, M. M. & MORARU, E. 2015. Coexistence of celiac disease and systemic lupus erythematosus in a 6-year-old girl-case report. *Rev Med Chir Soc Med Nat Iasi*, 119, 87-91.
- D'ERCOLE, C., ZULLO, A., BRAGAZZI, M. C. & CAMPO, S. M. 2012. Sarcoidosis and coeliac disease: do not forget the association! *Intern Emerg Med*, 7 Suppl 1, S25.
- DAHELE, A. & GHOSH, S. 2001. Vitamin B12 deficiency in untreated celiac disease. *Am J Gastroenterol*, 96, 745-50.
- DAHLBOM, I., OLSSON, M., FOROOZ, N. K., SJOHOLM, A. G., TRUEDSSON, L. & HANSSON, T. 2005. Immunoglobulin G (IgG) anti-tissue transglutaminase antibodies used as markers for IgA-deficient celiac disease patients. *Clin Diagn Lab Immunol*, 12, 254-8.
- DE ANDRADE, M., JACKOW, C. M., DAHM, N., HORDINSKY, M., REVEILLE, J. D. & DUVIC, M. 1999. Alopecia areata in families: association with the HLA locus. *J Invest Dermatol Symp Proc*, 4, 220-3.
- DE OLIVEIRA, P., DE CARVALHO, D. R., BRANDI, I. V. & PRATESI, R. 2016. Serological prevalence of celiac disease in Brazilian population of multiple sclerosis, neuromyelitis optica and myelitis. *Mult Scler Relat Disord*, 9, 125-8.
- DE RE, V., MAGRIS, R. & CANNIZZARO, R. 2017. New Insights into the Pathogenesis of Celiac Disease. *Front Med (Lausanne)*, 4, 137.
- DECKER, E., ENGELMANN, G., FINDEISEN, A., GERNER, P., LAASS, M., NEY, D., POSOVSKY, C., HOY, L. & HORNEF, M. W. 2010. Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children. *Pediatrics*, 125, e1433-40.

- DECKER, E., HORNEF, M. & STOCKINGER, S. 2011. Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children. *Gut Microbes*, 2, 91-8.
- DI MARIO, U., ANASTASI, E., MARIANI, P., BALLATI, G., PERFETTI, R., TRIGLIONE, P., MORELLINI, M. & BONAMICO, M. 1992. Diabetes-related autoantibodies do appear in children with coeliac disease. *Acta Paediatr*, 81, 593-7.
- DI STEFANO, M., MENGOLI, C., BERGONZI, M. & CORAZZA, G. R. 2013. Bone mass and mineral metabolism alterations in adult celiac disease: pathophysiology and clinical approach. *Nutrients*, 5, 4786-99.
- DIAZ-AMIGO, C. & POPPING, B. 2012. Gluten and gluten-free: issues and considerations of labeling regulations, detection methods, and assay validation. *JAOAC Int*, 95, 337-48.
- DICKE, W. K., WEIJERS, H. A. & VAN DE KAMER, J. H. 1953. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatr*, 42, 34-42.
- DICKEY, W. & KEARNEY, N. 2006. Overweight in celiac disease: prevalence, clinical characteristics, and effect of a gluten-free diet. *Am J Gastroenterol*, 101, 2356-9.
- DICKSON, B. C., STREUTKER, C. J. & CHETTY, R. 2006. Coeliac disease: an update for pathologists. *J Clin Pathol*, 59, 1008-16.
- DIETERICH, W., LAAG, E., SCHOPPER, H., VOLTA, U., FERGUSON, A., GILLET, H., RIECKEN, E. O. & SCHUPPAN, D. 1998. Autoantibodies to tissue transglutaminase as predictors of celiac disease. *Gastroenterology*, 115, 1317-21.
- DITAH, I. C., MARIETTA, E., NADEAU, A., BRANTNER, T. L., RUBIO-TAPIA, A. & MURRAY, J. A. 2014. Su1446 Despite an Increased Use of a Gluten Free Diet, Most Celiac Disease Remains Undiagnosed in the USA. *Gastroenterology*, 146, S-471.
- DOGAN, Y., YILDIRMAZ, S. & OZERCAN, I. H. 2012. Prevalence of celiac disease among first-degree relatives of patients with celiac disease. *J Pediatr Gastroenterol Nutr*, 55, 205-8.
- DONALDSON, M. R., FIRTH, S. D., WIMPEE, H., LEIFERMAN, K. M., ZONE, J. J., HORSLEY, W., O'GORMAN, M. A., JACKSON, W. D., NEUHAUSEN, S. L., HULL, C. M. & BOOK, L. S. 2007. Correlation of duodenal histology with tissue transglutaminase and endomysial antibody levels in paediatric celiac disease. *Clin Gastroenterol Hepatol*, 5, 567-73.
- DUBE, C., ROSTOM, A., SY, R., CRANNEY, A., SALOOJEE, N., GARRITTY, C., SAMPSON, M., ZHANG, L., YAZDI, F., MAMALADZE, V., PAN, I., MACNEIL, J., MACK, D., PATEL, D. & MOHER, D. 2005. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology*, 128, S57-67.
- DUBOIS, P. C., TRYNKA, G., FRANKE, L., HUNT, K. A., ROMANOS, J., CURTOTTI, A., ZHERNAKOVA, A., HEAP, G. A., ADANY, R., AROMAA, A., BARDELLA, M. T., VAN DEN BERG, L. H., BOCKETT, N. A., DE LA CONCHA, E. G., DEMA, B., FEHRMANN, R. S., FERNANDEZ-ARQUERO, M., FIATAL, S., GRANDONE, E., GREEN, P. M., GROEN, H. J., GWILLIAM, R., HOUWEN, R. H., HUNT, S. E., KAUKINEN, K., KELLEHER, D., KORPONAY-SZABO, I., KURPPA, K., MACMATHUNA, P., MAKI, M., MAZZILLI, M. C., MCCANN, O. T., MEARIN, M. L., MEIN, C. A., MIRZA, M. M., MISTRY, V., MORA, B., MORLEY, K. I., MULDER, C. J., MURRAY, J. A., NUNEZ, C., OOSTEROM, E., OPHOFF, R. A., POLANCO, I., PELTONEN, L., PLATTEEL, M., RYBAK, A., SALOMAA, V., SCHWEIZER, J. J., SPERANDEO, M. P., TACK, G. J., TURNER, G., VELDINK, J. H., VERBEEK, W. H., WEERSMA, R. K., WOLTERS, V. M., URCELAY, E., CUKROWSKA, B., GRECO, L., NEUHAUSEN, S. L., MCMANUS, R., BARISANI, D., DELOUKAS, P., BARRETT, J. C., SAAVALAINEN, P., WIJMENGA, C. & VAN HEEL, D. A. 2010. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*, 42, 295-302.

- DUBOIS, P. C. & VAN HEEL, D. A. 2008. Translational mini-review series on the immunogenetics of gut disease: immunogenetics of coeliac disease. *Clin Exp Immunol*, 153, 162-73.
- DUDBRIDGE, F. 2013. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, 9, e1003348.
- DUNTAS, L. H. 2009. Does celiac disease trigger autoimmune thyroiditis? *Nat Rev Endocrinol*, 5, 190-1.
- DUNTAS, L. H. 2015. The Role of Iodine and Selenium in Autoimmune Thyroiditis. *Horm Metab Res*, 47, 721-6.
- DUTTA, A. K. & CHACKO, A. 2016. Influence of environmental factors on the onset and course of inflammatory bowel disease. *World J Gastroenterol*, 22, 1088-100.
- EATON, W. W., ROSE, N. R., KALAYDJIAN, A., PEDERSEN, M. G. & MORTENSEN, P. B. 2007. Epidemiology of autoimmune diseases in Denmark. *J Autoimmun*, 29, 1-9.
- EBER, O. & LANGSTEGGER, W. 1994. [Clinical aspects of autoimmune thyroid diseases]. *Acta Med Austriaca*, 21, 1-7.
- ELFSTROM, P., MONTGOMERY, S. M., KAMPE, O., EKBOM, A. & LUDVIGSSON, J. F. 2008. Risk of thyroid disease in individuals with celiac disease. *J Clin Endocrinol Metab*, 93, 3915-21.
- ELKAN, A. C., SJOBERG, B., KOLSRUD, B., RINGERTZ, B., HAFSTROM, I. & FROSTEGARD, J. 2008. Gluten-free vegan diet induces decreased LDL and oxidized LDL levels and raised atheroprotective natural antibodies against phosphorylcholine in patients with rheumatoid arthritis: a randomized study. *Arthritis Res Ther*, 10, R34.
- ELLI, L., BONURA, A., GARAVAGLIA, D., RULLI, E., FLORIANI, I., TAGLIABUE, G., CONTIERO, P. & BARDELLA, M. T. 2012. Immunological comorbidity in coeliac disease: associations, risk factors and clinical implications. *J Clin Immunol*, 32, 984-90.
- ELSON, C. O. & CONG, Y. 2012. Host-microbiota interactions in inflammatory bowel disease. *Gut Microbes*, 3, 332-44.
- EMILSSON, L., WIJMENGA, C., MURRAY, J. A. & LUDVIGSSON, J. F. 2015. Autoimmune Disease in First-Degree Relatives and Spouses of Individuals With Celiac Disease. *Clin Gastroenterol Hepatol*, 13, 1271-1277.e2.
- ESCUDERO-HERNANDEZ, C., PENA, A. S. & BERNARDO, D. 2016. Immunogenetic Pathogenesis of Celiac Disease and Non-celiac Gluten Sensitivity. *Curr Gastroenterol Rep*, 18, 36.
- ESPGHAN 1990. Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch Dis Child*, 65, 909-11.
- FALLAHI, P., FERRARI, S. M., RUFFILLI, I., ELIA, G., BIRICOTTI, M., VITA, R., BENVENGA, S. & ANTONELLI, A. 2016. The association of other autoimmune diseases in patients with autoimmune thyroiditis: Review of the literature and report of a large series of patients. *Autoimmun Rev*.
- FARAGE, P., DE MEDEIROS NOBREGA, Y. K., PRATESI, R., GANDOLFI, L., ASSUNCAO, P. & ZANDONADI, R. P. 2017. Gluten contamination in gluten-free bakery products: a risk for coeliac disease patients. *Public Health Nutr*, 20, 413-416.
- FASANO, A. 2006. Systemic autoimmune disorders in celiac disease. *Curr Opin Gastroenterol*, 22, 674-9.
- FASANO, A. 2011. Zonulin and its regulation of intestinal barrier function: the biological door to inflammation, autoimmunity, and cancer. *Physiol Rev*, 91, 151-75.
- FASANO, A. 2012. Intestinal permeability and its regulation by zonulin: diagnostic and therapeutic implications. *Clin Gastroenterol Hepatol*, 10, 1096-100.
- FASANO, A., BERTI, I., GERARDUZZI, T., NOT, T., COLLETTI, R. B., DRAGO, S., ELITSUR, Y., GREEN, P. H., GUANDALINI, S., HILL, I. D., PIETZAK, M., VENTURA, A., THORPE, M., KRYSZAK, D., FORNAROLI, F., WASSERMAN, S. S., MURRAY, J. A. & HORVATH, K. 2003. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med*, 163, 286-92.

- FASANO, A., NOT, T., WANG, W., UZZAU, S., BERTI, I., TOMMASINI, A. & GOLDBLUM, S. E. 2000. Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease. *Lancet*, 355, 1518-9.
- FERNANDEZ-JIMENEZ, N., CASTELLANOS-RUBIO, A., PLAZA-IZURIETA, L., IRASTORZA, I., ELCOROARISTIZABAL, X., JAUREGI-MIGUEL, A., LOPEZ-EUBA, T., TUTAU, C., DE PANCORBO, M. M., VITORIA, J. C. & BILBAO, J. R. 2014. Coregulation and modulation of NFkappaB-related genes in celiac disease: uncovered aspects of gut mucosal inflammation. *Hum Mol Genet*, 23, 1298-310.
- FESSATOU, S., KOSTAKI, M. & KARPATIOS, T. 2003. Coeliac disease and alopecia areata in childhood. *J Paediatr Child Health*, 39, 152-4.
- FINGERLIN, T. E., HAMZE, N. & MAIER, L. A. 2015. Genetics of Sarcoidosis. *Clin Chest Med*, 36, 569-84.
- FINUCANE, M. M., STEVENS, G. A., COWAN, M. J., DANAIE, G., LIN, J. K., PACIOREK, C. J., SINGH, G. M., GUTIERREZ, H. R., LU, Y., BAHALIM, A. N., FARZADFAR, F., RILEY, L. M. & EZZATI, M. 2011. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet*, 377, 557-67.
- FITZPATRICK, D. J., RYAN, C. J., SHAH, N., GREENE, D., MOLONY, C. & SHIELDS, D. C. 2015. Genome-wide epistatic expression quantitative trait loci discovery in four human tissues reveals the importance of local chromosomal interactions governing gene expression. *BMC Genomics*, 16, 109.
- FORNARI, M. C., PEDREIRA, S., NIVELONI, S., GONZALEZ, D., DIEZ, R. A., VAZQUEZ, H., MAZURE, R., SUGAI, E., SMECUOL, E., BOERR, L., MAURINO, E. & BAI, J. C. 1998. Pre- and post-treatment serum levels of cytokines IL-1beta, IL-6, and IL-1 receptor antagonist in celiac disease. Are they related to the associated osteopenia? *Am J Gastroenterol*, 93, 413-8.
- FOURLANOS, S., VARNEY, M. D., TAIT, B. D., MORAHAN, G., HONEYMAN, M. C., COLMAN, P. G. & HARRISON, L. C. 2008. The rising incidence of type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. *Diabetes Care*, 31, 1546-9.
- FRANCIS, J., CARTY, J. E. & SCOTT, B. B. 2002. The prevalence of coeliac disease in rheumatoid arthritis. *Eur J Gastroenterol Hepatol*, 14, 1355-6.
- FREEDMAN, S. N., SHAHI, S. K. & MANGALAM, A. K. 2017. The "Gut Feeling": Breaking Down the Role of Gut Microbiome in Multiple Sclerosis. *Neurotherapeutics*.
- FREEMAN, H. J. 2008. Adult celiac disease in the elderly. *World J Gastroenterol*, 14, 6911-4.
- FREEMAN, H. J. 2017. Dietary compliance in celiac disease. *World J Gastroenterol*, 23, 2635-2639.
- FUCHS, V., KURPPA, K., HUHTALA, H., COLLIN, P., MAKI, M. & KAUKINEN, K. 2014. Factors associated with long diagnostic delay in celiac disease. *Scand J Gastroenterol*, 49, 1304-10.
- FUCHTENBUSCH, M., ZIEGLER, A. G. & HUMMEL, M. 2004. Elimination of dietary gluten and development of type 1 diabetes in high risk subjects. *Rev Diabet Stud*, 1, 39-41.
- FUNG, T. T., CHIUVE, S. E., MCCULLOUGH, M. L., REXRODE, K. M., LOGROSCINO, G. & HU, F. B. 2008. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med*, 168, 713-20.
- GALLI, G., ESPOSITO, G., LAHNER, E., PILOZZI, E., CORLETO, V. D., DI GIULIO, E., ALOE SPIRITI, M. A. & ANNIBALE, B. 2014. Histological recovery and gluten-free diet adherence: a prospective 1-year follow-up study of adult patients with coeliac disease. *Aliment Pharmacol Ther*, 40, 639-47.
- GARBER, J. R., COBIN, R. H., GHARIB, H., HENNESSEY, J. V., KLEIN, I., MECHANICK, J. I., PESSAH-POLLACK, R., SINGER, P. A. & WOEBER, K. A. 2012. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocr Pract*, 18, 988-1028.

- GARCIA-HORSMAN, J. A., VENALAINEN, J. I., LOHI, O., AURIOLA, I. S., KORPONAY-SZABO, I. R., KAUKINEN, K., MAKI, M. & MANNISTO, P. T. 2007. Deficient activity of mammalian prolyl oligopeptidase on the immunoactive peptide digestion in coeliac disease. *Scand J Gastroenterol*, 42, 562-71.
- GARNER, C., AHN, R., DING, Y. C., STEELE, L., STOVEN, S., GREEN, P. H., FASANO, A., MURRAY, J. A. & NEUHAUSEN, S. L. 2014. Genome-wide association study of celiac disease in North America confirms FRMD4B as new celiac locus. *PLoS One*, 9, e101428.
- GARNER, C. P., MURRAY, J. A., DING, Y. C., TIEN, Z., VAN HEEL, D. A. & NEUHAUSEN, S. L. 2009. Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet*, 18, 4219-25.
- GENT, A. E., HELLIER, M. D., GRACE, R. H., SWARBRICK, E. T. & COGGON, D. 1994. Inflammatory bowel disease and domestic hygiene in infancy. *Lancet*, 343, 766-7.
- GHAFOOR, R. & ANWAR, M. I. 2017. Vitamin D Deficiency in Alopecia Areata. *J Coll Physicians Surg Pak*, 27, 200-202.
- GOEL G, KING T, DAVESON AJ, ANDREWS JM, KRISHNARAJAH J, KRAUSE R, ET AL. 2017. Epitope-specific immunotherapy targeting CD4-positive T cells in coeliac disease: two randomised, double-blind, placebo-controlled phase 1 studies. *The lancet Gastroenterology & hepatology*, 2(7):479-93.
- GOLAN, D., LANDER, E. S. & ROSSET, S. 2014. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A*, 111, E5272-81.
- GOPALAKRISHNAN, S., TRIPATHI, A., TAMIZ, A. P., ALKAN, S. S. & PANDEY, N. B. 2012. Larazotide acetate promotes tight junction assembly in epithelial cells. *Peptides*, 35, 95-101.
- GRAY, A. M. & PAPANICOLAS, I. N. 2010. Impact of symptoms on quality of life before and after diagnosis of coeliac disease: results from a UK population survey. *BMC Health Serv Res*, 10, 105.
- GRECO, D., PISCIOTTA, M., GAMBINA, F. & MAGGIO, F. 2013. Celiac disease in subjects with type 1 diabetes mellitus: a prevalence study in western Sicily (Italy). *Endocrine*, 43, 108-111.
- GRECO, L., PERCOPO, S., CLOT, F., BOUGUERRA, F., BABRON, M. C., ELIAOU, J. F., FRANZESE, C., TRONCONE, R. & CLERGET-DARPOUX, F. 1998. Lack of correlation between genotype and phenotype in celiac disease. *J Pediatr Gastroenterol Nutr*, 26, 286-90.
- GRECO, L., ROMINO, R., COTO, I., DI COSMO, N., PERCOPO, S., MAGLIO, M., PAPARO, F., GASPERI, V., LIMONGELLI, M. G., COTICHINI, R., D'AGATE, C., TINTO, N., SACCHETTI, L., TOSI, R. & STAZI, M. A. 2002. The first large population based twin study of coeliac disease. *Gut*, 50, 624-8.
- GREEN, P. H. & JABRI, B. 2003. Coeliac disease. *Lancet*, 362, 383-91.
- GUDJONSDOTTIR, A. H., NILSSON, S., NALUAI, A. T., EK, J., AMUNDSEN, S. S., WAHLSTROM, J. & ASCHER, H. 2009. Association between genotypes and phenotypes in coeliac disease. *J Pediatr Gastroenterol Nutr*, 49, 165-9.
- GUNESH, S., THOMAS, G. A., WILLIAMS, G. T., ROBERTS, A. & HAWTHORNE, A. B. 2008. The incidence of Crohn's disease in Cardiff over the last 75 years: an update for 1996-2005. *Aliment Pharmacol Ther*, 27, 211-9.
- GUTIERREZ-ACHURY, J., DE ALMEIDA, R. C. & WIJMENGA, C. 2011. Shared genetics in coeliac disease and other immune-mediated diseases. *Journal of Internal Medicine*, 269, 591-603.
- HADITHI, M., DE BOER, H., MEIJER, J. W., WILLEKENS, F., KERCKHAERT, J. A., HEIJMANS, R., PENA, A. S., STEHOUWER, C. D. & MULDER, C. J. 2007a. Coeliac disease in Dutch patients with Hashimoto's thyroiditis and vice versa. *World J Gastroenterol*, 13, 1715-22.
- HADITHI, M., VON BLOMBERG, B. M., CRUSIUS, J. B., BLOEMENA, E., KOSTENSE, P. J., MEIJER, J. W., MULDER, C. J., STEHOUWER, C. D. & PENA, A. S. 2007b. Accuracy of



- serologic tests and HLA-DQ typing for diagnosing celiac disease. *Ann Intern Med*, 147, 294-302.
- HADJIVASSILIOU, M., AESCHLIMANN, P., SANDERS, D. S., MAKI, M., KAUKINEN, K., GRUNEWALD, R. A., BANDMANN, O., WOODROOFE, N., HADDOCK, G. & AESCHLIMANN, D. P. 2013. Transglutaminase 6 antibodies in the diagnosis of gluten ataxia. *Neurology*, 80, 1740-5.
- HADJIVASSILIOU, M., AESCHLIMANN, P., STRIGUN, A., SANDERS, D. S., WOODROOFE, N. & AESCHLIMANN, D. 2008. Autoantibodies in gluten ataxia recognize a novel neuronal transglutaminase. *Ann Neurol*, 64, 332-43.
- HADJIVASSILIOU, M., GRUNEWALD, R. A. & DAVIES-JONES, G. A. 2002. Gluten sensitivity as a neurological illness. *J Neurol Neurosurg Psychiatry*, 72, 560-3.
- HADJIVASSILIOU, M., MAKI, M., SANDERS, D. S., WILLIAMSON, C. A., GRUNEWALD, R. A., WOODROOFE, N. M. & KORPONAY-SZABO, I. R. 2006. Autoantibody targeting of brain and intestinal transglutaminase in gluten ataxia. *Neurology*, 66, 373-7.
- HADJIVASSILIOU, M., RAO, D. G., GRUNEWALD, R. A., AESCHLIMANN, D. P., SARRIGIANNIS, P. G., HOGGARD, N., AESCHLIMANN, P., MOONEY, P. D. & SANDERS, D. S. 2016. Neurological Dysfunction in Coeliac Disease and Non-Coeliac Gluten Sensitivity. *Am J Gastroenterol*, 111, 561-7.
- HADLEY, D., HAGOPIAN, W., LIU, E., SHE, J. X., SIMELL, O., AKOLKAR, B., ZIEGLER, A. G., REWERS, M., KRISCHER, J. P., CHEN, W. M., ONENGUT-GUMUSCU, S., BUGAWAN, T. L., RICH, S. S., ERLICH, H. & AGARDH, D. 2015. HLA-DPB1\*04:01 Protects Genetically Susceptible Children from Celiac Disease Autoimmunity in the TEDDY Study. *Am J Gastroenterol*, 110, 915-20.
- HAERE, P., HOIE, O., SCHULZ, T., SCHONHARDT, I., RAKI, M. & LUNDIN, K. E. 2016. Long-term mucosal recovery and healing in celiac disease is the rule - not the exception. *Scand J Gastroenterol*, 51, 1439-1446.
- HAFSTROM, I., RINGERTZ, B., SPANGBERG, A., VON ZWEIGBERGK, L., BRANNEMARK, S., NYLANDER, I., RONNELID, J., LAASONEN, L. & KLARESKOG, L. 2001. A vegan diet free of gluten improves the signs and symptoms of rheumatoid arthritis: the effects on arthritis correlate with a reduction in antibodies to food antigens. *Rheumatology (Oxford)*, 40, 1175-9.
- HAINES, M. L., ANDERSON, R. P. & GIBSON, P. R. 2008. Systematic review: The evidence base for long-term management of coeliac disease. *Aliment Pharmacol Ther*, 28, 1042-66.
- HAKANEN, M., LUOTOLA, K., SALMI, J., LAIPPALA, P., KAUKINEN, K. & COLLIN, P. 2001. Clinical and subclinical autoimmune thyroid disease in adult celiac disease. *Dig Dis Sci*, 46, 2631-5.
- HALFDANARSON, T. R., LITZOW, M. R. & MURRAY, J. A. 2007. Hematologic manifestations of celiac disease. *Blood*, 109, 412-21.
- HALL, N. J., RUBIN, G. & CHARNOCK, A. 2009. Systematic review: adherence to a gluten-free diet in adult patients with coeliac disease. *Aliment Pharmacol Ther*, 30, 315-30.
- HANKEY, G. L. & HOLMES, G. K. 1994. Coeliac disease in the elderly. *Gut*, 35, 65-7.
- HANSEN, C. H., KRYCH, L., BUSCHARD, K., METZDORFF, S. B., NELLEMAN, C., HANSEN, L. H., NIELSEN, D. S., FROKIAER, H., SKOV, S. & HANSEN, A. K. 2014. A maternal gluten-free diet reduces inflammation and diabetes incidence in the offspring of NOD mice. *Diabetes*, 63, 2821-32.
- HARJUTSALO, V., SJOBERG, L. & TUOMILEHTO, J. 2008. Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet*, 371, 1777-82.
- HARLEY, I. T. & KARP, C. L. 2012. Obesity and the gut microbiome: Striving for causality. *Mol Metab*, 1, 21-31.
- HARPER, J. W., HOLLERAN, S. F., RAMAKRISHNAN, R., BHAGAT, G. & GREEN, P. H. 2007. Anemia in celiac disease is multifactorial in etiology. *Am J Hematol*, 82, 996-1000.
- HARPSOE, M. C., BASIT, S., ANDERSSON, M., NIELSEN, N. M., FRISCH, M., WOHLFAHRT, J., NOHR, E. A., LINNEBERG, A. & JESS, T. 2014. Body mass index and risk of

- autoimmune diseases: a study within the Danish National Birth Cohort. *Int J Epidemiol*, 43, 843-55.
- HEIKKILA, K., HELIOVAARA, M., IMPIVAARA, O., KROGER, H., KNEKT, P., RISSANEN, H., MAKI, M. & KAUKINEN, K. 2015a. Celiac disease autoimmunity and hip fracture risk: findings from a prospective cohort study. *J Bone Miner Res*, 30, 630-6.
- HEIKKILA, K., PEARCE, J., MAKI, M. & KAUKINEN, K. 2015b. Celiac disease and bone fractures: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 100, 25-34.
- HERFARTH, H. H., MARTIN, C. F., SANDLER, R. S., KAPPELMAN, M. D. & LONG, M. D. 2014. Prevalence of a gluten-free diet and improvement of clinical symptoms in patients with inflammatory bowel diseases. *Inflamm Bowel Dis*, 20, 1194-7.
- HERNLUND, E., SVEDBOM, A., IVERGARD, M., COMPSTON, J., COOPER, C., STENMARK, J., MCCLOSKEY, E. V., JONSSON, B. & KANIS, J. A. 2013. Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). *Arch Osteoporos*, 8, 136.
- HILL, P. G. & HOLMES, G. K. 2008. Coeliac disease: a biopsy is not always necessary for diagnosis. *Aliment Pharmacol Ther*, 27, 572-7.
- HISCHENHUBER, C., CREVEL, R., JARRY, B., MAKI, M., MONERET-VAUTRIN, D. A., ROMANO, A., TRONCONE, R. & WARD, R. 2006. Review article: safe amounts of gluten for patients with wheat allergy or coeliac disease. *Aliment Pharmacol Ther*, 23, 559-75.
- HONEYMAN, M. C., COULSON, B. S., STONE, N. L., GELLERT, S. A., GOLDWATER, P. N., STEELE, C. E., COUPER, J. J., TAIT, B. D., COLMAN, P. G. & HARRISON, L. C. 2000. Association between rotavirus infection and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes. *Diabetes*, 49, 1319-24.
- HOOKWAY, C., BUCKNER, S., CROSLAND, P. & LONGSON, D. 2015. Irritable bowel syndrome in adults in primary care: summary of updated NICE guidance. *Bmj*, 350, h701.
- HOPPER, A. D., HADJIVASSILIOU, M., HURLSTONE, D. P., LOBO, A. J., MCALINDON, M. E., EGNER, W., WILD, G. & SANDERS, D. S. 2008. What is the role of serologic testing in celiac disease? A prospective, biopsy-confirmed study with economic analysis. *Clin Gastroenterol Hepatol*, 6, 314-20.
- HU, C., WONG, F. S. & WEN, L. 2015. Type 1 diabetes and gut microbiota: Friend or foe? *Pharmacol Res*, 98, 9-15.
- HUMBERT, P., BIDET, A., TREFFEL, P., DROBACHEFF, C. & AGACHE, P. 1991. Intestinal permeability in patients with psoriasis. *J Dermatol Sci*, 2, 324-6.
- HUSBY, S., KOLETZKO, S., KORPONAY-SZABO, I. R., MEARIN, M. L., PHILLIPS, A., SHAMIR, R., TRONCONE, R., GIERSEPPEN, K., BRANSKI, D., CATASSI, C., LELGEMAN, M., MAKI, M., RIBES-KONINCKX, C., VENTURA, A. & ZIMMER, K. P. 2012. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr*, 54, 136-60.
- HWANG, E., MCBRIDE, R., NEUGUT, A. I. & GREEN, P. H. 2008. Sarcoidosis in patients with celiac disease. *Dig Dis Sci*, 53, 977-81.
- ICEN, M., CROWSON, C. S., MCEVOY, M. T., DANN, F. J., GABRIEL, S. E. & MARADIT KREMERS, H. 2009. Trends in incidence of adult-onset psoriasis over three decades: a population-based study. *J Am Acad Dermatol*, 60, 394-401.
- IERARDI, E., LOSURDO, G., PISCITELLI, D., GIORGIO, F., SORRENTINO, C., PRINCIPI, M., MONTENEGRO, L., AMORUSO, A. & DI LEO, A. 2015. Seronegative celiac disease: where is the specific setting? *Gastroenterol Hepatol Bed Bench*, 8, 110-6.
- IMPERATORE, N., RISPO, A., CAPONE, P., DONETTO, S., DE PALMA, G. D., GERBINO, N., REA, M., CAPORASO, N. & TORTORA, R. 2016. Gluten-free diet does not influence the occurrence and the Th1/Th17-Th2 nature of immune-mediated diseases in patients with coeliac disease. *Dig Liver Dis*, 48, 740-4.

- INNALA, L., BERGLIN, E., MOLLER, B., LJUNG, L., SMEDBY, T., SODERGREN, A., MAGNUSSON, S., RANTAPAA-DAHLQVIST, S. & WALLBERG-JONSSON, S. 2014. Age at onset determines severity and choice of treatment in early rheumatoid arthritis: a prospective study. *Arthritis Res Ther*, 16, R94.
- IVARSSON, A. 2005. The Swedish epidemic of coeliac disease explored using an epidemiological approach--some lessons to be learnt. *Best Pract Res Clin Gastroenterol*, 19, 425-40.
- IVARSSON, A., PERSSON, L. A., NYSTROM, L., ASCHER, H., CAVELL, B., DANIELSSON, L., DANNAEUS, A., LINDBERG, T., LINDQUIST, B., STENHAMMAR, L. & HERNELL, O. 2000. Epidemic of coeliac disease in Swedish children. *Acta Paediatr*, 89, 165-71.
- JACKSON, J. R., EATON, W. W., CASCELLA, N. G., FASANO, A. & KELLY, D. L. 2012. Neurologic and psychiatric manifestations of celiac disease and gluten sensitivity. *Psychiatr Q*, 83, 91-102.
- JACOBSON, D. L., GANGE, S. J., ROSE, N. R. & GRAHAM, N. M. 1997. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. *Clin Immunol Immunopathol*, 84, 223-43.
- JAGER, S., STANGE, E. F. & WEHKAMP, J. 2013. Inflammatory bowel disease: an impaired barrier disease. *Langenbecks Arch Surg*, 398, 1-12.
- JANSSEN, G., CHRISTIS, C., KOOY-WINKELAAR, Y., EDENS, L., SMITH, D., VAN VEELLEN, P. & KONING, F. 2015. Ineffective degradation of immunogenic gluten epitopes by currently available digestive enzyme supplements. *PLoS One*, 10, e0128065.
- JANSSENS, A. C. 2008. Is the time right for translation research in genomics? *Eur J Epidemiol*, 23, 707-10.
- JOHN, J. A. & DRAPER, N. R. 1980. An Alternative Family of Transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29, 190-197.
- JOSTINS, L., RIPKE, S., WEERSMA, R. K., DUERR, R. H., MCGOVERN, D. P., HUI, K. Y., LEE, J. C., SCHUMM, L. P., SHARMA, Y., ANDERSON, C. A., ESSERS, J., MITROVIC, M., NING, K., CLEYNEN, I., THEATRE, E., SPAIN, S. L., RAYCHAUDHURI, S., GOYETTE, P., WEI, Z., ABRAHAM, C., ACHKAR, J. P., AHMAD, T., AMININEJAD, L., ANANTHAKRISHNAN, A. N., ANDERSEN, V., ANDREWS, J. M., BAIDOO, L., BALSCHUN, T., BAMPTON, P. A., BITTON, A., BOUCHER, G., BRAND, S., BUNING, C., COHAIN, A., CICHON, S., D'AMATO, M., DE JONG, D., DEVANEY, K. L., DUBINSKY, M., EDWARDS, C., ELLINGHAUS, D., FERGUSON, L. R., FRANCHIMONT, D., FRANSEN, K., GEARRY, R., GEORGES, M., GIEGER, C., GLAS, J., HARITUNIAN, T., HART, A., HAWKEY, C., HEDL, M., HU, X., KARLSEN, T. H., KUPCINSKAS, L., KUGATHASAN, S., LATIANO, A., LAUKENS, D., LAWRENCE, I. C., LEES, C. W., LOUIS, E., MAHY, G., MANSFIELD, J., MORGAN, A. R., MOWAT, C., NEWMAN, W., PALMIERI, O., PONSIOEN, C. Y., POTOCHNIK, U., PRESCOTT, N. J., REGUEIRO, M., ROTTER, J. I., RUSSELL, R. K., SANDERSON, J. D., SANS, M., SATSANGI, J., SCHREIBER, S., SIMMS, L. A., SVENTORAITYTE, J., TARGAN, S. R., TAYLOR, K. D., TREMELLING, M., VERSPAGET, H. W., DE VOS, M., WIJMEGA, C., WILSON, D. C., WINKELMANN, J., XAVIER, R. J., ZEISSIG, S., ZHANG, B., ZHANG, C. K., ZHAO, H., SILVERBERG, M. S., ANNESE, V., HAKONARSON, H., BRANT, S. R., RADFORD-SMITH, G., MATHEW, C. G., RIOUX, J. D., SCHADT, E. E., et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 491, 119-24.
- JOZKOW, P., LWOW, F., SLOWINSKA-LISOWSKA, M. & MEDRAS, M. 2017. Trends in the prevalence of autoimmune thyroiditis in the leading private health-care provider in Poland. *Adv Clin Exp Med*, 26, 497-503.
- JUNKER, Y., ZEISSIG, S., KIM, S. J., BARISANI, D., WIESER, H., LEFFLER, D. A., ZEVALLOS, V., LIBERMANN, T. A., DILLON, S., FREITAG, T. L., KELLY, C. P. & SCHUPPAN, D. 2012. Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J Exp Med*, 209, 2395-408.
- JURAN, B. D., HIRSCHFELD, G. M., INVERNIZZI, P., ATKINSON, E. J., LI, Y., XIE, G., KOSOY, R., RANSOM, M., SUN, Y., BIANCHI, I., SCHLICHT, E. M., LLEO, A., COLTESCU, C.,

- BERNUZZI, F., PODDA, M., LAMMERT, C., SHIGETA, R., CHAN, L. L., BALSCHUN, T., MARCONI, M., CUSI, D., HEATHCOTE, E. J., MASON, A. L., MYERS, R. P., MILKIEWICZ, P., ODIN, J. A., LUKETIC, V. A., BACON, B. R., BODENHEIMER, H. C., JR., LIAKINA, V., VINCENT, C., LEVY, C., FRANKE, A., GREGERSEN, P. K., BOSSA, F., GERSHWIN, M. E., DEANDRADE, M., AMOS, C. I., LAZARIDIS, K. N., SELDIN, M. F. & SIMINOVITCH, K. A. 2012. ImmunoChip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum Mol Genet*, 21, 5209-21.
- KABAKCHIEV, B. & SILVERBERG, M. S. 2013. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology*, 144, 1488-96, 1496.e1-3.
- KABBANI, T. A., GOLDBERG, A., KELLY, C. P., PALLAV, K., TARIQ, S., PEER, A., HANSEN, J., DENNIS, M. & LEFFLER, D. A. 2012. Body mass index and the risk of obesity in coeliac disease treated with the gluten-free diet. *Aliment Pharmacol Ther*, 35, 723-9.
- KAGNOFF, M. F. 2007. Celiac disease: pathogenesis of a model immunogenetic disease. *J Clin Invest*, 117, 41-9.
- KANG, J. Y., KANG, A. H., GREEN, A., GWEE, K. A. & HO, K. Y. 2013. Systematic review: worldwide variation in the frequency of coeliac disease and changes over time. *Aliment Pharmacol Ther*, 38, 226-45.
- KAPLAN, G. G. & NG, S. C. 2016. Globalisation of inflammatory bowel disease: perspectives from the evolution of inflammatory bowel disease in the UK and China. *Lancet Gastroenterol Hepatol*, 1, 307-316.
- KARELL, K., LOUKA, A. S., MOODIE, S. J., ASCHER, H., CLOT, F., GRECO, L., CICLITIRA, P. J., SOLLID, L. M. & PARTANEN, J. 2003. HLA types in celiac 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.
- KASARDA, D. D. 2013. Can an increase in celiac disease be attributed to an increase in the gluten content of wheat as a consequence of wheat breeding? *J Agric Food Chem*, 61, 1155-9.
- KAUKINEN, K., PERAAHO, M., LINDFORS, K., PARTANEN, J., WOOLLEY, N., PIKKARAINEN, P., KARVONEN, A. L., LAASANEN, T., SIEVANEN, H., MAKI, M. & COLLIN, P. 2007. Persistent small bowel mucosal villous atrophy without symptoms in coeliac disease. *Aliment Pharmacol Ther*, 25, 1237-45.
- KAUKINEN, K., SULKANEN, S., MAKI, M. & COLLIN, P. 2002. IgA-class transglutaminase antibodies in evaluating the efficacy of gluten-free diet in coeliac disease. *Eur J Gastroenterol Hepatol*, 14, 311-5.
- KELLY, C. P., GREEN, P. H., MURRAY, J. A., DIMARINO, A., COLATRELLA, A., LEFFLER, D. A., ALEXANDER, T., ARSENESCU, R., LEON, F., JIANG, J. G., ARTERBURN, L. A., PATERSON, B. M. & FEDORAK, R. N. 2013. Larazotide acetate in patients with coeliac disease undergoing a gluten challenge: a randomised placebo-controlled study. *Aliment Pharmacol Ther*, 37, 252-62.
- KEMPPAINEN, K. M., LYNCH, K. F., LIU, E., LONNROT, M., SIMELL, V., BRIESE, T., KOLETZKO, S., HAGOPIAN, W., REWERS, M., SHE, J. X., SIMELL, O., TOPPARI, J., ZIEGLER, A. G., AKOLKAR, B., KRISCHER, J. P., LERNMARK, A., HYOTY, H., TRIPLETT, E. W. & AGARDH, D. 2017. Factors That Increase Risk of Celiac Disease Autoimmunity After a Gastrointestinal Infection in Early Life. *Clin Gastroenterol Hepatol*, 15, 694-702.e5.
- KEMPPAINEN, T., KROGER, H., JANATUINEN, E., ARNALA, I., KOSMA, V. M., PIKKARAINEN, P., JULKUNEN, R., JURVELIN, J., ALHAVA, E. & UUSITUPA, M. 1999. Osteoporosis in adult patients with celiac disease. *Bone*, 24, 249-55.
- KHALEGHI, S., JU, J. M., LAMBA, A. & MURRAY, J. A. 2016. The potential utility of tight junction regulation in celiac disease: focus on larazotide acetate. *Therap Adv Gastroenterol*, 9, 37-49.

- KIESLICH, M., ERRAZURIZ, G., POSSELT, H. G., MOELLER-HARTMANN, W., ZANELLA, F. & BOEHLES, H. 2001. Brain white-matter lesions in celiac disease: a prospective study of 75 diet-treated patients. *Pediatrics*, 108, E21.
- KIMMONS, J. E., BLANCK, H. M., TOHILL, B. C., ZHANG, J. & KHAN, L. K. 2006. Associations between body mass index and the prevalence of low micronutrient levels among US adults. *MedGenMed*, 8, 59.
- KIVELA, L., KAUKINEN, K., LAHDEAHO, M. L., HUHTALA, H., ASHORN, M., RUUSKA, T., HILTUNEN, P., VISAKORPI, J., MAKI, M. & KURPPA, K. 2015. Presentation of Celiac Disease in Finnish Children Is No Longer Changing: A 50-Year Perspective. *J Pediatr*.
- KOCSIS, D., MIHELLER, P., LORINCZY, K., HERSZENYI, L., TULASSAY, Z., RACZ, K. & JUHASZ, M. 2013. Coeliac disease in a 15-year period of observation (1997 and 2011) in a Hungarian referral centre. *Eur J Intern Med*, 24, 461-7.
- KOCSIS, D., TOTH, Z., CSONTOS, A. A., MIHELLER, P., PAK, P., HERSZENYI, L., TOTH, M., TULASSAY, Z. & JUHASZ, M. 2015. Prevalence of inflammatory bowel disease among coeliac disease patients in a Hungarian coeliac centre. *BMC Gastroenterol*, 15, 141.
- KOLETZKO, S., LEE, H. S., BEYERLEIN, A., ARONSSON, C. A., HUMMEL, M., LIU, E., SIMELL, V., KURPPA, K., LERNMARK, A., HAGOPIAN, W., REWERS, M., SHE, J. X., SIMELL, O., TOPPARI, J., ZIEGLER, A. G., KRISCHER, J. & AGARDH, D. 2017. Caesarean Section on The Risk of Celiac Disease in the Offspring: The Teddy Study. *J Pediatr Gastroenterol Nutr*.
- KONDRASHOVA, A., MUSTALAHTI, K., KAUKINEN, K., VISKARI, H., VOLODICHEVA, V., HAAPALA, A. M., ILONEN, J., KNIP, M., MAKI, M. & HYOTY, H. 2008. Lower economic status and inferior hygienic environment may protect against celiac disease. *Ann Med*, 40, 223-31.
- KONING, F., THOMAS, R., ROSSJOHN, J. & TOES, R. E. 2015. Coeliac disease and rheumatoid arthritis: similar mechanisms, different antigens. *Nat Rev Rheumatol*, 11, 450-61.
- KORPONAY-SZABO, I., KOVACS, J., LORINCZ, M., TOROK, E. & GORACZ, G. 1998. Families with multiple cases of gluten-sensitive enteropathy. *Z Gastroenterol*, 36, 553-8.
- KORPONAY-SZABO, I. R., HALTTUNEN, T., SZALAI, Z., LAURILA, K., KIRALY, R., KOVACS, J. B., FESUS, L. & MAKI, M. 2004. In vivo targeting of intestinal and extraintestinal transglutaminase 2 by coeliac autoantibodies. *Gut*, 53, 641-8.
- KRISHNAREDDY, S., STIER, K., RECANATI, M., LEBWOHL, B. & GREEN, P. H. 2017. Commercially available glutenases: a potential hazard in coeliac disease. *Therap Adv Gastroenterol*, 10, 473-481.
- KRISTENSEN, K. & HENRIKSEN, L. 2016. Cesarean section and disease associated with immune function. *J Allergy Clin Immunol*, 137, 587-90.
- KUPFER, S. S. & JABRI, B. 2012. Pathophysiology of celiac disease. *Gastrointest Endosc Clin N Am*, 22, 639-60.
- KURADA, S., YADAV, A. & LEFFLER, D. A. 2016. Current and novel therapeutic strategies in celiac disease. *Expert Rev Clin Pharmacol*, 9, 1211-23.
- KURPPA, K., HIETIKKO, M., SULIC, A. M., KAUKINEN, K. & LINDFORS, K. 2014. Current status of drugs in development for celiac disease. *Expert Opin Investig Drugs*, 23, 1079-91.
- KURPPA, K., LINDFORS, K., COLLIN, P., SAAVALAINEN, P., PARTANEN, J., HAIMILA, K., HUHTALA, H., LAURILA, K., MAKI, M. & KAUKINEN, K. 2011. Antibodies against deamidated gliadin peptides in early-stage celiac disease. *J Clin Gastroenterol*, 45, 673-8.
- KURPPA, K., TAAVELA, J., SAAVALAINEN, P., KAUKINEN, K. & LINDFORS, K. 2016. Novel diagnostic techniques for celiac disease. *Expert Rev Gastroenterol Hepatol*, 10, 795-805.

- LAHDEAHO, M. L., MAKI, M., LAURILA, K., HUHTALA, H. & KAUKINEN, K. 2011. Small-bowel mucosal changes and antibody responses after low- and moderate-dose gluten challenge in celiac disease. *BMC Gastroenterol*, 11, 129.
- LAIIVORANTA-NYMAN, S., MOTTONEN, T., HERMANN, R., TUOKKO, J., LUUKKAINEN, R., HAKALA, M., HANNONEN, P., KORPELA, M., YLI-KERTTULA, U., TOIVANEN, A. & ILONEN, J. 2004. HLA-DR-DQ haplotypes and genotypes in Finnish patients with rheumatoid arthritis. *Ann Rheum Dis*, 63, 1406-12.
- LAMMERS, K. M., LU, R., BROWNLEY, J., LU, B., GERARD, C., THOMAS, K., RALLABHANDI, P., SHEA-DONOHUE, T., TAMIZ, A., ALKAN, S., NETZEL-ARNETT, S., ANTALIS, T., VOGEL, S. N. & FASANO, A. 2008. Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3. *Gastroenterology*, 135, 194-204.e3.
- LARUSSA, T., SURACI, E., NAZIONALE, I., ABENAVOLI, L., IMENEO, M. & LUZZA, F. 2012. Bone mineralization in celiac disease. *Gastroenterol Res Pract*, 2012, 198025.
- LAURET, E. & RODRIGO, L. 2013. Celiac disease and autoimmune-associated conditions. *Biomed Res Int*, 2013, 127589.
- LEBWOHL, B., BLASER, M. J., LUDVIGSSON, J. F., GREEN, P. H., RUNDLE, A., SONNENBERG, A. & GENTA, R. M. 2013. Decreased risk of celiac disease in patients with *Helicobacter pylori* colonization. *Am J Epidemiol*, 178, 1721-30.
- LEBWOHL, B., BLASER, M. J., LUDVIGSSON, J. F., GREEN, P. H., RUNDLE, A., SONNENBERG, A. & GENTA, R. M. 2014a. Re: "Decreased risk of celiac disease in patients with *Helicobacter pylori* colonization". The authors reply. *Am J Epidemiol*, 179, 1275-6.
- LEBWOHL, B., LUDVIGSSON, J. F. & GREEN, P. H. 2014b. The unfolding story of celiac disease risk factors. *Clin Gastroenterol Hepatol*, 12, 632-5.
- LEE, H. J., ANDERSON, Z. & RYU, D. 2014. Gluten contamination in foods labeled as "gluten free" in the United States. *J Food Prot*, 77, 1830-3.
- LEE, S. H., HAROLD, D., NYHOLT, D. R., GODDARD, M. E., ZONDERVAN, K. T., WILLIAMS, J., MONTGOMERY, G. W., WRAY, N. R. & VISSCHER, P. M. 2013. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet*, 22, 832-41.
- LEEDS, J. S., HOROLDT, B. S., SIDHU, R., HOPPER, A. D., ROBINSON, K., TOULSON, B., DIXON, L., LOBO, A. J., MCALINDON, M. E., HURLSTONE, D. P. & SANDERS, D. S. 2007. Is there an association between coeliac disease and inflammatory bowel diseases? A study of relative prevalence in comparison with population controls. *Scand J Gastroenterol*, 42, 1214-20.
- LEFFLER, D. A., DENNIS, M., EDWARDS GEORGE, J. B., JAMMA, S., MAGGE, S., COOK, E. F., SCHUPPAN, D. & KELLY, C. P. 2009. A simple validated gluten-free diet adherence survey for adults with celiac disease. *Clin Gastroenterol Hepatol*, 7, 530-6, 536.e1-2.
- LEFFLER, D. A., DENNIS, M., HYETT, B., KELLY, E., SCHUPPAN, D. & KELLY, C. P. 2007a. Etiologies and predictors of diagnosis in nonresponsive celiac disease. *Clin Gastroenterol Hepatol*, 5, 445-50.
- LEFFLER, D. A., EDWARDS GEORGE, J. B., DENNIS, M., COOK, E. F., SCHUPPAN, D. & KELLY, C. P. 2007b. A prospective comparative study of five measures of gluten-free diet adherence in adults with coeliac disease. *Aliment Pharmacol Ther*, 26, 1227-35.
- LEFFLER, D. A., GREEN, P. H. & FASANO, A. 2015a. Extraintestinal manifestations of coeliac disease. *Nat Rev Gastroenterol Hepatol*, 12, 561-71.
- LEFFLER, D. A., KELLY, C. P., GREEN, P. H., FEDORAK, R. N., DIMARINO, A., PERROW, W., RASMUSSEN, H., WANG, C., BERCIK, P., BACHIR, N. M. & MURRAY, J. A. 2015b. Larazotide acetate for persistent symptoms of celiac disease despite a gluten-free diet: a randomized controlled trial. *Gastroenterology*, 148, 1311-9.e6.
- LEFFLER, D. A. & SCHUPPAN, D. 2010. Update on serologic testing in celiac disease. *Am J Gastroenterol*, 105, 2520-4.
- LEONARD, M. M. & FASANO, A. 2016. The microbiome as a possible target to prevent celiac disease. *Expert Rev Gastroenterol Hepatol*, 10, 555-6.

- LERNER, A. & MATTHIAS, T. 2015a. Changes in intestinal tight junction permeability associated with industrial food additives explain the rising incidence of autoimmune disease. *Autoimmun Rev*, 14, 479-89.
- LERNER, A. & MATTHIAS, T. 2015b. Rheumatoid arthritis-celiac disease relationship: joints get that gut feeling. *Autoimmun Rev*, 14, 1038-47.
- LEVY, M., KOLODZIEJCZYK, A. A., THAISS, C. A. & ELINAV, E. 2017. Dysbiosis and the immune system. *Nat Rev Immunol*, 17, 219-232.
- LINDFORS, K., KOSKINEN, O., KURPPA, K., LAURILA, K., COLLIN, P., HAIMILA, K., PARTANEN, J., SAAVALAINEN, P., MAKI, M. & KAUKINEN, K. 2011. Serodiagnostic assays for celiac disease based on the open or closed conformation of the autoantigen, transglutaminase 2. *J Clin Immunol*, 31, 436-42.
- LINDFORS, K., LAHDEAHO, M. L., KALLIOKOSKI, S., KURPPA, K., COLLIN, P., MAKI, M. & KAUKINEN, K. 2012. Future treatment strategies for celiac disease. *Expert Opin Ther Targets*, 16, 665-75.
- LIONETTI, E., CASTELLANETA, S., FRANCAVILLA, R., PULVIRENTI, A., TONUTTI, E., AMARRI, S., BARBATO, M., BARBERA, C., BARERA, G., BELLANTONI, A., CASTELLANO, E., GUARISO, G., LIMONGELLI, M. G., PELLEGRINO, S., POLLONI, C., UGHI, C., ZUIN, G., FASANO, A. & CATASSI, C. 2014. Introduction of gluten, HLA status, and the risk of celiac disease in children. *N Engl J Med*, 371, 1295-303.
- LIU, E., LEE, H. S., ARONSSON, C. A., HAGOPIAN, W. A., KOLETZKO, S., REWERS, M. J., EISENBARTH, G. S., BINGLEY, P. J., BONIFACIO, E., SIMELL, V. & AGARDH, D. 2014. Risk of paediatric celiac disease according to HLA haplotype and country. *N Engl J Med*, 371, 42-9.
- LIU, E., LI, M., EMERY, L., TAKI, I., BARRIGA, K., TIBERTI, C., EISENBARTH, G. S., REWERS, M. J. & HOFFENBERG, E. J. 2007. Natural history of antibodies to deamidated gliadin peptides and transglutaminase in early childhood celiac disease. *J Pediatr Gastroenterol Nutr*, 45, 293-300.
- LOHI, S., MUSTALAHTI, K., KAUKINEN, K., LAURILA, K., COLLIN, P., RISSANEN, H., LOHI, O., BRAVI, E., GASPARIN, M., REUNANEN, A. & MAKI, M. 2007. Increasing prevalence of coeliac disease over time. *Aliment Pharmacol Ther*, 26, 1217-25.
- LUDVIGSSON, J. F., BAI, J. C., BIAGI, F., CARD, T. R., CIACCI, C., CICLITIRA, P. J., GREEN, P. H., HADJIVASSILIOU, M., HOLDOWAY, A., VAN HEEL, D. A., KAUKINEN, K., LEFFLER, D. A., LEONARD, J. N., LUNDIN, K. E., MCGOUGH, N., DAVIDSON, M., MURRAY, J. A., SWIFT, G. L., WALKER, M. M., ZINGONE, F. & SANDERS, D. S. 2014a. Diagnosis and management of adult coeliac disease: guidelines from the British Society of Gastroenterology. *Gut*, 63, 1210-28.
- LUDVIGSSON JF, CIACCI C, GREEN PH, KAUKINEN K, KORPONAY-SZABO IR, KURPPA K, ET AL. Outcome measures in coeliac disease trials: the Tampere recommendations. *Gut*. 2018;67(8):1410-24.
- LUDVIGSSON, J. F., BAI, J. C., BIAGI, F., CARD, T. R., CIACCI, C., CICLITIRA, P. J., GREEN, P. H., HADJIVASSILIOU, M., HOLDOWAY, A., VAN HEEL, D. A., KAUKINEN, K., LEFFLER, D. A., LEONARD, J. N., LUNDIN, K. E., MCGOUGH, N., DAVIDSON, M., MURRAY, J. A., SWIFT, G. L., WALKER, M. M., ZINGONE, F. & SANDERS, D. S. 2014b. Diagnosis and management of adult coeliac disease: Guidelines from the British society of gastroenterology. *Gut*, 63, 1210-1228.
- LUDVIGSSON, J. F., CARD, T. R., KAUKINEN, K., BAI, J., ZINGONE, F., SANDERS, D. S. & MURRAY, J. A. 2015. Screening for celiac disease in the general population and in high-risk groups. *United European Gastroenterol J*, 3, 106-20.
- LUDVIGSSON, J. F., LEFFLER, D. A., BAI, J. C., BIAGI, F., FASANO, A., GREEN, P. H., HADJIVASSILIOU, M., KAUKINEN, K., KELLY, C. P., LEONARD, J. N., LUNDIN, K. E., MURRAY, J. A., SANDERS, D. S., WALKER, M. M., ZINGONE, F. & CIACCI, C. 2013a. The Oslo definitions for coeliac disease and related terms. *Gut*, 62, 43-52.
- LUDVIGSSON, J. F., LINDELOF, B., ZINGONE, F. & CIACCI, C. 2011. Psoriasis in a nationwide cohort study of patients with celiac disease. *J Invest Dermatol*, 131, 2010-6.

- LUDVIGSSON, J. F., LUDVIGSSON, J., EKBOM, A. & MONTGOMERY, S. M. 2006. Celiac disease and risk of subsequent type 1 diabetes: a general population cohort study of children and adolescents. *Diabetes Care*, 29, 2483-8.
- LUDVIGSSON, J. F., RUBIO-TAPIA, A., CHOWDHARY, V., MURRAY, J. A. & SIMARD, J. F. 2012. Increased Risk of Systemic Lupus Erythematosus in 29,000 Patients with Biopsy-verified Celiac Disease. *Journal of Rheumatology*, 39, 1964-1970.
- LUDVIGSSON, J. F., RUBIO-TAPIA, A., VAN DYKE, C. T., MELTON, L. J., 3RD, ZINSMEISTER, A. R., LAHR, B. D. & MURRAY, J. A. 2013b. Increasing incidence of celiac disease in a North American population. *Am J Gastroenterol*, 108, 818-24.
- LUDVIGSSON, J. F., WAHLSTROM, J., GRUNEWALD, J., EKBOM, A. & MONTGOMERY, S. M. 2007. Coeliac disease and risk of sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis*, 24, 121-6.
- LUNDIN, K. E. & WIJMENGA, C. 2015. Coeliac disease and autoimmune disease-genetic overlap and screening. *Nat Rev Gastroenterol Hepatol*, 12, 507-15.
- MAAHS, D. M., WEST, N. A., LAWRENCE, J. M. & MAYER-DAVIS, E. J. 2010. Epidemiology of type 1 diabetes. *Endocrinol Metab Clin North Am*, 39, 481-97.
- MAKI, M. 2014. Celiac disease treatment: gluten-free diet and beyond. *J Pediatr Gastroenterol Nutr*, 59 Suppl 1, S15-7.
- MALAMUT, G., MERESSE, B., CELLIER, C. & CERF-BENSUSSAN, N. 2012a. Refractory celiac disease: from bench to bedside. *Semin Immunopathol*, 34, 601-13.
- MALAMUT, G., MURRAY, J. A. & CELLIER, C. 2012b. Refractory celiac disease. *Gastrointest Endosc Clin N Am*, 22, 759-72.
- MANKERTZ, J. & SCHULZKE, J. D. 2007. Altered permeability in inflammatory bowel disease: pathophysiology and clinical implications. *Curr Opin Gastroenterol*, 23, 379-83.
- MANSUETO, P., SEIDITA, A., D'ALCAMO, A. & CARROCCIO, A. 2014. Non-celiac gluten sensitivity: literature review. *J Am Coll Nutr*, 33, 39-54.
- MARAI, I., SHOENFELD, Y., BIZZARO, N., VILLALTA, D., DORIA, A., TONUTTI, E. & TOZZOLI, R. 2004. IgA and IgG tissue transglutaminase antibodies in systemic lupus erythematosus. *Lupus*, 13, 241-4.
- MARIETTA EV, GOMEZ AM, YEOMAN C, TILAHUN AY, CLARK CR, LUCKEY DH, ET AL. 2013 Low incidence of spontaneous type 1 diabetes in non-obese diabetic mice raised on gluten-free diets is associated with changes in the intestinal microbiome. *PloS one*, 8(11):e78687.
- MARILD, K., LUDVIGSSON, J., SANZ, Y. & LUDVIGSSON, J. F. 2014. Antibiotic exposure in pregnancy and risk of coeliac disease in offspring: a cohort study. *BMC Gastroenterol*, 14, 75.
- MARILD, K., STEPHANSSON, O., MONTGOMERY, S., MURRAY, J. A. & LUDVIGSSON, J. F. 2012. Pregnancy outcome and risk of celiac disease in offspring: a nationwide case-control study. *Gastroenterology*, 142, 39-45.e3.
- MARILD, K., YE, W., LEBWOHL, B., GREEN, P. H., BLASER, M. J., CARD, T. & LUDVIGSSON, J. F. 2013. Antibiotic exposure and the development of coeliac disease: a nationwide case-control study. *BMC Gastroenterol*, 13, 109.
- MARSH, M. N. 1992. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology*, 102, 330-54.
- MAYBERRY, J. F., SMART, H. L. & TOGHILL, P. J. 1986. Familial association between coeliac disease and ulcerative colitis: preliminary communication. *J R Soc Med*, 79, 204-5.
- MCLEOD, D. S. & COOPER, D. S. 2012. The incidence and prevalence of thyroid autoimmunity. *Endocrine*, 42, 252-65.
- MEGIORNI, F., MORA, B., BONAMICO, M., BARBATO, M., NENNA, R., MAIELLA, G., LULLI, P. & MAZZILLI, M. C. 2009. HLA-DQ and risk gradient for celiac disease. *Hum Immunol*, 70, 55-9.



- MELONI, A., MANDAS, C., JORES, R. D. & CONGIA, M. 2009. Prevalence of autoimmune thyroiditis in children with celiac disease and effect of gluten withdrawal. *J Pediatr*, 155, 51-5, 55.e1.
- METSO, S., HYYTIA-ILMONEN, H., KAUKINEN, K., HUHTALA, H., JAATINEN, P., SALMI, J., TAURIO, J. & COLLIN, P. 2012. Gluten-free diet and autoimmune thyroiditis in patients with celiac disease. A prospective controlled study. *Scand J Gastroenterol*, 47, 43-8.
- MIRANDA, D. M., MASSOM, J. N., CATARINO, R. M., SANTOS, R. T., TOYODA, S. S., MARONE, M. M., TOMIMORI, E. K. & MONTE, O. 2015. Impact of nutritional iodine optimization on rates of thyroid hypoechogenicity and autoimmune thyroiditis: a cross-sectional, comparative study. *Thyroid*, 25, 118-24.
- MITEA, C., HAVENAAR, R., DRIJFHOUT, J. W., EDENS, L., DEKKING, L. & KONING, F. 2008. Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease. *Gut*, 57, 25-32.
- MOLBERG, O., UHLEN, A. K., JENSEN, T., FLAETE, N. S., FLECKENSTEIN, B., ARENTZ-HANSEN, H., RAKI, M., LUNDIN, K. E. & SOLLID, L. M. 2005. Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. *Gastroenterology*, 128, 393-401.
- MONES, R. L. 2009. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *J Pediatr Gastroenterol Nutr*, 48, 645-6.
- MONTALTO, M., CUOCO, L., RICCI, R., MAGGIANO, N., VECCHIO, F. M. & GASBARRINI, G. 2002. Immunohistochemical analysis of ZO-1 in the duodenal mucosa of patients with untreated and treated celiac disease. *Digestion*, 65, 227-33.
- MORENO, M. L., RODRIGUEZ-HERRERA, A., SOUSA, C. & COMINO, I. 2017. Biomarkers to Monitor Gluten-Free Diet Compliance in Celiac Patients. *Nutrients*, 9.
- MURRAY, J. A., BREANN DAN MOORE, S., VAN DYKE, C. T., LAHR, B. D., DIERKHISING, R. A., ZINSMEISTER, A. R., JOSEPH MELTON, L., KRONING, C. M., EL-YOUSSEFF, M. & CZAJA, A. J. 2007. HLA DQ Gene Dosage and Risk and Severity of Celiac Disease. *Clin Gastroenterol Hepatol*, 5, 1406-12.
- MURRAY, J. A., VAN DYKE, C., PLEVAK, M. F., DIERKHISING, R. A., ZINSMEISTER, A. R. & MELTON, L. J., 3RD 2003. Trends in the identification and clinical features of celiac disease in a North American community, 1950-2001. *Clin Gastroenterol Hepatol*, 1, 19-27.
- MURRAY, J. A., WATSON, T., CLEARMAN, B. & MITROS, F. 2004. Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease. *Am J Clin Nutr*, 79, 669-73.
- MUSTALAHTI, K., CATASSI, C., REUNANEN, A., FABIANI, E., HEIER, M., MCMILLAN, S., MURRAY, L., METZGER, M. H., GASPARIN, M., BRAVI, E. & MAKI, M. 2010. The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med*, 42, 587-95.
- MUSTALAHTI, K., HOLOPAINEN, P., KARELL, K., MAKI, M. & PARTANEN, J. 2002. Genetic dissection between silent and clinically diagnosed symptomatic forms of coeliac disease in multiplex families. *Dig Liver Dis*, 34, 842-5.
- MYASOEDOVA, E., CROWSON, C. S., KREMERS, H. M., THERNEAU, T. M. & GABRIEL, S. E. 2010. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis Rheum*, 62, 1576-82.
- MYLOTTE, M., EGAN-MITCHELL, B., MCCARTHY, C. F. & MCNICHOLL, B. 1973. Incidence of coeliac disease in the West of Ireland. *Br Med J*, 1, 703-5.
- NAIYER, A. J., SHAH, J., HERNANDEZ, L., KIM, S. Y., CIACCIO, E. J., CHENG, J., MANAVALAN, S., BHAGAT, G. & GREEN, P. H. 2008. Tissue transglutaminase antibodies in individuals with celiac disease bind to thyroid follicles and extracellular matrix and may contribute to thyroid dysfunction. *Thyroid*, 18, 1171-8.
- NAWOOR, Z., BURNS, R., SMITH, D. F., SHEEHAN, S., O'HERLIHY, C. & SMYTH, P. P. 2006. Iodine intake in pregnancy in Ireland--a cause for concern? *Ir J Med Sci*, 175, 21-4.

- NEUHAUSEN, S. L., STEELE, L., RYAN, S., MOUSAVI, M., PINTO, M., OSANN, K. E., FLODMAN, P. & ZONE, J. J. 2008. Co-occurrence of celiac disease and other autoimmune diseases in celiacs and their first-degree relatives. *J Autoimmun*, 31, 160-5.
- NEUNHAM, E. D., SHEPHERD, S. J., STRAUSS, B. J., HOSKING, P. & GIBSON, P. R. 2016. Adherence to the gluten-free diet can achieve the therapeutic goals in almost all patients with coeliac disease: A 5-year longitudinal study from diagnosis. *J Gastroenterol Hepatol*, 31, 342-9.
- NGO, S. T., STEYN, F. J. & MCCOMBE, P. A. 2014. Gender differences in autoimmune disease. *Front Neuroendocrinol*, 35, 347-69.
- NICOLAE, D. L., GAMAZON, E., ZHANG, W., DUAN, S., DOLAN, M. E. & COX, N. J. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*, 6, e1000888.
- NICOLETTI, A., PATTI, F., LO FERMO, S., SCIACCA, A., LAISA, P., LIBERTO, A., LANZAFAME, S., CONTRAFFATTO, D., D'AGATE, C., RUSSO, A. & ZAPPIA, M. 2008. Frequency of celiac disease is not increased among multiple sclerosis patients. *Mult Scler*, 14, 698-700.
- NISTICO, L., FAGNANI, C., COTO, I., PERCOPO, S., COTICHINI, R., LIMONGELLI, M. G., PAPARO, F., D'ALFONSO, S., GIORDANO, M., SFERLAZZAS, C., MAGAZZU, G., MOMIGLIANO-RICHIARDI, P., GRECO, L. & STAZI, M. A. 2006. Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut*, 55, 803-8.
- NORSTROM, F., LINDHOLM, L., SANDSTROM, O., NORDYKE, K. & IVARSSON, A. 2011. Delay to celiac disease diagnosis and its implications for health-related quality of life. *BMC Gastroenterol*, 11, 118.
- OBERHUBER, G., GRANDITSCH, G. & VOGELSANG, H. 1999. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol*, 11, 1185-94.
- OERTLI, M., SUNDQUIST, M., HITZLER, I., ENGLER, D. B., ARNOLD, I. C., REUTER, S., MAXEINER, J., HANSSON, M., TAUBE, C., QUIDING-JARBRINK, M. & MULLER, A. 2012. DC-derived IL-18 drives Treg differentiation, murine *Helicobacter pylori*-specific immune tolerance, and asthma protection. *J Clin Invest*, 122, 1082-96.
- OKADA, H., KUHN, C., FEILLET, H. & BACH, J. F. 2010. The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clin Exp Immunol*, 160, 1-9.
- OLEN, O., MONTGOMERY, S. M., MARCUS, C., EKBOM, A. & LUDVIGSSON, J. F. 2009. Coeliac disease and body mass index: a study of two Swedish general population-based registers. *Scand J Gastroenterol*, 44, 1198-206.
- OLIVARES, M., NEEF, A., CASTILLEJO, G., PALMA, G. D., VAREA, V., CAPILLA, A., PALAU, F., NOVA, E., MARCOS, A., POLANCO, I., RIBES-KONINCKX, C., ORTIGOSA, L., IZQUIERDO, L. & SANZ, Y. 2015. The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. *Gut*, 64, 406-17.
- ORGAZ-MOLINA, J., BUENDIA-EISMAN, A., ARRABAL-POLO, M. A., RUIZ, J. C. & ARIAS-SANTIAGO, S. 2012. Deficiency of serum concentration of 25-hydroxyvitamin D in psoriatic patients: a case-control study. *J Am Acad Dermatol*, 67, 931-8.
- OSTENSSON, M., MONTEN, C., BACELIS, J., GUDJONSDOTTIR, A. H., ADAMOVIC, S., EK, J., ASCHER, H., POLLAK, E., ARNELL, H., BROWALDH, L., AGARDH, D., WAHLSTROM, J., NILSSON, S. & TORINSSON-NALUAI, A. 2013. A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease. *PLoS One*, 8, e70174.
- OUAKA-KCHAOU, A., ENNAIFER, R., ELLOUMI, H., GARGOURI, D., HEFAIEDH, R., KOCHLEF, A., ROMANI, M., KILANI, A., KHARRAT, J. & GHORBEL, A. 2008. Autoimmune diseases in coeliac disease: effect of gluten exposure. *Therap Adv Gastroenterol*, 1, 169-72.

- OXENTENKO, A. S. & MURRAY, J. A. 2015. Celiac Disease: Ten Things That Every Gastroenterologist Should Know. *Clin Gastroenterol Hepatol*, 13, 1396-404; quiz e127-9.
- PALLAV, K., LEFFLER, D. A., BENNETT, M., TARIQ, S., XU, H., KABBANI, T., MOSS, A. C., DENNIS, M., KELLY, C. P. & SCHUPPAN, D. 2012. Open conformation tissue transglutaminase testing for celiac dietary assessment. *Dig Liver Dis*, 44, 375-8.
- PALMA, G. D., CAPILLA, A., NOVA, E., CASTILLEJO, G., VAREA, V., POZO, T., GARROTE, J. A., POLANCO, I., LOPEZ, A., RIBES-KONINCKX, C., MARCOS, A., GARCIA-NOVO, M. D., CALVO, C., ORTIGOSA, L., PENA-QUINTANA, L., PALAU, F. & SANZ, Y. 2012. Influence of milk-feeding type and genetic risk of developing coeliac disease on intestinal microbiota of infants: the PROFICEL study. *PLoS One*, 7, e30791.
- PANNARAJ, P. S., LI, F., CERINI, C., BENDER, J. M., YANG, S., ROLLIE, A., ADISETIYO, H., ZABIH, S., LINCEZ, P. J., BITTINGER, K., BAILEY, A., BUSHMAN, F. D., SLEASMAN, J. W. & ALDROVANDI, G. M. 2017. Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr*, 171, 647-654.
- PASCUAL, V., DIELI-CRIMI, R., LOPEZ-PALACIOS, N., BODAS, A., MEDRANO, L. M. & NUNEZ, C. 2014. Inflammatory bowel disease and celiac disease: overlaps and differences. *World J Gastroenterol*, 20, 4846-56.
- PASTORE, M. R., BAZZIGALUPPI, E., BELLONI, C., ARCOVIO, C., BONIFACIO, E. & BOSI, E. 2003. Six months of gluten-free diet do not influence autoantibody titers, but improve insulin secretion in subjects at high risk for type 1 diabetes. *J Clin Endocrinol Metab*, 88, 162-5.
- PATEL, D., KALKAT, P., BAISCH, D. & ZIPSER, R. 2005. Celiac disease in the elderly. *Gerontology*, 51, 213-4.
- PATERSON, B. M., LAMMERS, K. M., ARRIETA, M. C., FASANO, A. & MEDDINGS, J. B. 2007. The safety, tolerance, pharmacokinetic and pharmacodynamic effects of single doses of AT-1001 in coeliac disease subjects: a proof of concept study. *Aliment Pharmacol Ther*, 26, 757-66.
- PAZIANAS, M., BUTCHER, G. P., SUBHANI, J. M., FINCH, P. J., ANG, L., COLLINS, C., HEANEY, R. P., ZAIDI, M. & MAXWELL, J. D. 2005. Calcium absorption and bone mineral density in celiacs after long term treatment with gluten-free diet and adequate calcium intake. *Osteoporos Int*, 16, 56-63.
- PENAGINI, F., DILILLO, D., MENEGHIN, F., MAMELI, C., FABIANO, V. & ZUCCOTTI, G. V. 2013. Gluten-free diet in children: an approach to a nutritionally adequate and balanced diet. *Nutrients*, 5, 4553-65.
- PENGIRAN TENGAH, C. D., LOCK, R. J., UNSWORTH, D. J. & WILLS, A. J. 2004. Multiple sclerosis and occult gluten sensitivity. *Neurology*, 62, 2326-7.
- PESCHKEN, C. A. & HITCHON, C. A. 2012. Rising prevalence of systemic autoimmune rheumatic disease: increased awareness, increased disease or increased survival? *Arthritis Research & Therapy*, 14, A20-A20.
- PETAROS, P., MARTELOSSI, S., TOMMASINI, A., TORRE, G., CARADONNA, M. & VENTURA, A. 2002. Prevalence of autoimmune disorders in relatives of patients with celiac disease. *Dig Dis Sci*, 47, 1427-31.
- PETUKHOVA, L. & CHRISTIANO, A. M. 2016. Functional Interpretation of Genome-Wide Association Study Evidence in Alopecia Areata. *J Invest Dermatol*, 136, 314-7.
- PETUKHOVA, L., DUVIC, M., HORDINSKY, M., NORRIS, D., PRICE, V., SHIMOMURA, Y., KIM, H., SINGH, P., LEE, A., CHEN, W. V., MEYER, K. C., PAUS, R., JAHODA, C. A., AMOS, C. I., GREGERSEN, P. K. & CHRISTIANO, A. M. 2010. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature*, 466, 113-7.
- PINKAS, D. M., STROP, P., BRUNGER, A. T. & KHOSLA, C. 2007. Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol*, 5, e327.

- PRATESI, R., GANDOLFI, L., FRIEDMAN, H., FARAGE, L., DE CASTRO, C. A. & CATASSI, C. 1998. Serum IgA antibodies from patients with coeliac disease react strongly with human brain blood-vessel structures. *Scand J Gastroenterol*, 33, 817-21.
- PURCELL, S. M., WRAY, N. R., STONE, J. L., VISSCHER, P. M., O'DONOVAN, M. C., SULLIVAN, P. F. & SKLAR, P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460, 748-52.
- RAHMATI, A., SHAKERI, R., SOHRABI, M., ALIPOUR, A., BOGHRATIAN, A., SETAREH, M. & ZAMANI, F. 2014. Correlation of tissue transglutaminase antibody with duodenal histologic marsh grading. *Middle East J Dig Dis*, 6, 131-6.
- RAMAKRISHNA, B. S., MAKHARIA, G. K., CHETRI, K., DUTTA, S., MATHUR, P., AHUJA, V., AMARCHAND, R., BALAMURUGAN, R., CHOWDHURY, S. D., DANIEL, D., DAS, A., GEORGE, G., GUPTA, S. D., KRISHNAN, A., PRASAD, J. H., KAUR, G., PUGAZHENDHI, S., PULIMOOD, A., RAMAKRISHNA, K. & VERMA, A. K. 2016. Prevalence of Adult Celiac Disease in India: Regional Variations and Associations. *Am J Gastroenterol*, 111, 115-23.
- RAMPERTAB, S. D., POORAN, N., BRAR, P., SINGH, P. & GREEN, P. H. 2006. Trends in the presentation of celiac disease. *Am J Med*, 119, 355.e9-14.
- RASCH, E. K., HIRSCH, R., PAULOSE-RAM, R. & HOCHBERG, M. C. 2003. Prevalence of rheumatoid arthritis in persons 60 years of age and older in the United States: effect of different methods of case classification. *Arthritis Rheum*, 48, 917-26.
- RASHID, M. & LEE, J. 2016. Serologic testing in celiac disease: Practical guide for clinicians. *Can Fam Physician*, 62, 38-43.
- RASHTAK, S. & MURRAY, J. A. 2009. Celiac disease in the elderly. *Gastroenterol Clin North Am*, 38, 433-46.
- RAVIKUMARA, M., TUTHILL, D. P. & JENKINS, H. R. 2006. The changing clinical presentation of coeliac disease. *Arch Dis Child*, 91, 969-71.
- REICHEL, K. L. & JENSEN, D. 2004. IgA antibodies against gliadin and gluten in multiple sclerosis. *Acta Neurol Scand*, 110, 239-41.
- REID, J. R. & WHEELER, S. F. 2005. Hyperthyroidism: diagnosis and treatment. *Am Fam Physician*, 72, 623-30.
- REILLY, N. R., FASANO, A. & GREEN, P. H. 2012. Presentation of celiac disease. *Gastrointest Endosc Clin N Am*, 22, 613-21.
- REILLY, N. R. & GREEN, P. H. 2012. Epidemiology and clinical presentations of celiac disease. *Semin Immunopathol*, 34, 473-8.
- REUNALA, T. & LOKKI, J. 1978. Dermatitis herpetiformis in Finland. *Acta Derm Venereol*, 58, 505-10.
- REUNALA, T., SALMI, T. T., HERVONEN, K., LAURILA, K., KAUTIAINEN, H., COLLIN, P. & KAUKINEN, K. 2015. IgA antiepidermal transglutaminase antibodies in dermatitis herpetiformis: a significant but not complete response to a gluten-free diet treatment. *Br J Dermatol*, 172, 1139-41.
- RICANO-PONCE, I., WIJMENGA, C. & GUTIERREZ-ACHURY, J. 2015. Genetics of celiac disease. *Best Pract Res Clin Gastroenterol*, 29, 399-412.
- RICHES, P. L., MCRORIE, E., FRASER, W. D., DETERMANN, C., VAN'T HOF, R. & RALSTON, S. H. 2009. Osteoporosis associated with neutralizing autoantibodies against osteoprotegerin. *N Engl J Med*, 361, 1459-65.
- RIDDLE, M. S., MURRAY, J. A., CASH, B. D., PIMENTEL, M. & PORTER, C. K. 2013. Pathogen-specific risk of celiac disease following bacterial causes of foodborne illness: a retrospective cohort study. *Dig Dis Sci*, 58, 3242-5.
- RIDDLE, M. S., MURRAY, J. A. & PORTER, C. K. 2012. The incidence and risk of celiac disease in a healthy US adult population. *Am J Gastroenterol*, 107, 1248-55.
- RODRIGO, L., HERNANDEZ-LAHOZ, C., FUENTES, D., ALVAREZ, N., LOPEZ-VAZQUEZ, A. & GONZALEZ, S. 2011. Prevalence of celiac disease in multiple sclerosis. *BMC Neurol*, 11, 31.

- ROMANOS, J., ROSEN, A., KUMAR, V., TRYNKA, G., FRANKE, L., SZPERL, A., GUTIERREZ-ACHURY, J., VAN DIEMEN, C. C., KANNINGA, R., JANKIPERSADSING, S. A., STECK, A., EISENBARTH, G., VAN HEEL, D. A., CUKROWSKA, B., BRUNO, V., MAZZILLI, M. C., NUNEZ, C., BILBAO, J. R., MEARIN, M. L., BARISANI, D., REWERS, M., NORRIS, J. M., IVARSSON, A., BOEZEN, H. M., LIU, E. & WIJMENGA, C. 2014. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. *Gut*, 63, 415-22.
- ROMANOS, J., VAN DIEMEN, C. C., NOLTE, I. M., TRYNKA, G., ZHERNAKOVA, A., FU, J., BARDELLA, M. T., BARISANI, D., MCMANUS, R., VAN HEEL, D. A. & WIJMENGA, C. 2009. Analysis of HLA and non-HLA alleles can identify individuals at high risk for coeliac disease. *Gastroenterology*, 137, 834-40, 840 e1-3.
- ROSE, A. M. & BELL, L. C. 2012. Epistasis and immunity: the role of genetic interactions in autoimmune diseases. *Immunology*, 137, 131-8.
- ROSTAMI, K. & VILLANACCI, V. 2009. Microscopic enteritis: novel prospect in coeliac disease clinical and immuno-histogenesis. Evolution in diagnostic and treatment strategies. *Dig Liver Dis*, 41, 245-52.
- ROSTOM, A., DUBE, C., CRANNEY, A., SALOOJEE, N., SY, R., GARRITTY, C., SAMPSON, M., ZHANG, L., YAZDI, F., MAMALADZE, V., PAN, I., MACNEIL, J., MACK, D., PATEL, D. & MOHER, D. 2005. The diagnostic accuracy of serologic tests for coeliac disease: a systematic review. *Gastroenterology*, 128, S38-46.
- ROY, A., LASZKOWSKA, M., SUNDSTROM, J., LEBWOHL, B., GREEN, P. H., KAMPE, O. & LUDVIGSSON, J. F. 2016. Prevalence of Celiac Disease in Patients with Autoimmune Thyroid Disease: A Meta-Analysis. *Thyroid*, 26, 880-90.
- RUBIO-TAPIA, A., HILL, I. D., KELLY, C. P., CALDERWOOD, A. H. & MURRAY, J. A. 2013. ACG clinical guidelines: diagnosis and management of coeliac disease. *Am J Gastroenterol*, 108, 656-76; quiz 677.
- RUBIO-TAPIA, A., KYLE, R. A., KAPLAN, E. L., JOHNSON, D. R., PAGE, W., ERDTMANN, F., BRANTNER, T. L., KIM, W. R., PHELPS, T. K., LAHR, B. D., ZINSMEISTER, A. R., MELTON, L. J., 3RD & MURRAY, J. A. 2009. Increased prevalence and mortality in undiagnosed coeliac disease. *Gastroenterology*, 137, 88-93.
- RUBIO-TAPIA, A., LUDVIGSSON, J. F., BRANTNER, T. L., MURRAY, J. A. & EVERHART, J. E. 2012. The prevalence of coeliac disease in the United States. *Am J Gastroenterol*, 107, 1538-44; quiz 1537, 1545.
- RUBIO-TAPIA, A. & MURRAY, J. A. 2010. Classification and management of refractory coeliac disease. *Gut*, 59, 547-57.
- RUBIO-TAPIA, A., VAN DYKE, C. T., LAHR, B. D., ZINSMEISTER, A. R., EL-YOUSSEF, M., MOORE, S. B., BOWMAN, M., BURGART, L. J., MELTON, L. J., 3RD & MURRAY, J. A. 2008. Predictors of family risk for coeliac disease: a population-based study. *Clin Gastroenterol Hepatol*, 6, 983-7.
- SAINSBURY, A., SANDERS, D. S. & FORD, A. C. 2013. Prevalence of irritable bowel syndrome-type symptoms in patients with coeliac disease: a meta-analysis. *Clin Gastroenterol Hepatol*, 11, 359-65.e1.
- SALEEM, A., CONNOR, H. J. & REGAN, P. O. 2012. Adult coeliac disease in Ireland: a case series. *Ir J Med Sci*, 181, 225-9.
- SALMI, T. T., COLLIN, P., REUNALA, T., MAKI, M. & KAUKINEN, K. 2010. Diagnostic methods beyond conventional histology in coeliac disease diagnosis. *Dig Liver Dis*, 42, 28-32.
- SALMI, T. T., HERVONEN, K., KAUTIAINEN, H., COLLIN, P. & REUNALA, T. 2011. Prevalence and incidence of dermatitis herpetiformis: a 40-year prospective study from Finland. *Br J Dermatol*, 165, 354-9.
- SALVATORE, S., FINAZZI, S., GHEZZI, A., TOSI, A., BARASSI, A., LUINI, C., BETTINI, B., ZIBETTI, A., NESPOLI, L. & MELZI D'ERIL, G. V. 2004. Multiple sclerosis and coeliac disease: is there an increased risk? *Mult Scler*, 10, 711-2.
- SANZ, Y. 2015. Microbiome and Gluten. *Ann Nutr Metab*, 67 Suppl 2, 28-41.

- SAPONE, A., DE MAGISTRIS, L., PIETZAK, M., CLEMENTE, M. G., TRIPATHI, A., CUCCA, F., LAMPIS, R., KRYSZAK, D., CARTENI, M., GENEROSO, M., IAFUSCO, D., PRISCO, F., LAGHI, F., RIEGLER, G., CARRATU, R., COUNTS, D. & FASANO, A. 2006. Zonulin upregulation is associated with increased gut permeability in subjects with type 1 diabetes and their relatives. *Diabetes*, 55, 1443-9.
- SARDY, M., KARPATI, S., MERKL, B., PAULSSON, M. & SMYTH, N. 2002. Epidermal transglutaminase (TGase 3) is the autoantigen of dermatitis herpetiformis. *J Exp Med*, 195, 747-57.
- SATEGNA GUIDETTI, C., SOLERIO, E., SCAGLIONE, N., AIMO, G. & MENGOZZI, G. 2001. Duration of gluten exposure in adult coeliac disease does not correlate with the risk for autoimmune disorders. *Gut*, 49, 502-5.
- SATEGNA-GUIDETTI, C., FRANCO, E., MARTINI, S. & BOBBIO, M. 2004. Binding by serum IgA antibodies from patients with coeliac disease to monkey heart tissue. *Scand J Gastroenterol*, 39, 540-3.
- SATEGNA-GUIDETTI, C., VOLTA, U., CIACCI, C., USAI, P., CARLINO, A., DE FRANCESCHI, L., CAMERA, A., PELLI, A. & BROSSA, C. 2001. Prevalence of thyroid disorders in untreated adult celiac disease patients and effect of gluten withdrawal: an Italian multicenter study. *Am J Gastroenterol*, 96, 751-7.
- SCHMID, S., KOCZWARA, K., SCHWINGHAMMER, S., LAMPASONA, V., ZIEGLER, A. G. & BONIFACIO, E. 2004. Delayed exposure to wheat and barley proteins reduces diabetes incidence in non-obese diabetic mice. *Clin Immunol*, 111, 108-18.
- SCHOSLER, L., CHRISTENSEN, L. A. & HVAS, C. L. 2016. Symptoms and findings in adult-onset celiac disease in a historical Danish patient cohort. *Scand J Gastroenterol*, 51, 288-94.
- SCHUPPAN, D. & ZEVALLOS, V. 2015. Wheat amylase trypsin inhibitors as nutritional activators of innate immunity. *Dig Dis*, 33, 260-3.
- SCHYUM, A. C. & RUMESSEN, J. J. 2013. Serological testing for celiac disease in adults. *United European Gastroenterol J*, 1, 319-25.
- SECONDULFO, M., IAFUSCO, D., CARRATU, R., DEMAGISTRIS, L., SAPONE, A., GENEROSO, M., MEZZOGIOMO, A., SASSO, F. C., CARTENI, M., DE ROSA, R., PRISCO, F. & ESPOSITO, V. 2004. Ultrastructural mucosal alterations and increased intestinal permeability in non-celiac, type I diabetic patients. *Dig Liver Dis*, 36, 35-45.
- SELLITTO, M., BAI, G., SERENA, G., FRICKE, W. F., STURGEON, C., GAJER, P., WHITE, J. R., KOENIG, S. S., SAKAMOTO, J., BOOTHE, D., GICQUELAIS, R., KRYSZAK, D., PUPPA, E., CATASSI, C., RAVEL, J. & FASANO, A. 2012. Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One*, 7, e33387.
- SERENA, G., CAMHI, S., STURGEON, C., YAN, S. & FASANO, A. 2015. The Role of Gluten in Celiac Disease and Type 1 Diabetes. *Nutrients*, 7, 7143-62.
- SHAN, L., MOLBERG, O., PARROT, I., HAUSCH, F., FILIZ, F., GRAY, G. M., SOLLID, L. M. & KHOSLA, C. 2002. Structural basis for gluten intolerance in celiac sprue. *Science*, 297, 2275-9.
- SHARKEY, L. M., CORBETT, G., CURRIE, E., LEE, J., SWEENEY, N. & WOODWARD, J. M. 2013. Optimising delivery of care in coeliac disease - comparison of the benefits of repeat biopsy and serological follow-up. *Aliment Pharmacol Ther*, 38, 1278-91.
- SHIVASHANKAR, R., TREMAINE, W. J., HARMSSEN, W. S. & LOFTUS, E. V., JR. 2016. Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol*.
- SHOR, D. B., BARZILAI, O., RAM, M., IZHAKY, D., PORAT-KATZ, B. S., CHAPMAN, J., BLANK, M., ANAYA, J. M. & SHOENFELD, Y. 2009. Gluten sensitivity in multiple sclerosis: experimental myth or clinical truth? *Ann N Y Acad Sci*, 1173, 343-9.
- SILDORF, S. M., FREDHEIM, S., SVENSSON, J. & BUSCHARD, K. 2012. Remission without insulin therapy on gluten-free diet in a 6-year old boy with type 1 diabetes mellitus. *BMJ Case Rep*, 2012.

- SIMON-VECSEI, Z., KIRALY, R., BAGOSSI, P., TOTH, B., DAHLBOM, I., CAJA, S., CSOSZ, E., LINDFORS, K., SBLATTERO, D., NEMES, E., MAKI, M., FESUS, L. & KORPONAY-SZABO, I. R. 2012. A single conformational transglutaminase 2 epitope contributed by three domains is critical for celiac antibody binding and effects. *Proc Natl Acad Sci U S A*, 109, 431-6.
- SIMRE, K., UIBO, O., PEET, A., TILLMANN, V., KOOL, P., HAMALAINEN, A. M., HARKONEN, T., SILJANDER, H., VIRTANEN, S., ILONEN, J., KNIP, M. & UIBO, R. 2016. Exploring the risk factors for differences in the cumulative incidence of coeliac disease in two neighbouring countries: the prospective DIABIMMUNE study. *Dig Liver Dis*, 48, 1296-1301.
- SINGH, P. & GARBER, J. J. 2016. Implementation and adherence to osteoporosis screening guidelines among coeliac disease patients. *Dig Liver Dis*, 48, 1451-1456.
- SINGH, P., KURRAY, L., AGNIHOTRI, A., DAS, P., VERMA, A. K., SREENIVAS, V., DATTAGUPTA, S. & MAKHARIA, G. K. 2014. Titers of Anti-tissue Transglutaminase Antibody Correlate Well With Severity of Villous Abnormalities in Celiac Disease. *J Clin Gastroenterol*.
- SJOBERG, V., SANDSTROM, O., HEDBERG, M., HAMMARSTROM, S., HERNELL, O. & HAMMARSTROM, M. L. 2013. Intestinal T-cell responses in celiac disease - impact of celiac disease associated bacteria. *PLoS One*, 8, e53414.
- SMUKALLA, S., LEBWOHL, B., MEARS, J. G., LESLIE, L. A. & GREEN, P. H. 2014. How often do hematologists consider celiac disease in iron-deficiency anemia? Results of a national survey. *Clin Adv Hematol Oncol*, 12, 100-5.
- SMYTH, P. & O'HERLIHY, C. 2012. Dietary iodine intake in pregnancy: an update. *Ir Med J*, 105, 5-6.
- SMYTH, P. P., HETHERTON, A. M., O'CARROLL, D., SMITH, D. F., O'HALLORAN, M. J. & O'DONOVAN, D. K. 1988. Iodine intake in Ireland--an update. *Ir J Med Sci*, 157, 377-9.
- SNOW, C. F. 1999. Laboratory diagnosis of vitamin B12 and folate deficiency: a guide for the primary care physician. *Arch Intern Med*, 159, 1289-98.
- SNYDER, M. R. & MURRAY, J. A. 2015. Celiac Disease: Advances in Diagnosis. *Expert Rev Clin Immunol*.
- SO, H. C. & SHAM, P. C. 2017. Exploring the predictive power of polygenic scores derived from genome-wide association studies: a study of 10 complex traits. *Bioinformatics*, 33, 886-892.
- SOLLID, L. M. 2002. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol*, 2, 647-55.
- SPATOLA BN, KAUKINEN K, COLLIN P, MAKI M, KAGNOFF MF, DAUGHERTY PS. 2014 Persistence of elevated deamidated gliadin peptide antibodies on a gluten-free diet indicates nonresponsive coeliac disease. *Alimentary pharmacology & therapeutics*, 39(4):407-17.
- STAHL, E. A., WEGMANN, D., TRYNKA, G., GUTIERREZ-ACHURY, J., DO, R., VOIGHT, B. F., KRAFT, P., CHEN, R., KALLBERG, H. J., KURREEMAN, F. A., KATHIRESAN, S., WIJMENGA, C., GREGERSEN, P. K., ALFREDSSON, L., SIMINOVITCH, K. A., WORTHINGTON, J., DE BAKKER, P. I., RAYCHAUDHURI, S. & PLENGE, R. M. 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet*, 44, 483-9.
- STECK, A. K., ARMSTRONG, T. K., BABU, S. R. & EISENBARTH, G. S. 2011. Stepwise or linear decrease in penetrance of type 1 diabetes with lower-risk HLA genotypes over the past 40 years. *Diabetes*, 60, 1045-9.
- STENE, L. C., HONEYMAN, M. C., HOFFENBERG, E. J., HAAS, J. E., SOKOL, R. J., EMERY, L., TAKI, I., NORRIS, J. M., ERLICH, H. A., EISENBARTH, G. S. & REWERS, M. 2006. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol*, 101, 2333-40.

- SYAGE JA, MURRAY JA, GREEN PHR, KHOSLA C. 2017 Latiglutenase Improves Symptoms in Seropositive Celiac Disease Patients While on a Gluten-Free Diet. *Digestive diseases and sciences*, 62(9):2428-32.
- SZAJEWSKA, H., SHAMIR, R., MEARIN, M. L., KONINCKX, C. R., CATASSI, C., DOMELLOF, M., FEWTRELL, M. S., HUSBY, S., PAPADOPOULOU, A., VANDENPLAS, Y., CASTILLEJO, G., KOLACEK, S., KOLETZKO, S., KORPONAY-SZABO, I. R., LIONETTI, E., POLANCO, I. & TRONCONE, R. 2016. Gluten Introduction and The Risk of Coeliac Disease. A Position Paper By The European Society For Paediatric Gastroenterology, Hepatology & Nutrition. *J Pediatr Gastroenterol Nutr*.
- TAAVELA, J., KURPPA, K., COLLIN, P., LAHDEAHO, M. L., SALMI, T., SAAVALAINEN, P., HAIMILA, K., HUHTALA, H., LAURILA, K., SIEVANEN, H., MAKI, M. & KAUKINEN, K. 2013. Degree of damage to the small bowel and serum antibody titers correlate with clinical presentation of patients with celiac disease. *Clin Gastroenterol Hepatol*, 11, 166-71.e1.
- TACK, G. J., VAN DE WATER, J. M., BRUINS, M. J., KOOY-WINKELAAR, E. M., VAN BERGEN, J., BONNET, P., VREUGDENHIL, A. C., KORPONAY-SZABO, I., EDENS, L., VON BLOMBERG, B. M., SCHREURS, M. W., MULDER, C. J. & KONING, F. 2013. Consumption of gluten with gluten-degrading enzyme by celiac patients: a pilot-study. *World J Gastroenterol*, 19, 5837-47.
- TAJUDDIN, T., RAZIF, S., DHAR, R., THORNE, J. & MURRAY, F. E. 2011. Clinical presentation of adult coeliac disease. *Ir Med J*, 104, 20-2.
- TAVAKKOLI, A., DIGIACOMO, D., GREEN, P. H. & LEBWOHL, B. 2013. Vitamin D status and concomitant autoimmunity in celiac disease. *J Clin Gastroenterol*, 47, 515-9.
- TELEGA, G., BENNET, T. R. & WERLIN, S. 2008. Emerging new clinical patterns in the presentation of celiac disease. *Arch Pediatr Adolesc Med*, 162, 164-8.
- TERESI, S., CRAPISI, M., VALLEJO, M. D., CASTELLANETA, S. P., FRANCAVILLA, R., IACONO, G., RAVELLI, A., MENEGAZZI, P., LOUALI, M. & CATASSI, C. 2010. Celiac disease seropositivity in Saharawi children: a follow-up and family study. *J Pediatr Gastroenterol Nutr*, 50, 506-9.
- THOMPSON, T., LEE, A. R. & GRACE, T. 2010. Gluten contamination of grains, seeds, and flours in the United States: a pilot study. *J Am Diet Assoc*, 110, 937-40.
- TILG, H., KOCH, R. & MOSCHEN, A. R. 2013. Proinflammatory wheat attacks on the intestine: alpha-amylase trypsin inhibitors as new players. *Gastroenterology*, 144, 1561-3; discussion 1563-4.
- TILG, H., MOSCHEN, A. R., KASER, A., PINES, A. & DOTAN, I. 2008. Gut, inflammation and osteoporosis: basic and clinical concepts. *Gut*, 57, 684-94.
- TOLLEFSON, M. M., CROWSON, C. S., MCEVOY, M. T. & MARADIT KREMERS, H. 2010. Incidence of psoriasis in children: a population-based study. *J Am Acad Dermatol*, 62, 979-87.
- TOSCANO, V., CONTI, F. G., ANASTASI, E., MARIANI, P., TIBERTI, C., POGGI, M., MONTUORI, M., MONTI, S., LAURETI, S., CIPOLLETTA, E., GEMME, G., CAIOLA, S., DI MARIO, U. & BONAMICO, M. 2000. Importance of gluten in the induction of endocrine autoantibodies and organ dysfunction in adolescent celiac patients. *Am J Gastroenterol*, 95, 1742-8.
- TRYNKA, G., HUNT, K. A., BOCKETT, N. A., ROMANOS, J., MISTRY, V., SZPERL, A., BAKKER, S. F., BARDELLA, M. T., BHAW-ROSUN, L., CASTILLEJO, G., DE LA CONCHA, E. G., DE ALMEIDA, R. C., DIAS, K. R. M., VAN DIEMEN, C. C., DUBOIS, P. C. A., DUERR, R. H., EDKINS, S., FRANKE, L., FRANSEN, K., GUTIERREZ, J., HEAP, G. A. R., HRDLICKOVA, B., HUNT, S., IZURIETA, L. P., IZZO, V., JOOSTEN, L. A. B., LANGFORD, C., MAZZILLI, M. C., MEIN, C. A., MIDAHA, V., MITROVIC, M., MORA, B., MORELLI, M., NUTLAND, S., NUNEZ, C., ONENGUT-GUMUSCU, S., PEARCE, K., PLATTEEL, M., POLANCO, I., POTTER, S., RIBES-KONINCKX, C., RICANO-PONCE, I., RICH, S. S., RYBAK, A., SANTIAGO, J. L., SENAPATI, S., SOOD, A., SZAJEWSKA, H., TRONCONE, R., VARADE, J., WALLACE, C., WOLTERS, V. M., ZHERNAKOVA, A., THELMA, B. K., CUKROWSKA,



- B., URCELAY, E., BILBAO, J. R., MEARIN, M. L., BARISANI, D., BARRETT, J. C., PLAGNOL, V., DELOUKAS, P., WIJMENGA, C., VAN HEEL, D. A., SPANISH CONSORTIUM GENETICS, C., PREVENT, C. D. S. G. & WTCCC 2011. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*, 43, 1193-U45.
- TSOI, L. C., SPAIN, S. L., ELLINGHAUS, E., STUART, P. E., CAPON, F., KNIGHT, J., TEJASVI, T., KANG, H. M., ALLEN, M. H., LAMBERT, S., STOLL, S. W., WEIDINGER, S., GUDJONSSON, J. E., KOKS, S., KINGO, K., ESKO, T., DAS, S., METSPALU, A., WEICHENTHAL, M., ENERBACK, C., KRUEGER, G. G., VOORHEES, J. J., CHANDRAN, V., ROSEN, C. F., RAHMAN, P., GLADMAN, D. D., REIS, A., NAIR, R. P., FRANKE, A., BARKER, J. N., ABECASIS, G. R., TREMBATH, R. C. & ELDER, J. T. 2015. Enhanced meta-analysis and replication studies identify five new psoriasis susceptibility loci. *Nat Commun*, 6, 7001.
- TUOMILEHTO, J. 2013. The emerging global epidemic of type 1 diabetes. *Curr Diab Rep*, 13, 795-804.
- TURSI, A., BRANDIMARTE, G. & GIORGETTI, G. M. 2003. Lack of usefulness of anti-transglutaminase antibodies in assessing histologic recovery after gluten-free diet in celiac disease. *J Clin Gastroenterol*, 37, 387-91.
- UNGPRASERT, P., CARMONA, E. M., UTZ, J. P., RYU, J. H., CROWSON, C. S. & MATTESON, E. L. 2016. Epidemiology of Sarcoidosis 1946-2013: A Population-Based Study. *Mayo Clin Proc*, 91, 183-8.
- UNGPRASERT, P., WIJARNPREECHA, K. & KITTANAMONGKOLCHAI, W. 2017. Psoriasis and Risk of Celiac Disease: A Systematic Review and Meta-analysis. *Indian J Dermatol*, 62, 41-46.
- VAARALA, O. 2008. Leaking gut in type 1 diabetes. *Curr Opin Gastroenterol*, 24, 701-6.
- VAHEDI, K., MASCART, F., MARY, J. Y., LABERENNE, J. E., BOUHNİK, Y., MORIN, M. C., OCMANT, A., VELLY, C., COLOMBEL, J. F. & MATUCHANSKY, C. 2003. Reliability of antitransglutaminase antibodies as predictors of gluten-free diet compliance in adult celiac disease. *Am J Gastroenterol*, 98, 1079-87.
- VAN GERVEN, N. M., BAKKER, S. F., DE BOER, Y. S., WITTE, B. I., BONTKES, H., VAN NIEUWKERK, C. M., MULDER, C. J. & BOUMA, G. 2014. Seroprevalence of celiac disease in patients with autoimmune hepatitis. *Eur J Gastroenterol Hepatol*, 26, 1104-7.
- VAN HEEL, D. A., FRANKE, L., HUNT, K. A., Gwilliam, R., ZHERNAKOVA, A., INOUE, M., WAPENAAR, M. C., BARNARDO, M., BETHEL, G., HOLMES, G. K. T., FEIGHERY, C., JEWELL, D., KELLEHER, D., KUMAR, P., TRAVIS, S., WALTERS, J. R. F., SANDERS, D. S., HOWDLE, P., SWIFT, J., PLAYFORD, R. J., MCLAREN, W. M., MEARIN, M. L., MULDER, C. J., MCMANUS, R., MCGINNIS, R., CARDON, L. R., DELOUKAS, P. & WIJMENGA, C. 2007. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nature Genetics*, 39, 827-829.
- VANGA, R. R. & KELLY, C. P. 2014. Novel therapeutic approaches for celiac disease. *Discov Med*, 17, 285-93.
- VASQUEZ, H., MAZURE, R., GONZALEZ, D., FLORES, D., PEDREIRA, S., NIVELONI, S., SMECUOL, E., MAURINO, E. & BAI, J. C. 2000. Risk of fractures in celiac disease patients: a cross-sectional, case-control study. *Am J Gastroenterol*, 95, 183-9.
- VAZQUEZ-ROQUE MI, CAMILLERI M, SMYRK T, MURRAY JA, MARIETTA E, O'NEILL J, ET AL. 2013 A controlled trial of gluten-free diet in patients with irritable bowel syndrome-diarrhea: effects on bowel frequency and intestinal function. *Gastroenterology*, 144(5):903-11.e3.
- VECSEI, E., STEINWENDNER, S., KOGLER, H., INNERHOFER, A., HAMMER, K., HAAS, O. A., AMANN, G., CHOTT, A., VOGELSANG, H., SCHOENLECHNER, R., HUF, W. & VECSEI, A. 2014. Follow-up of paediatric celiac disease: value of antibodies in predicting mucosal healing, a prospective cohort study. *BMC Gastroenterol*, 14, 28.

- VEHIK, K., HAMMAN, R. F., LEZOTTE, D., NORRIS, J. M., KLINGENSMITH, G. J., REWERS, M. & DABELEA, D. 2008. Trends in high-risk HLA susceptibility genes among Colorado youth with type 1 diabetes. *Diabetes Care*, 31, 1392-6.
- VENTURA, A., MAGAZZU, G. & GRECO, L. 1999. Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. *Gastroenterology*, 117, 297-303.
- VENTURA, A., NERI, E., UGHI, C., LEOPALDI, A., CITTA, A. & NOT, T. 2000. Gluten-dependent diabetes-related and thyroid-related autoantibodies in patients with celiac disease. *J Pediatr*, 137, 263-5.
- VERDU, E. F., GALIPEAU, H. J. & JABRI, B. 2015. Novel players in coeliac disease pathogenesis: role of the gut microbiota. *Nat Rev Gastroenterol Hepatol*, 12, 497-506.
- VILJAMAA, M., KAUKINEN, K., HUHTALA, H., KYRONPALO, S., RASMUSSEN, M. & COLLIN, P. 2005. Coeliac disease, autoimmune diseases and gluten exposure. *Scand J Gastroenterol*, 40, 437-43.
- VILPPULA, A., COLLIN, P., MAKI, M., VALVE, R., LUOSTARINEN, M., KREKELA, I., PATRIKAINEN, H., KAUKINEN, K. & LUOSTARINEN, L. 2008. Undetected coeliac disease in the elderly: a biopsy-proven population-based study. *Dig Liver Dis*, 40, 809-13.
- VILPPULA, A., KAUKINEN, K., LUOSTARINEN, L., KREKELA, I., PATRIKAINEN, H., VALVE, R., LUOSTARINEN, M., LAURILA, K., MAKI, M. & COLLIN, P. 2011. Clinical benefit of gluten-free diet in screen-detected older celiac disease patients. *BMC Gastroenterol*, 11, 136.
- VILPPULA, A., KAUKINEN, K., LUOSTARINEN, L., KREKELA, I., PATRIKAINEN, H., VALVE, R., MAKI, M. & COLLIN, P. 2009. Increasing prevalence and high incidence of celiac disease in elderly people: a population-based study. *BMC Gastroenterol*, 9, 49.
- VISSCHER, T. L., VIET, A. L., KROESBERGEN, I. H. & SEIDELL, J. C. 2006. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity (Silver Spring)*, 14, 2054-63.
- VISSER, J., ROZING, J., SAPONE, A., LAMMERS, K. & FASANO, A. 2009. Tight junctions, intestinal permeability, and autoimmunity: celiac disease and type 1 diabetes paradigms. *Ann N Y Acad Sci*, 1165, 195-205.
- VIVAS, S., RUIZ DE MORALES, J. M., FERNANDEZ, M., HERNANDO, M., HERRERO, B., CASQUEIRO, J. & GUTIERREZ, S. 2008. Age-related clinical, serological, and histopathological features of celiac disease. *Am J Gastroenterol*, 103, 2360-5; quiz 2366.
- VIVAS, S., VAQUERO, L., RODRIGUEZ-MARTIN, L. & CAMINERO, A. 2015. Age-related differences in celiac disease: Specific characteristics of adult presentation. *World J Gastrointest Pharmacol Ther*, 6, 207-12.
- VOLTA, U., BARDAZZI, F., ZAULI, D., DEFRANCESCHI, L., TOSTI, A., MOLINARO, N., GHETTI, S., TETTA, C., GRASSI, A. & BIANCHI, F. B. 1997. Serological screening for coeliac disease in vitiligo and alopecia areata. *Br J Dermatol*, 136, 801-2.
- VOLTA, U., CAIO, G., STANGHELLINI, V. & DE GIORGIO, R. 2014. The changing clinical profile of celiac disease: a 15-year experience (1998-2012) in an Italian referral center. *BMC Gastroenterol*, 14, 194.
- VRIEZINGA, S. L., AURICCHIO, R., BRAVI, E., CASTILLEJO, G., CHMIELEWSKA, A., CRESPO ESCOBAR, P., KOLACEK, S., KOLETZKO, S., KORPONAY-SZABO, I. R., MUMMERT, E., POLANCO, I., PUTTER, H., RIBES-KONINCKX, C., SHAMIR, R., SZAJEWSKA, H., WERKSTETTER, K., GRECO, L., GYIMESI, J., HARTMAN, C., HOGEN ESCH, C., HOPMAN, E., IVARSSON, A., KOLTAI, T., KONING, F., MARTINEZ-OJINAGA, E., TE MARVELDE, C., PAVIC, A., ROMANOS, J., STOOPMAN, E., VILLANACCI, V., WIJMEGA, C., TRONCONE, R. & MEARIN, M. L. 2014. Randomized feeding intervention in infants at high risk for celiac disease. *N Engl J Med*, 371, 1304-15.

- WAHAB, P. J., MEIJER, J. W. & MULDER, C. J. 2002. Histologic follow-up of people with celiac disease on a gluten-free diet: slow and incomplete recovery. *Am J Clin Pathol*, 118, 459-63.
- WAHNSCHAFTE, U., SCHULZKE, J. D., ZEITZ, M. & ULLRICH, R. 2007. Predictors of clinical response to gluten-free diet in patients diagnosed with diarrhoea-predominant irritable bowel syndrome. *Clin Gastroenterol Hepatol*, 5, 844-50; quiz 769.
- WAKIM-FLEMING, J., PAGADALA, M. R., LEMYRE, M. S., LOPEZ, R., KUMARAVEL, A., CAREY, W. D. & ZEIN, N. N. 2013. Diagnosis of celiac disease in adults based on serology test results, without small-bowel biopsy. *Clin Gastroenterol Hepatol*, 11, 511-6.
- WALKER, W. A. 2017. The importance of appropriate initial bacterial colonization of the intestine in newborn, child, and adult health. *Pediatr Res*, 82, 387-395.
- WATANABE, C., KOMOTO, S., HOKARI, R., KURIHARA, C., OKADA, Y., HOZUMI, H., HIGASHIYAMA, M., SAKURABA, A., TOMITA, K., TSUZUKI, Y., KAWAGUCHI, A., NAGAO, S., OGATA, S. & MIURA, S. 2014. Prevalence of serum celiac antibody in patients with IBD in Japan. *J Gastroenterol*, 49, 825-34.
- WEST, J., FLEMING, K. M., TATA, L. J., CARD, T. R. & CROOKS, C. J. 2014. Incidence and prevalence of celiac disease and dermatitis herpetiformis in the UK over two decades: population-based study. *Am J Gastroenterol*, 109, 757-68.
- WEST, J., LOGAN, R. F., CARD, T. R., SMITH, C. & HUBBARD, R. 2004. Risk of vascular disease in adults with diagnosed coeliac disease: a population-based study. *Aliment Pharmacol Ther*, 20, 73-9.
- WESTRA, H. J., PETERS, M. J., ESKO, T., YAGHOOTKAR, H., SCHURMANN, C., KETTUNEN, J., CHRISTIANSEN, M. W., FAIRFAX, B. P., SCHRAMM, K., POWELL, J. E., ZHERNAKOVA, A., ZHERNAKOVA, D. V., VELDINK, J. H., VAN DEN BERG, L. H., KARJALAINEN, J., WITHOFF, S., UITTERLINDEN, A. G., HOFMAN, A., RIVADENEIRA, F., HOEN, P. A. C., REINMAA, E., FISCHER, K., NELIS, M., MILANI, L., MELZER, D., FERRUCCI, L., SINGLETON, A. B., HERNANDEZ, D. G., NALLS, M. A., HOMUTH, G., NAUCK, M., RADKE, D., VOLKER, U., PEROLA, M., SALOMAA, V., BRODY, J., SUCHY-DICEY, A., GHARIB, S. A., ENQUOBAHRIE, D. A., LUMLEY, T., MONTGOMERY, G. W., MAKINO, S., PROKISCH, H., HERDER, C., RODEN, M., GRALLERT, H., MEITINGER, T., STRAUCH, K., LI, Y., JANSEN, R. C., VISSCHER, P. M., KNIGHT, J. C., PSATY, B. M., RIPATTI, S., TEUMER, A., FRAYLING, T. M., METSPALU, A., VAN MEURS, J. B. J. & FRANKE, L. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*, 45, 1238-1243.
- WHYTE, L. A., KOTCHA, S., WATKINS, W. J. & JENKINS, H. R. 2014. Coeliac disease is more common in children with high socio-economic status. *Acta Paediatr*, 103, 289-94.
- WIJMENG, C. & GUTIERREZ-ACHURY, J. 2014. Celiac disease genetics: past, present and future challenges. *J Pediatr Gastroenterol Nutr*, 59 Suppl 1, S4-7.
- WILD, D., ROBINS, G. G., BURLEY, V. J. & HOWDLE, P. D. 2010. Evidence of high sugar intake, and low fibre and mineral intake, in the gluten-free diet. *Aliment Pharmacol Ther*, 32, 573-81.
- WISE, J. 2016. NICE calls for wider use of blood tests to diagnose coeliac disease. *Bmj*, 355, i5660.
- WITHOFF, S., LI, Y., JONKERS, I. & WIJMENG, C. 2016. Understanding Celiac Disease by Genomics. *Trends Genet*, 32, 295-308.
- YANG, A., CHEN, Y., SCHERL, E., NEUGUT, A. I., BHAGAT, G. & GREEN, P. H. 2005. Inflammatory bowel disease in patients with celiac disease. *Inflamm Bowel Dis*, 11, 528-32.
- YOSIPOVITCH, G., DEVORE, A. & DAWN, A. 2007. Obesity and the skin: skin physiology and skin manifestations of obesity. *J Am Acad Dermatol*, 56, 901-16; quiz 917-20.
- ZANWAR, V. G., PAWAR, S. V., GAMBHIRE, P. A., JAIN, S. S., SURUDE, R. G., SHAH, V. B., CONTRACTOR, Q. Q. & RATHI, P. M. 2016. Symptomatic improvement with gluten restriction in irritable bowel syndrome: a prospective, randomized, double blinded placebo controlled trial. *Intest Res*, 14, 343-350.

- ZEVALLOS, V. F., RAKER, V., TENZER, S., JIMENEZ-CALVENTE, C., ASHFAQ-KHAN, M., RUSSEL, N., PICKERT, G., SCHILD, H., STEINBRINK, K. & SCHUPPAN, D. 2017. Nutritional Wheat Amylase-Trypsin Inhibitors Promote Intestinal Inflammation via Activation of Myeloid Cells. *Gastroenterology*, 152, 1100-1113.e12.
- ZINGONE, F., BUCCI, C., TORTORA, R., SANTONICOLA, A., CAPPELLO, C., FRANZESE, M. D., PASSANANTI, V. & CIACCI, C. 2009. Body mass index and prevalence of skin diseases in adults with untreated coeliac disease. *Digestion*, 80, 18-24.
- ZUBILLAGA, P., VIDALES, M. C., ZUBILLAGA, I., ORMAECHEA, V., GARCIA-URKIA, N. & VITORIA, J. C. 2002. HLA-DQA1 and HLA-DQB1 genetic markers and clinical presentation in celiac disease. *J Pediatr Gastroenterol Nutr*, 34, 548-54.

## Appendices

## Appendix I. Template used to extract data from patients' charts.

1. MRN: \_\_\_\_\_

2. Date of Birth: \_\_\_\_\_

3. Gender (please circle where appropriate): Male      Female

4. How old were you when diagnosed with coeliac disease? \_\_\_\_\_

5. Serology at diagnosis: tTG;              Range;              Date;

6. Biopsy at diagnosis & Date;

7. Relatives with coeliac disease;    Yes              No

8. If yes to the previous question please state who in the patient's family has coeliac disease

9. Weight at diagnosis;                      Weight one year after dx;                      Most recent weight;  
Height;

10. Last outpatient visit;

11. Symptoms at diagnosis (Please tick where appropriate):

|                                |  |                   |  |
|--------------------------------|--|-------------------|--|
| Diarrhoea                      |  | Flatulence        |  |
| Pale stools difficult to flush |  | Weight loss       |  |
| Abdominal Pain                 |  | Fatigue           |  |
| Constipation                   |  | Tiredness         |  |
| Vomiting                       |  | Anaemia           |  |
| Heartburn                      |  | Indigestion       |  |
| Bloating                       |  | None of the above |  |

12. Any other symptoms at diagnosis

\_\_\_\_\_

**12. Did the patient ever had any of these conditions?**

|                        |  |                          |  |
|------------------------|--|--------------------------|--|
| Osteoporosis           |  | Dermatitis Herpetiformis |  |
| Osteopenia             |  | Lymphoma                 |  |
| Iron deficiency        |  | Depression               |  |
| B12 deficiency         |  | Ataxia                   |  |
| Folic acid deficiency  |  | Neuropathy               |  |
| Recurrent mouth ulcers |  | Eczema                   |  |
| Reflux oesophagitis    |  | Barrett's oesophagus     |  |

**13. If the patient has osteoporosis or osteopenia;**

**T-score at diagnosis and area affected;**

**Bone Mineral Density at diagnosis for each affected area;**

**Year of diagnosis with osteoporosis/osteopenia;**

**14. Does the patient have any of these conditions?**

|                            |  |                              |  |
|----------------------------|--|------------------------------|--|
| Type 1 Diabetes            |  | Pernicious anaemia           |  |
| Type 2 Diabetes            |  | Vitiligo                     |  |
| Thyroid Disease            |  | Psoriasis                    |  |
| Rheumatoid arthritis       |  | Systemic Lupus Erythematosus |  |
| Inflammatory Bowel Disease |  | Reiter's syndrome            |  |
| Ischaemic Heart Disease    |  |                              |  |

**13. If you ticked any of the conditions above in question 13, how old was the patient when diagnosed with it?**

---

**14. Family history of autoimmune disease?** (Please state condition and relationship i.e. father type 1 diabetes, sister thyroid disease, etc.)

---

**15. If you ticked Thyroid disease in question 13, what type of thyroid disease did you have at diagnosis?**

---

## Appendix II. Questionnaire used during patient visit to CD clinic.

### Association of genotype and phenotype in coeliac disease

The aim of this study is to analyse the clinical presentation of coeliac disease and its relationship to the genotype and gene expression of the affected individuals. Your participation is voluntary and your personal data will be anonymized and kept confidential at all times.

---

1. Name: \_\_\_\_\_

2. Date of Birth: \_\_\_\_\_

3. Gender (please circle where appropriate): Male      Female

4. How old were you when diagnosed with coeliac disease? \_\_\_\_\_

5. When you were diagnosed with coeliac disease were blood tests done to check for antibody levels?

(Please circle where appropriate):      Yes      No

6. Was a biopsy performed to confirm the diagnosis? (Please circle where appropriate)      Yes      No

7. Were your grandparents born in Ireland? (Please circle where appropriate)      Yes      No

8. Does anybody else's in your family have coeliac disease?

(Please circle where appropriate):      Yes      No

9. If you answered yes to the previous question please state who in your family has coeliac disease (please state your relationship i.e. brother, father, etc...)

---

10. When you were diagnosed with coeliac disease which symptoms did you have? (Please tick where appropriate):

|                                |  |                   |  |
|--------------------------------|--|-------------------|--|
| Diarrhoea                      |  | Flatulence        |  |
| Pale stools difficult to flush |  | Weight loss       |  |
| Abdominal Pain                 |  | Fatigue           |  |
| Constipation                   |  | Tiredness         |  |
| Vomiting                       |  | Anaemia           |  |
| Heartburn                      |  | Indigestion       |  |
| Bloating                       |  | None of the above |  |

11. Did you have any other symptoms? (Please state which ones)

---

12. When did you start feeling the symptoms described in questions 10 and 11? (Please state year or age)



---

turn over)

(Please

**12. Did you ever have any of these conditions?** (Please tick where appropriate)

|                        |  |                          |  |
|------------------------|--|--------------------------|--|
| Osteoporosis           |  | Dermatitis Herpetiformis |  |
| Osteopenia             |  | Lymphoma                 |  |
| Iron deficiency        |  | Depression               |  |
| B12 deficiency         |  | Ataxia                   |  |
| Folic acid deficiency  |  | Neuropathy               |  |
| Recurrent mouth ulcers |  | Eczema                   |  |
| Reflux oesophagitis    |  | Barrett's oesophagus     |  |

**13. Do you have any of these conditions?** (Please tick where appropriate)

|                            |  |                              |  |
|----------------------------|--|------------------------------|--|
| Type 1 Diabetes            |  | Pernicious anaemia           |  |
| Type 2 Diabetes            |  | Vitiligo                     |  |
| Thyroid Disease            |  | Psoriasis                    |  |
| Rheumatoid arthritis       |  | Systemic Lupus Erythematosus |  |
| Inflammatory Bowel Disease |  | Reiter's syndrome            |  |
| Ischaemic Heart Disease    |  |                              |  |

**14. If you ticked any of the conditions above in question 13, how old were you when you were diagnosed with it?**

**15. Has anybody in your family been diagnosed with any of the conditions in question 13? (Please state condition and relationship i.e. father type 1 diabetes, sister thyroid disease, etc.)**

---

**16. If you ticked Thyroid disease in question 13, what type of thyroid disease did you have at diagnosis?**

---

**17. If you ticked Thyroid disease in question 13, were blood tests done to check for thyroid antibody levels and were these positive?**

(Please circle where appropriate):    Yes        No

**18. Have you ever smoked?**

(Please circle where appropriate):    Yes        No

**19. In order to clarify aspects of your symptoms or to keep our information relating to coeliac disease up to date, would you allow us to contact you in the future by post or telephone (we envisage a maximum of 1 or 2 calls annually) if required?**

(Please circle where appropriate):    Yes        No

**Thank you for participating in this study.**

This questionnaire is a critical part of our study, which seeks to achieve a better understanding of coeliac disease and its manifestations in order to improve the quality of life of affected individuals.

Appendix III. Supplementary Table 1. The Oslo classification of coeliac disease.

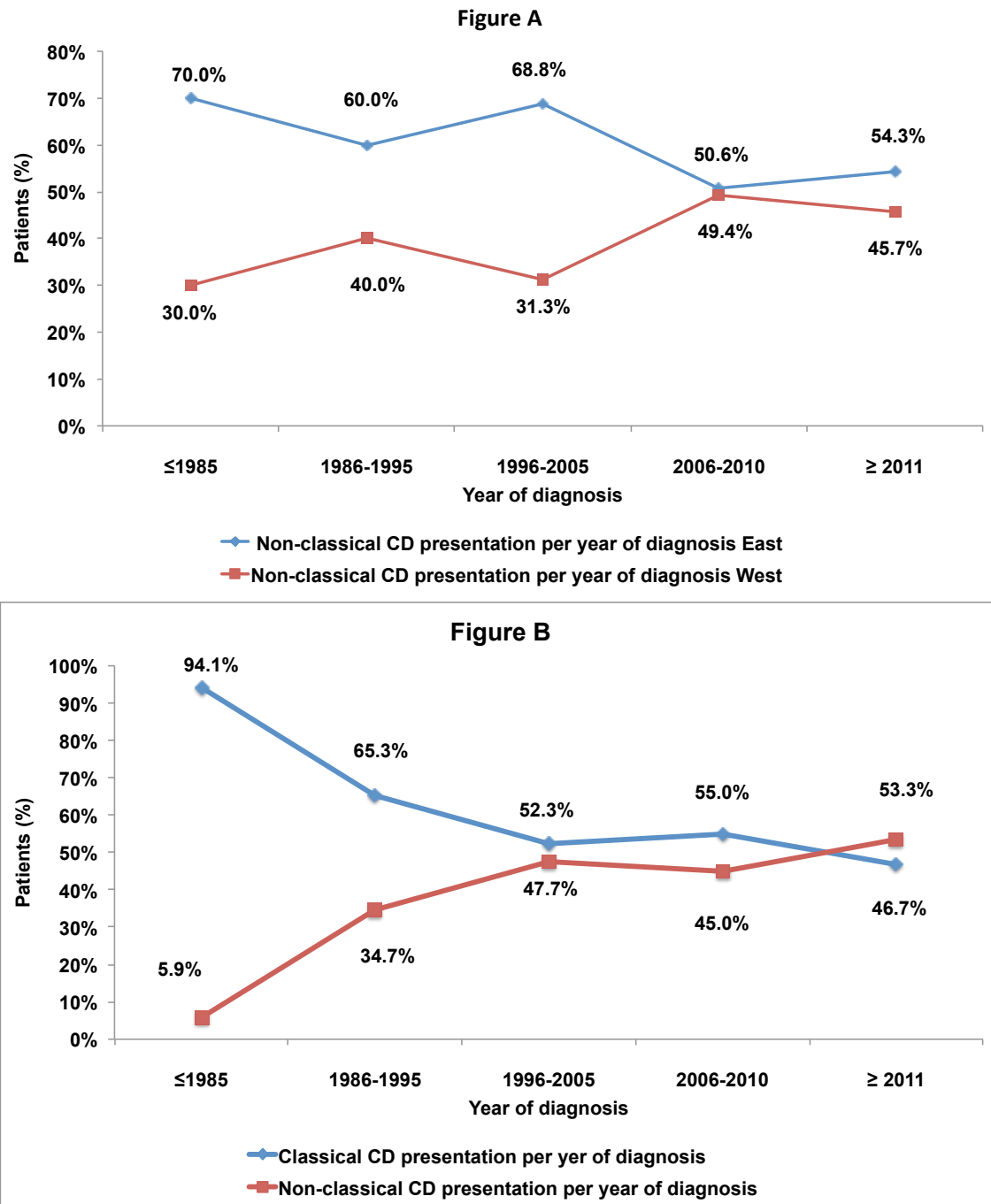
| <b>Oslo Classification of coeliac disease</b> |                                                                                                                                                                              |
|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Coeliac disease presentation</b>           | <b>Characteristics</b>                                                                                                                                                       |
| <b>Classical coeliac disease</b>              | Malabsorption present; diarrhoea, steatorrhoea, weight loss or growth failure required at presentation                                                                       |
| <b>Non-classical coeliac disease</b>          | No signs and symptoms of malabsorption at presentation (i.e. patients with constipation and abdominal pain). Also monosymptomatic disease other than diarrhoea, steatorrhoea |
| <b>Subclinical coeliac disease</b>            | Coeliac disease that is below the threshold of clinical detection                                                                                                            |

**Appendix IV. Supplementary Figure 1. Geographical division of CD patients sample in the present study (n=749).**



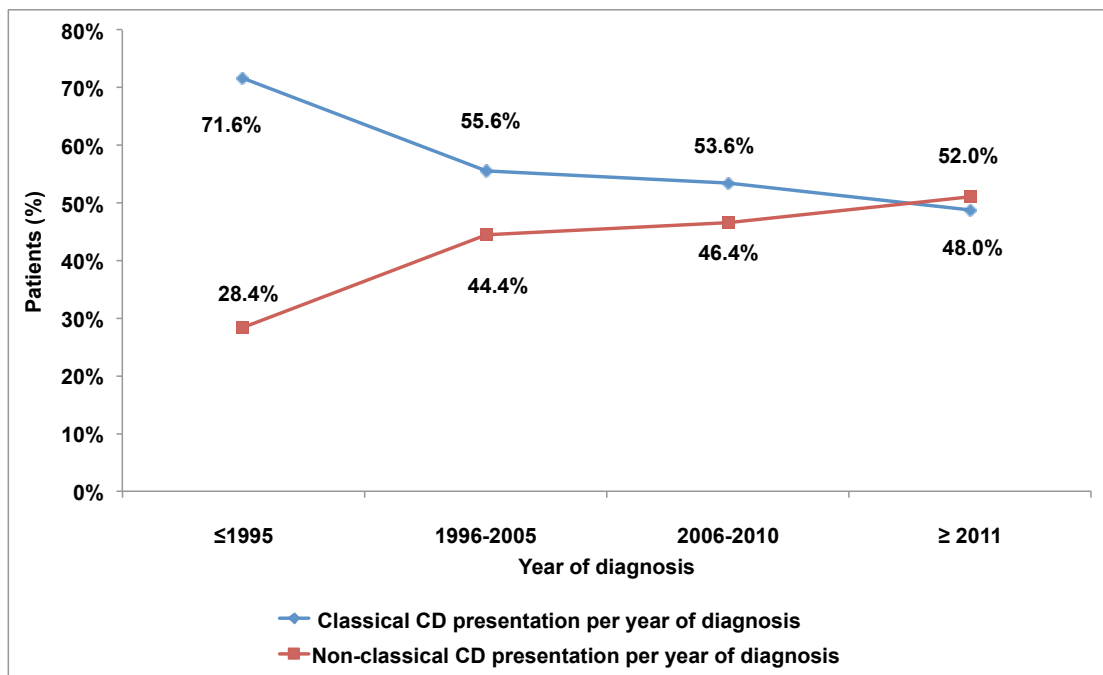
## Appendix V. Supplementary Figure 2

Changes on clinical presentation of CD patients at the time of diagnosis in adulthood diagnosed in Galway University Hospital in the west of Ireland (Figure A) ( $p=0.41$ ), and other participating centres on the east coast (Figure B) ( $p=0.004$ ) over 5 time periods.



### Appendix VI. Supplementary Figure 3

Changes on clinical presentation of CD at the time of diagnosis in adulthood over 4 time periods (n=602).



Subjects diagnosed as adults showed a significant decrease in classical CD presentation and a concomitant increase in non-classical/subclinical presentation over time ( $p=0.006$ ) (Supplementary Figure 4.2).

## Appendix VII. Supplementary Table 2

Gluten free diet adherence scale used by expert coeliac dietitian to classify CD patients compliance with diet a year after diagnosis.

| Expert Dietician Assesment of GFD Adherence |                                                         |
|---------------------------------------------|---------------------------------------------------------|
| <b>1- Excellent;</b>                        | Participant eats gluten less than 3 times per year      |
| <b>2- Good;</b>                             | Participant eats gluten 1 time per month                |
| <b>3- Fair;</b>                             | Participant eats gluten 2-3 times per month             |
| <b>4- Poor;</b>                             | Participant eats gluten 1-2 times per week              |
| <b>5- Very Poor;</b>                        | Participant eats gluten greater than two times per week |
| <b>6- Participant</b>                       | Participant does not follow the gluten free diet        |

## Appendix VIII. Supplementary Table 2

Characteristics of the sample diagnosed in adulthood segregated by the age at diagnosis with coeliac disease; ≥18-35 years, >35-55 years and >55 years.

| Characteristic                                                                  | Median or %†        |                     |                  | P-value          |
|---------------------------------------------------------------------------------|---------------------|---------------------|------------------|------------------|
|                                                                                 | ≥18-35 years, n=165 | >35-55 years, n=282 | >55 years, n=205 |                  |
| <b>Females/Males, n=652 (%)</b>                                                 | 78.8/21.2           | 66.7/33.3           | 60.5/39.5        |                  |
| <b>Clinical presentation at diagnosis**</b> , n 602 (%)                         |                     |                     |                  |                  |
| Classical celiac disease                                                        | 59.3                | 54.8                | 53.5             | 0.55             |
| Non-classical/Subclinical celiac disease                                        | 40.7                | 45.2                | 46.5             |                  |
| <b>Monosymptomatic or subclinical celiac disease</b> , n=652 (%)                |                     |                     |                  |                  |
| <b>Manifestations at diagnosis‡</b> , n 602 (%)                                 |                     |                     |                  |                  |
| Diarrhea                                                                        | 48.3                | 42.1                | 39.9             | 0.285            |
| Steatorrhea                                                                     | 4.8                 | 8.5                 | 6.1              | 0.33             |
| Abdominal Pain                                                                  | 49.0                | 30.5                | 27.8             | <b>&lt;0.001</b> |
| Constipation                                                                    | 22.1                | 11.2                | 10.6             | <b>0.003</b>     |
| Vomiting                                                                        | 10.3                | 5.4                 | 4.0              | <b>0.045</b>     |
| Bloating                                                                        | 37.2                | 23.2                | 18.2             | <b>&lt;0.001</b> |
| Weight Loss                                                                     | 21.4                | 27.4                | 34.3             | <b>0.029</b>     |
| Fatigue/Tiredness                                                               | 34.5                | 33.6                | 26.3             | 0.16             |
| Iron Deficiency Anaemia                                                         | 21.4                | 29.3                | 35.9             | <b>0.015</b>     |
| Other                                                                           | 47.6                | 44.8                | 44.9             | 0.85             |
| <b>Biopsy at diagnosis</b> , n=355 (%)                                          |                     |                     |                  |                  |
| Marsh ≤ 2                                                                       | 6.5                 | 4.7                 | 6.2              | 0.75             |
| Marsh 3a                                                                        | 24.7                | 20.1                | 20.2             |                  |
| Marsh 3b                                                                        | 42.9                | 38.9                | 44.2             |                  |
| Marsh 3c                                                                        | 26.0                | 36.2                | 29.5             |                  |
| <b>Tissue transglutaminase§</b> , n=346 (%)                                     | 2.1                 | 2.0                 | 2.1              | 0.62             |
| <b>Presence of celiac disease extra-intestinal manifestations λ</b> , n 652 (%) |                     |                     |                  |                  |
| Yes                                                                             | 53.3                | 73.0                | 80.0             | <b>&lt;0.001</b> |
| <b>Presence of other immune-mediated conditions</b> , n 652 (%)                 |                     |                     |                  |                  |
| Yes                                                                             | 24.8                | 32.3                | 35.6             | 0.079            |

\* Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into: Classical, Non-classical and Subclinical

‡ Only those diagnosed in adulthood included in analysis

† Percentages provided are valid percentages

§Tissue transglutaminase antibodies represented as log10 of fold increase above cut-off for equivocal result, constant 1 added to remove negative values

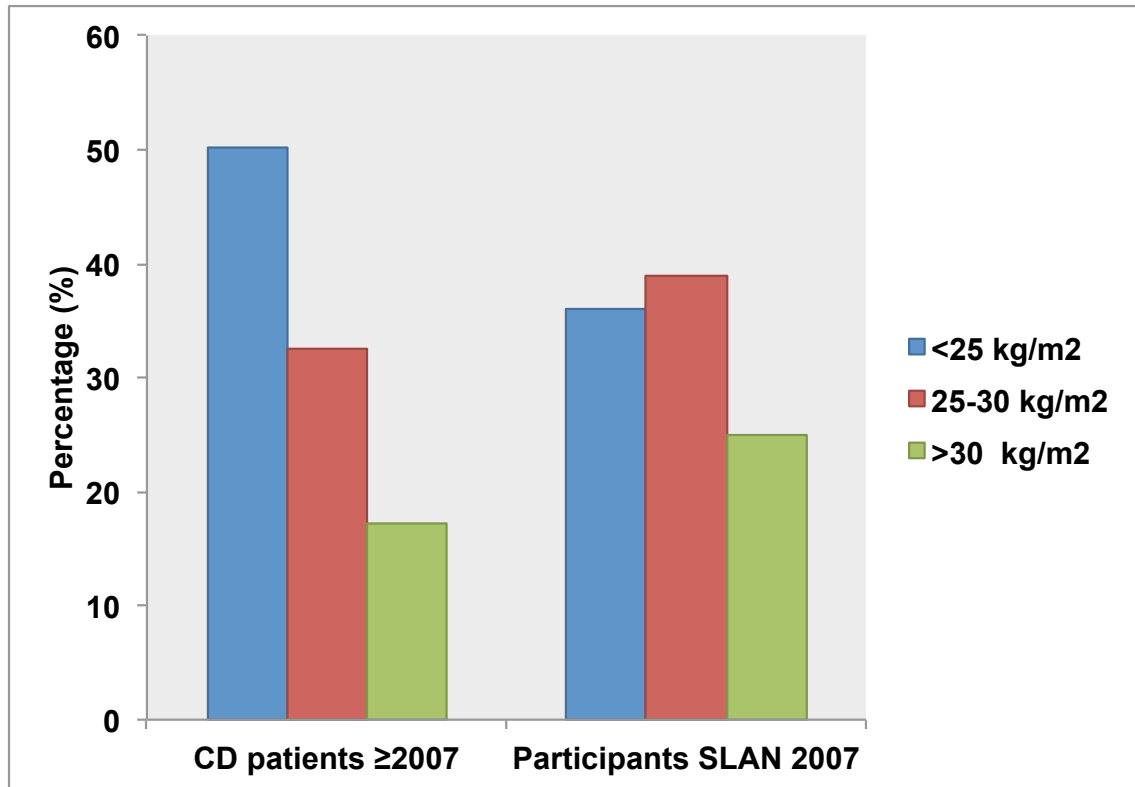
λ Coeliac disease extra-intestinal manifestations considered; Osteoporosis/Osteopenia, Iron deficiency, B12/Folate deficiency and Depression



The female/male ratio reduces as CD is diagnosed later in life being 3.7:1 for those diagnosed between 18-35 years and 1.5:1 for those diagnosed over 55 years. Coeliac disease is also less symptomatic for patients >55 presenting with less abdominal pain, constipation, vomiting and bloating ( $p<0.001$ ;  $p=0.003$ ;  $p=0.045$  and  $p<0.001$  respectively), however this group significantly presents with more weight loss at diagnosis ( $p=0.015$ ). Patients diagnosed over 55 also present significantly more often with iron deficiency anaemia ( $p=0.001$ ).

#### Appendix IX. Supplementary Figure 4.

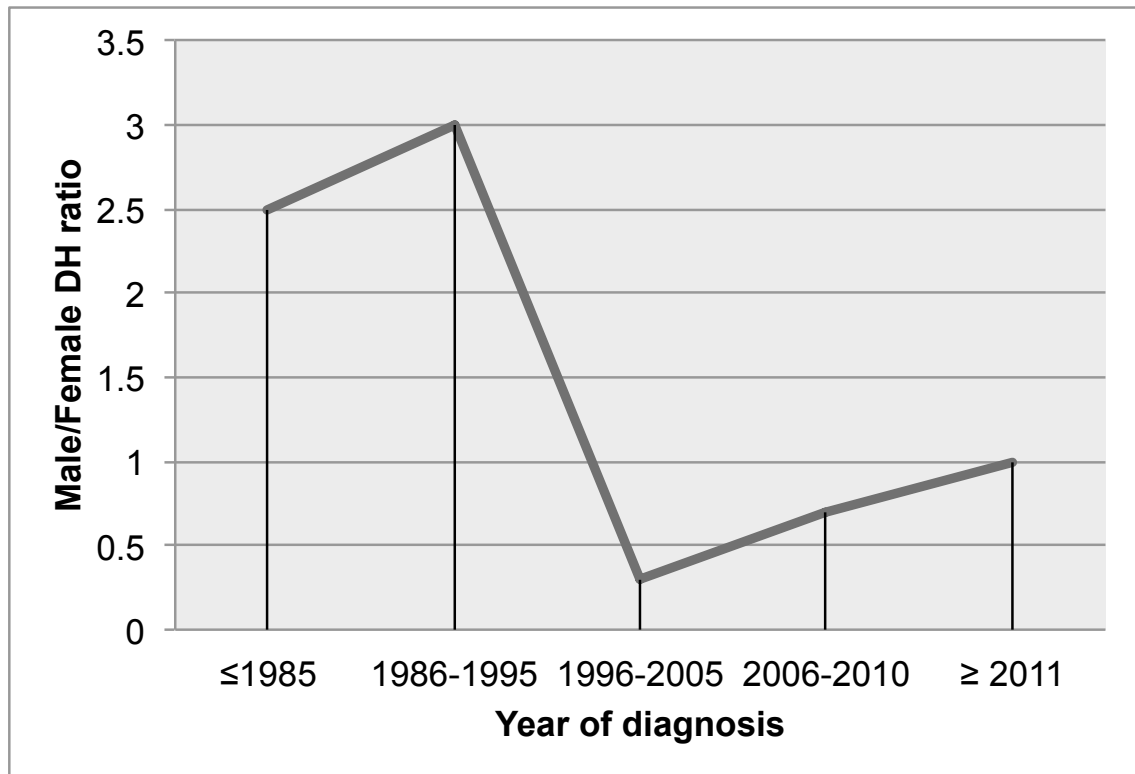
BMI at diagnosis diagnosed  $\geq 2007$  compared to a subcohort of general population data from the survey of lifestyles attitudes and nutrition (Hookway et al.) 2007 with measured BMIs.



When we compared the measured BMIs from this SLAN 2007 subcohort to compare it to the BMIs at diagnosis for our patients diagnosed  $\geq 2007$ , it can be observed that the percentage of SLAN participants over 25 kg/m<sup>2</sup> is 64% in this cohort as opposed to 49.8% in CD patients.

## Appendix X. Supplementary Figure 5.

Male/Female ratio for CD patients diagnosed in adulthood and with co-existent dermatitis herpetiformis over 5 time periods (n=51).



There is a non-significant tendency for the Male/Female ratio of CD patients diagnosed in adulthood and with co-existent DH to decrease over time ( $p=0.082$ ). However, the ratio decreases from 2.5  $\leq 1985$  to 1.0  $\geq 2011$ .

**Appendix XI. Supplementary Table 3. Characteristics of the 436 patients with genotypic information.**

| Characteristic                                        | % <sup>†</sup> | Median (range) |
|-------------------------------------------------------|----------------|----------------|
| <b>Age of diagnosis (years), n 436</b>                |                | 43 (0.5-81.0)  |
| <b>Females/Males, n 436 (%)</b>                       | 69.5/30.5      |                |
| <b>Family history of celiac disease, n 436 (%)</b>    |                |                |
| Yes                                                   | 35.3           |                |
| <b>Diagnosis by age range, n 436 (%)</b>              |                |                |
| < 18 years                                            | 15.4           |                |
| ≥18-35 years                                          | 21.3           |                |
| >35-55 years                                          | 36.2           |                |
| >55                                                   | 27.1           |                |
| <b>Clinical presentation at diagnosis*, n 342 (%)</b> |                |                |
| Classical CD                                          | 57.0           |                |
| Non-classical CD                                      | 43.0           |                |
| <b>% HLA DQ2 Homozygous, n 434</b>                    |                |                |
| Yes                                                   | 14.0           |                |

\* Clinical presentation at diagnosis was classified according to the Oslo definition of celiac disease into; Classical, Non-classical and Subclinical

† Percentages provided are valid percentages

**Appendix XII. Analysis of HLA & Non-HLA (Table A) and HLA only (Table B) and PRS changes over time using a binary time variable.**

**Table A. Analysis of HLA & non-HLA PRS at different p-value levels over time using binary time variable  $\leq 2005$  and  $\geq 2006$ .**

| HLA & Non-HLA PRS    |                                      |                |                   |                    |
|----------------------|--------------------------------------|----------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value                          | t-test p-value | Beta coefficient  | Regression p-value |
|                      | <b><math>5 \times 10^{-8}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.20840                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.20689                              | 0.208          | - 0.060           | 0.208              |
|                      | <b><math>1 \times 10^{-7}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.19924                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.19817                              | 0.362          | - 0.044           | 0.362              |
|                      | <b><math>1 \times 10^{-6}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.17635                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.17545                              | 0.376          | - 0.043           | 0.376              |
|                      | <b><math>1 \times 10^{-5}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.16345                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.16275                              | 0.485          | - 0.034           | 0.453              |
|                      | <b><math>1 \times 10^{-4}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.14342                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.14277                              | 0.298          | - 0.050           | 0.298              |
|                      | <b><math>1 \times 10^{-3}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.10620                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.10584                              | 0.357          | - 0.044           | 0.357              |
|                      | <b>0.01</b>                          |                |                   |                    |
| $\leq 2005$ n=252    | 0.07124                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.07110                              | 0.406          | - 0.040           | 0.406              |
|                      | <b>0.05</b>                          |                |                   |                    |
| $\leq 2005$ n=252    | 0.049265                             |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.04918                              | 0.285          | - 0.051           | 0.285              |
|                      | <b>0.1</b>                           |                |                   |                    |
| $\leq 2005$ n=252    | 0.04146                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.041385                             | 0.164          | - 0.067           | 0.164              |
|                      | <b>0.5</b>                           |                |                   |                    |
| $\leq 2005$ n=252    | 0.02570                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.02566                              | 0.085          | - 0.082           | 0.085              |

1.0\$= Reference group

**Table B. Analysis of HLA only PRS at different p-value levels over time using binary time variable  $\leq 2005$  and  $\geq 2006$ .**

| HLA only PRS         |                                      |                |                   |                    |
|----------------------|--------------------------------------|----------------|-------------------|--------------------|
| Time period mean PRS | PRS p-value                          | t-test p-value | Beta coefficient  | Regression p-value |
|                      | <b><math>5 \times 10^{-8}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.29602                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.29336                              | 0.135          | - 0.072           | 0.135              |
|                      | <b><math>1 \times 10^{-7}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.29424                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.29167                              | 0.146          | - 0.070           | 0.146              |
|                      | <b><math>1 \times 10^{-6}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.28278                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.27995                              | 0.087          | - 0.082           | 0.087              |
|                      | <b><math>1 \times 10^{-5}</math></b> |                |                   |                    |
| $\leq 2005$ n=252    | 0.26886                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.266275                             | 0.128          | - 0.073           | 0.128              |
|                      | <b><math>1 \times 10^{-4}</math></b> |                |                   |                    |
| $\leq 2005$ n=133    | 0.25908                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.25671                              | 0.134          | - 0.072           | 0.134              |
|                      | <b><math>1 \times 10^{-3}</math></b> |                |                   |                    |
| $\leq 2005$ n=133    | 0.25315                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.25103                              | 0.156          | - 0.068           | 0.156              |
|                      | <b>0.01</b>                          |                |                   |                    |
| $\leq 2005$ n=133    | 0.244805                             |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.24264                              | 0.144          | - 0.070           | 0.144              |
|                      | <b>0.05</b>                          |                |                   |                    |
| $\leq 2005$ n=133    | 0.243515                             |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.24129                              | 0.110          | - 0.077           | 0.110              |
|                      | <b>0.1</b>                           |                |                   |                    |
| $\leq 2005$ n=133    | 0.24340                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.24104                              | 0.063          | - 0.089           | 0.063              |
|                      | <b>0.5</b>                           |                |                   |                    |
| $\leq 2005$ n=133    | 0.23098                              |                | 1.0 <sup>\$</sup> |                    |
| $\geq 2006$ n=184    | 0.22890                              | 0.094          | - 0.080           | 0.094              |

1.0<sup>\$</sup>= Reference group

In the previous tables (Fung et al.) it can be observed how the p-values lower when considering our sample as a binary category. A trend towards statistical significance can be seen when considering HLA & non-HLA and HLA only PRS at p-values of 0.1 & 0.5. Statistical significance is reached for non-HLA PRS at a p-level of  $1 \times 10^{-6}$  (Table C).

**Appendix XIII. Supplementary Table 3. HLA SNPs which differ  $\geq 0.04$  in frequency as shown in Figure 4.12.**

| Marker                                                                                                | Freq Diff | PRS weight | r2   | Position | Regulome | Ref | Alt | Location                        | Expression     | Reg motif                           | Proteins bound   | eQTL                                                          | Allele Freq | Conserved |
|-------------------------------------------------------------------------------------------------------|-----------|------------|------|----------|----------|-----|-----|---------------------------------|----------------|-------------------------------------|------------------|---------------------------------------------------------------|-------------|-----------|
| <b>rs6916301</b>                                                                                      | 0.065     | 0.1145     |      | 27074479 | 5        | C   | T   | 35kb 3' of U2                   | -              | 5 altered motifs                    | -                | HIST1H1C, BTN2A1, PRSS16                                      | 0.135       |           |
| <b>rs9357031</b>                                                                                      |           |            | 1    | 27078975 | 2b       | C   | T   | 31kb 3' of U2                   | 21 tissues     | Eomes,Pax-4                         | 7 bound proteins | HIST1H1C, BTN2A1,                                             | 0.11        |           |
| <b>rs477515</b>                                                                                       | 0.064     | 0.2853     |      | 32601914 | 6        | G   | A   | 12kb 5' of HLA-DRB1 within gene | BLD            | HDAC2                               | -                | DQA1, DQB1, DQB1-AS1, DQA2, DQB2, DRB1, DRB6, 250 other       | 0.278       |           |
| <b>rs556025</b>                                                                                       |           |            | 0.82 | 32603103 | 6        | C   | T   | 13kb 5' of HLA-DRB1 within gene | BLD            | BCL,Ets,Sin3Ak-20                   | -                | DQA1, DQB1, DQB1-AS1, DQA2, DQB2, DRB1, DRB6, 180 Other       | 0.325       |           |
| <b>rs2690110</b>     | 0.060     | 0.1152     |      | 25328339 | 6        | A   | G   | LRRC16A within gene             | 9 tissues      | 11 altered motifs                   | -                | LRRC16A                                                       | 0.365       |           |
| <b>rs9380351</b>                                                                                      | 0.059     | 0.0608     |      | 33161969 | 2b       | G   | A   | 711bp 3' of COL11A2             | 22 tissues     | GATA3, PU.1, Elf1, 2 altered motifs | -                | B3GALT4, HLA-DOA, HLA-DPB1, HSD17B8, RING1                    | 0.222       |           |
| <b>rs9394139</b>                                                                                      |           |            | 0.86 | 33143408 | 2b       | C   | T   | 1.4kb 5' of                     | BLD, other     | PRDM1                               | CTCF,POL2,YY1    | HLA-DPB1                                                      | 0.185       |           |
| <b>rs806977</b>                                                                                       | 0.049     | 0.1789     |      | 26189387 | 4        | G   | A   | 310bp 5' of HIST1H4D            | BLD, GI, other | Myb,Myf                             | BCL3,POL2        | BTN3A2, Various histones                                      | 0.451       |           |
| <b>rs1059486</b>                                                                                      |           |            | 0.86 | 26199855 | 4        | G   | A   | HIST1H2BF synonymous            | 48 tissues     | 9 altered motifs                    | 8 bound proteins |                                                               | 0.49        | Yes       |
| <b>rs2235233</b>                                                                                      | 0.047     | 0.2743     |      | 27312073 | 3a       | T   | C   | POM121L2 missense               | 12 tissues     | 4 altered motifs                    | MAX,USF1         | HIST1H2BK                                                     | 0.211       |           |
| <b>rs12174489</b>                                                                                     |           |            | 0.86 | 27320021 | 7        | C   | T   | 7.9kb 5' of POM121L2            | -              | Maf,Mef2,Pou2f2                     | -                | HIST1H2BK                                                     | 0.186       |           |
| <b>rs9261533</b> §                                                                                    | 0.046     | 0.3763     |      | 30153481 | 7        | T   | C   | TRIM10 3'-UTR                   | BLD, GI, LIV   | AP-3,HDAC2,p300                     | -                | HLA-J, HLA-K, IFITM4P, HLA-L, PPP1R11, RPP21, TRIM10, ZNRD1   | 0.251       |           |
| <b>rs929156</b>                                                                                       |           |            | 0.93 | 30171922 | 1f       | G   | A   | TRIM15 missense                 | 14 tissues     | Hsf,PTF1-beta,Smad3                 | CTCF             | 39 hits                                                       | 0.253       |           |
| <b>rs9261528</b>                                                                                      |           |            | 0.96 | 30151652 | 4        | T   | G   | 292bp 3' of TRIM10              | BLD, GI, BONE  | 5 altered motifs                    | GABP             | 39 hits                                                       | 0.254       |           |
| <b>rs9261503</b>                                                                                      |           |            | 0.89 | 30144086 | 3a       | G   | A   | TRIM40 within gene              | GI, LIV, other | CDP,THAP1                           | P300             | 44 hits                                                       | 0.233       |           |
| <b>rs3778638</b>                                                                                      | 0.045     | 0.3342     |      | 31124347 | 7        | G   | A   | PSORS1C1 within gene            | 8 tissues      | -                                   | -                | HLA-C, HLA-B, HCG27, HCG22                                    | 0.19        |           |
| <b>rs3129054</b>                                                                                      | 0.042     | 0.2571     |      | 29681279 | 5        | C   | T   | 168bp 5' of ZFP57               | BLD, GI, other | SEF-1                               | -                | HLA-A, HLA-F, HLA-G ZNRD1,                                    | 0.322       |           |
| <b>rs6917517</b>   | 0.040     | 0.2714     |      | 31114584 | 4        | C   | T   | 165bp 5' of PSORS1C1            | 19 tissues     | Irf,NF-I                            | 7 bound proteins | 70 hits VARSL, HCG27, VARSL, HLA-C, HLA-B, PSORS1C1, PSORS1C2 | 0.183       |           |
| <b>rs13217599</b>  | -0.040    | 0.1954     |      | 27618451 | 4        | T   | C   | 13kb 5' of TRNAI6               | BLD, GI, other | 4 altered motifs                    | 8 bound proteins | -                                                             | 0.072       |           |
| <b>rs6900224</b>                                                                                      | -0.041    | 0.2979     |      | 25437758 | 7        | T   | C   | LRRC16A within gene             | -              | GR                                  | -                | LRRC16A, BTN3A2, TRIM38                                       | 0.234       |           |
| <b>rs4712938</b>                                                                                      |           |            | 0.94 | 25449041 | 4        | C   | T   | LRRC16A within gene             | BLD, other     | Pitx2                               | -                | BTN3A2, TRIM38                                                | 0.229       |           |

### Appendix XIII. Supplementary Table 3 (cont.)

|                     |        |        |               |    |   |     |                               |                     |                            |                                   |                                                           |       |     |
|---------------------|--------|--------|---------------|----|---|-----|-------------------------------|---------------------|----------------------------|-----------------------------------|-----------------------------------------------------------|-------|-----|
| <b>rs9261558</b> \$ | -0.043 | 0.2901 | 30199517      | 6  | T | C   | TRIM26 within gene            | LIV, other          | 6 altered motifs           | -                                 | 66 hits                                                   | 0.177 |     |
| <b>rs3130385</b>    |        |        | 0.83 30211387 | 5  | C | T   | TRIM26 within gene            | BLD, GI, other      | -                          | -                                 | 77 hits; HCG4P5, HLA-A, HLA-J, MICD, HCG4P5               | 0.131 |     |
| <b>rs4148871</b>    | -0.043 | 0.3328 | 32835539      | 5  | G | A   | XXbac-BPG246D15.9 within gene | BLD, other          |                            |                                   | HLA-DOA, HLA-DOB, PSMB9, TAP2                             | 0.158 |     |
| <b>rs1053860</b>    | -0.044 | 0.3192 | 26325000      | 5  | C | T   | 28kb 3' of U6                 | BLD                 | 5 altered motifs           | -                                 | 8 hits                                                    | 0.447 |     |
| <b>rs6917995</b>    |        | 0.84   | 26327586      | 4  | T | C   | 25kb 3' of U6                 | BLD, GI, other      | Ets                        | ETS1, NFKB, P300, POL3, YY1, NFKB | BTN3A2, BTN2A1, HIST1H4H                                  | 0.472 |     |
| <b>rs1091031</b>    | -0.045 | 0.0344 | 24538911      | 7  | G | A   | 1.7kb 3' of ALDH5A1           | -                   | 4 altered motifs           | -                                 | ALDH5A1, GPLD1                                            | 0.485 |     |
| <b>rs4148874</b>    | -0.046 | 0.3845 | 32829711      | 4  | T | C   | XXbac-BPG246D15.9 within gene | IPSC                | 5 altered motifs           | ZNF263                            | HLA-DOB, HLA-DQA2, HLA-DQB2, HLA-DRB1, HLA-DRB5, HLA-DPB2 | 0.188 |     |
| <b>rs9379780</b>    | -0.049 | 0.2631 | 25706943      | 2b | A | G   | 5.2kb 3' of SCGN              | 4 tissues           | HNF4,Hoxa7,SRF             | CEBPB                             | BTN3A2, 7 hits                                            | 0.489 |     |
| <b>rs198809</b>     | -0.049 | 0.2598 | 26128538      | 4  | A | G   | HIST1H2AC                     | 15 tissues          | DEC,Egr-1,Nkx2             | CEBPB                             | BTN3A2, various histones, TRIM38                          | 0.451 |     |
| <b>rs198826</b>     |        | 0.99   | 26120925      | 1d | T | C   | HIST1H2BC                     | BLD, LIV, GI        | 4 altered motifs           | YY1                               | BTN3A2, various histones, TRIM38                          | 0.476 |     |
| <b>rs241420</b>     | -0.051 | 0.6760 | 32855616      | 3a | C | G,T | PSMB9 within gene             | BLD, GI, other      | -                          | -                                 | SKIV2L                                                    | 0.095 |     |
| <b>rs241419</b>     |        |        | 1 32856171    | 5  | G | A,C | PSMB9 missense                | BLD, GI, other      | -                          | -                                 | SKIV2L                                                    | 0.029 | Yes |
| <b>rs6907185</b>    | -0.052 | 0.1972 | 31876907      | 5  | A | G   | SLC44A4 within gene           | ADRL, GI            | BHLHE40,Zic                |                                   | SKIV2L, RDB, EHMT2, C6orf25                               | 0.139 |     |
| <b>rs511294</b>     |        | 0.82   | 31921092      | 2b | T | G   | C2 within gene                | 9 tissues           | RREB-1                     | CEBPB,FOXA1,T CF4                 | SKIV2L, RDBP                                              | 0.067 |     |
| <b>rs7748771</b>    | -0.054 | 0.0929 | 25354871      | 2b | A | G   | LRRC16A within gene           | FAT, GI             | HNF4,RXRA                  | -                                 | -                                                         | 0.102 |     |
| <b>rs78536796</b>   |        | 0.92   | 25370355      | 4  | G | A   | LRRC16A within gene           | BLD, GI, other      | Evi-1                      | -                                 | LRRC16A                                                   | 0.067 |     |
| <b>rs76619442</b>   |        | 0.86   | 25368484      | 5  | G | A   | LRRC16A within gene           | FAT, BLD, GI, other | GATA, GR, HDAC2, AIRE, BCL |                                   | -                                                         | 0.074 | Yes |
| <b>rs2296330</b>    | -0.055 | 0.2376 | 1 33682844    | 1f | G | A   | ITPR3                         | 4 tissues           | Nfkb, EBF, PU.1            | -                                 | ITPR3, UQCC2                                              | 0.165 |     |
| <b>rs2894342</b>    | -0.056 | 0.0488 | 33806617      | 6  | C | A   | 2.6kb 5' of MLN               | BLD                 | NF-AT, PPAR                | -                                 | ITPR3                                                     | 0.271 |     |
| <b>rs73412140</b>   |        | 0.89   | 33789237      | 2a | G | T   | 100bp 5' of LEMD2             | 53 tissues          | 5 altered motifs           | 38 bound proteins                 | ITPR3, UQCC2                                              | 0.268 | Yes |
| <b>rs9467160</b>    | -0.058 | 0.0392 | 24441518      | 4  | G | A   | GPLD1 within gene             | BLD, THY other      | -                          | CEBPB                             | MRS2, ALDH5A1                                             | 0.357 |     |
| <b>rs7762901</b>    |        | 0.92   | 24449411      | 5  | G | A   | GPLD1 within gene             | BLD, GI             | PU.1,Rad21                 | -                                 | MRS2, ALDH5A1                                             | 0.241 |     |
| <b>rs13194984</b>   | -0.063 | 0.1954 | 26500335      | 4  | G | T   | 885bp 5' of BTN1A1            | 4 tissues           | CTCF, IK-2, RAD21          | CTCF                              | BTN3A2, HIST1H4H, HMGN4, ZNF322                           | 0.117 |     |

BLD = Blood; GI= Gastrointestinal; LNG= Lung; CRVX= Cervix; BRN= Brain; LIV=Liver; THY=Thyroid; FAT=Fat tissue; ADRL=adrenal gland

\$ rs9261558 replaces rs116783382

§ rs9261533 replaces rs116285907



## Appendix XIV. Supplementary Table 4. Non-HLA SNPs which differ $\geq 0.04$ in frequency as shown in Figure 4.12.

| Marker     | Freq Diff | r2          | Chromosome | Position    | Regulome | Location           | Expression        | Reg motif               | Proteins bound                         | eQTL                       | MAF  | Constraint Disease interval or SNP   |
|------------|-----------|-------------|------------|-------------|----------|--------------------|-------------------|-------------------------|----------------------------------------|----------------------------|------|--------------------------------------|
| rs1353248  | -0.078    |             | chr3       | 159,623,558 | 7        | SCHIP1 (8.4kb 3')  | BLD, Skin, BRN    | CEBPB,HLF,Hsf           | -                                      | -                          | 0.29 | CD (OR 0.88) MS, PBC, SJO, SSC, SLE  |
| rs62270405 | -0.068    |             | chr3       | 159,594,856 | 7        | SCHIP1 within gene | 4 tissues         | Mef2,PLZF               | -                                      | -                          | 0.31 | CD, MS, PBC                          |
| rs62270407 |           | 1           | chr3       | 159,597,625 | 4        | SCHIP1 within gene | STRM, HRT, MUS    | HNF1,Pax-4              | RAD21                                  | -                          | 0.31 |                                      |
| rs56986069 |           | 0.98        | chr3       | 159,595,142 | 6        | SCHIP1 within gene | BRST, BRN         | 9 altered motifs        | -                                      | -                          | 0.31 |                                      |
| rs58957258 |           | 1           | chr3       | 159,595,383 | 7        | SCHIP1 within gene | 7 tissues         | Foxp1,HDAC2,p300        | -                                      | -                          | 0.31 |                                      |
| rs7651948  |           | 1           | chr3       | 159,597,280 | 7        | SCHIP1 within gene | BLD, 6 tissues    | RFX5                    | -                                      | -                          | 0.31 |                                      |
| rs4682103  | -0.067    |             | chr3       | 112,055,792 | 5        | CD200 within gene  | 12 tissues        | AP-1,HDAC2              | -                                      | n of dehydroisoandrosteron | 0.51 |                                      |
| rs1131199  |           | 0.99        | chr3       | 112,059,767 | 5        | CD200 - Missense   | BLD, 4 tissues    | -                       | -                                      | -                          | 0.51 | Yes                                  |
| rs56132007 | -0.06     |             | chr21      | 45,622,705  | 2b       | AP001057.1         | BLD, GI, other    | CTCF, p300              | 17 bound proteins                      | ICOSLG, Skin               | 0.18 | CD, AS, CRO, IBD, UC                 |
| rs6518350  |           | 1           | chr21      | 45,621,816  | 1f       | AP001057.1         | GI, BLD, other    | BATF                    | -                                      | ICOSLG, Blood, Skin        | 0.18 |                                      |
| rs6518351  |           | 1           | chr21      | 45,624,666  | 1f       | AP001057.1         | BLD, GI, other    | DMRT5                   | -                                      | ICOSLG, Skin               | 0.18 |                                      |
| rs56299324 |           | 1           | chr21      | 45,622,815  | 2b       | AP001057.1         | BLD, GI, other    | ATF3                    | 16 bound proteins                      | ICOSLG, Skin               | 0.18 |                                      |
| rs55965762 |           | 1           | chr21      | 45,622,608  | 4        | AP001057.1         | BLD, GI, other    | FXR, HDAC2, NRSF, Nanog | 12 bound proteins                      | ICOSLG, Skin               | 0.18 |                                      |
| rs9610686  | -0.051    |             | chr22      | 37,633,850  | 5        | RAC2 within gene   | BLD,GI, other     | 5 altered motifs        | -                                      | C1QTNF6, CYTH4, RAC2       | 0.39 | T1D, VIT                             |
| rs5995404  |           | 0.99        | chr22      | 37,632,937  | 2b       | RAC2 within gene   | BLD, 5 other      | PLZF, Pou2f2            | -                                      | C1QTNF6, CYTH4, RAC2       | 0.40 |                                      |
| rs8135343  |           | 0.90        | chr22      | 37,627,844  | 2b       | RAC2 within gene   | GI, BLD, other    | Hic1, LBP9              | -                                      | C1QTNF6, CYTH4, RAC2       | 0.38 |                                      |
| rs739041   |           | 0.82        | chr22      | 37,624,998  | 2b       | RAC2 within gene   | GI, BLD, other    | ERa, AP2, 3 other       | POL2                                   | C1QTNF6, CYTH4, RAC2       | 0.43 |                                      |
| rs2157206  |           | 0.98        | chr22      | 37,636,350  | 3a       | RAC2 within gene   | BLD, GI, other    | MOVO-B, ZBTB33, ZBTB7A  | -                                      | -                          | 0.39 |                                      |
| rs11851414 | -0.05     |             | chr14      | 69,259,501  | 4        | ZFP36L1 intronic   | BLD, GI, other    | Ets, NRSF, TCF12        | POL24H8,POL2                           | -                          | 0.21 | CD (OR 1.13), CRO, IBD, JIA, MS, T1D |
| rs1051533  |           | 1           | chr14      | 69,259,661  | 2b       | ZFP36L1 synonymous | 53 tissues (DHS)  | -                       | POL2, OCT2, TCF12, NFkB/21 others      | -                          | 0.21 | Yes                                  |
| rs72731554 |           | 0.87        | chr14      | 69,260,370  | 2b       | ZFP36L1 intronic   | BLD, GI, other    | PPAR, Znf143            | POL24H8,POL2                           | -                          | 0.19 |                                      |
| rs4902649  |           | 0.93        | chr14      | 69,267,270  | 3a       | ZFP36L1 intronic   | 10 tissues        | 4 altered motifs        | 6 bound proteins                       | -                          | 0.20 |                                      |
| rs12934152 | -0.045    |             | chr16      | 27,403,018  | 4        | IL21R within gene  | BLD, GI, 9 other  | Irf6                    | POL24H8,POL2, PBX3                     | IL21R                      | 0.23 | -                                    |
| rs35646287 |           | 0.98        | chr16      | 27,402,072  | 3a       | IL21R promoter     | BLD, GI, 12 other | Egr1, BCL               | EGR1, ELF1, TCF12, NFkB, PU1, SRF, SP1 | IL21R (5' 89)              | 0.22 |                                      |
| rs270042   | -0.04     |             | chr6       | 127,863,939 | 7        | THEMIS within gene | BLD               | Foxj2, GATA, RXRa       | -                                      | THEMIS                     | 0.62 | CD, MS, CRO, IBD                     |
| rs270033   |           | 0.94        | chr6       | 127,855,808 | 4        | THEMIS within gene | BLD, BRN          | -                       | CTCF, RAD21, SMC3                      | THEMIS                     | 0.62 |                                      |
| rs802734   | -0.04     |             | chr6       | 127,957,653 | 7        | RP11-73O6.4        | BLD, other        | -                       | -                                      | -                          | 0.28 | CD (OR 1.17 3e-14), MS (OR 0.9 6e-9) |
| rs4559105  |           | 0.85        | chr6       | 127,971,247 | 7        | PTPRK within gene  | BLD               | 5 altered motifs        | -                                      | -                          | 0.25 |                                      |
| rs55743914 |           | 0.64 (0.97) | chr6       | 127,972,417 | 4        | PTPRK within gene  | BLD, GI, other    | NRSF                    | -                                      | -                          |      | CD (OR 1.2 1e-18), CRO, MS           |

## Appendix XIV. Supplementary Table 4 (cont.)

|            |           |                         |             |             |                   |                         |                   |                              |                                 |                      |      |                                                                               |
|------------|-----------|-------------------------|-------------|-------------|-------------------|-------------------------|-------------------|------------------------------|---------------------------------|----------------------|------|-------------------------------------------------------------------------------|
| rs4666716  | 0.93      | chr2                    | 182,061,995 | 3a          | AC104820.2 within | BLD, GI                 | 12 altered motifs | -                            | UBE2E3                          | 0.44                 |      |                                                                               |
| rs2061197  | 0.062     | (252 SNPs in LD r2=0.8) | chr3        | 47,001,350  | 5                 | CCDC12 within gene      | BLD, GI, other    | 4 altered motifs             | -                               | CCDC12, SCAP, NBEAL2 | 0.41 | -                                                                             |
| rs7633698  |           | 0.96                    | chr3        | 47,013,452  | 3a                | CCDC12 within gene      | BLD, GI, other    | 12 altered motifs            | EBF1, PU1/9 other               | CCDC12, SCAP, NBEAL2 | 0.59 |                                                                               |
| rs11720139 |           | 0.99                    | chr3        | 47,036,756  | 2b                | NBEAL2 missense         | BLD, GI, other    |                              | -                               | CCDC12, SCAP, NBEAL2 | 0.39 | Yes                                                                           |
| rs1108301  |           | 0.96                    | chr3        | 47,029,425  | 2b                | NBEAL2 within gene      | BLD, GI, other    | 11 altered motifs            | 37 bound proteins               | CCDC12, SCAP, NBEAL2 | 0.41 | Yes                                                                           |
| rs7557448  | 0.054     |                         | chr2        | 62,481,457  | 7                 | B3GNT2 (30kb 3')        | BLD               | -                            | -                               | -                    | 0.19 | AS                                                                            |
| rs77589207 | 0.053     |                         | chr2        | 182,096,324 | 6                 | AC104820.2 within gene  | 5 tissues         | 7 altered motifs             | -                               | CERKL, ITGA4         | 0.09 | CD, AS                                                                        |
| rs11801183 | 0.053     |                         | chr1        | 172,893,844 | 4                 | RP1-15D23.2 within gene | BLD, GI, Other    | AP1, GATA, PPAR3             | -                               | FASLG                | 0.18 | CD, CRO                                                                       |
| rs6684112  |           | 0.87                    | chr1        | 172,897,756 | 2b                | RP1-15D23.2 within gene | BLD, GI, Other    | 4 altered motifs             | POL24H8,CFOS,GATA2              | -                    | 0.17 |                                                                               |
| rs17365327 | 0.052     |                         | chr2        | 182,165,417 | 3a                | AC104820.2              | BLD, GI           | AIRE, Foxp1                  | MAX                             | UBE2E3               | 0.44 | AS, CD                                                                        |
| rs17456058 |           | 0.83                    | chr2        | 182,146,020 | 2c                | AC104820.2 within gene  | BLD, BRN, MUS, GI | Evi1                         | -                               | UBE2E3               | 0.43 | Yes                                                                           |
| rs13410543 |           | 0.91                    | chr2        | 182,166,142 | 5                 | AC104820.2              | BLD               | GATA, HDAC2                  | -                               | UBE2E3               | 0.42 |                                                                               |
| rs10800746 | 0.051     |                         | chr1        | 200,881,391 | 5                 | C1orf106 within gene    | GI                | Crx, NF-kB, Pitx3            | -                               | KIF21B, C1orf106     | 0.28 | CD (OR 0.9 3e-8), AS, CRO, UC, MS                                             |
| rs7554511  |           | 0.89                    | chr1        | 200,877,561 | 4                 | C1orf106 within gene    | GI, ileum         |                              | FOXA1                           | Many genes in ileum  | 0.26 | CD (OR 0.87 e-190), UC (OR 0.85 e-25), MS, CeD                                |
| rs12132298 | 0.89      |                         | chr1        | 200,875,094 | 2b                | C1orf106 within gene    | GI, BLD           | GATA, MAF                    | BATF, NF-kB, EBF1, STAT3        | C1orf106, MROH3P     | 0.26 | Yes                                                                           |
| rs296520   | 0.9       |                         | chr1        | 200,880,978 | 7                 | C1orf106 missense       | GI                | 7 altered motifs             | -                               | C1orf106, MROH3P     | 0.74 | Yes                                                                           |
| rs1323292  | 0.047     |                         | chr1        | 192,571,891 | 4                 | RGS1 (5' 3834)          | BLD, GI           | GATA, Hoxa7                  | -                               | RGS1, RP5-1011O1.2   | 0.81 | MS (OR 0.9 2e-8)                                                              |
| rs1359062  | 1<br>0.97 |                         | chr1        | 192,572,342 | 1f                | RGS1 (5' 3383)          | BLD, GI, other    | -                            | -                               | RGS1, RP5-1011O1.2   | 0.81 | CD (OR 0.77 e-25), MS (OR 0.87 e-20)                                          |
| rs2984920  |           |                         | chr1        | 192,544,794 | 2a                | RGS1 (5' 60)            | BLD, GI           | IRF, PU1                     | 16 bound proteins               | RGS1, RP5-1011O1.2   | 0.81 |                                                                               |
| rs7535818  | 0.93      |                         | chr1        | 192,575,969 | 3a                | RGS1, within gene       |                   | GATA, AIRE, Mef2, Irf, Foxj1 | POLs, POL24H8, USF2, SMC3, WHIP | RGS1, RP5-1011O1.2   | 0.81 |                                                                               |
| rs3184504  | 0.041     |                         | chr12       | 111,884,607 | 3a                | SH2B3 missense          | BLD, GI, other    | HES1, MTF1                   | -                               | SH2B3, ATXN2         | 0.53 | CD (OR 1.2 e-21), RA, CD, JIA (OR 1.2 e-9), PBC, SLE, T1DM (OR 0.8 e-21), CHD |

BLD = Blood; GI= Gastrointestinal; LNG= Lung; CRVX= Cervix; BRN= Brain; MUS= muscle; BRST=Breast

CRO= Crohn's; UC= Ulcerative colitis; MS= Multiple Sclerosis; PBC= Primary biliary cirrhosis; SLE=Systemic Lupus Erythematosus; RA=rheumatoid arthritis; SJO=Sjogren's syndrome; SSC=systemic sclerosis; JIA=juvenile idiopathic arthritis; CHD=coronary heart disease

**Appendix XV. Selected list of abstracts accepted for presentation at national and international meetings.**

- ***“A Cross-sectional Study of the Phenotype in Coeliac Disease in a Large Cohort of Irish Patients”***. Abstract based on preliminary results on clinical phenotype accepted for poster presentation at the Irish society of Gastroenterology (ISG) summer meeting in June 2015. This abstract was also presented at United European Gastroenterology (UEG) meeting in October 2015 in Barcelona. This abstract won best poster presentation at ISG meeting.
- ***“Co-existence of coeliac disease and other auto-immune conditions in a cohort of Irish coeliac patients”***. Abstract based on preliminary results on clinical phenotype accepted for poster presentation at ISG winter meeting in November 2015.
- ***“The Changing Face of Coeliac Disease Presentation; a 55 Year Perspective in Ireland”***. Abstract based on manuscript ‘The Changing Face of Coeliac Disease Presentation; a 55 Year Perspective in Ireland’ accepted for lecture presentation at Digestive Diseases Week (DDW) 2016. Published in *Gastroenterology*; Volume 150, Issue 4, Supplement 1, Pages S205–S206 (April 2016). This abstract was also presented at ISG summer meeting 2016 and won second prize best oral presentation.
- ***“Coeliac disease characteristics per body mass index category at diagnosis and weight changes during the first year”***. Abstract accepted for poster presentation at Irish Society for Clinical Nutrition and Metabolism (IrSPEN) 2017 conference.
- ***“Co-existence of celiac disease and other immune-mediated conditions in celiac patients”***. Abstract accepted for poster presentation at Digestive Diseases Week (DDW) 2017 conference. Published in *Gastroenterology*; Volume 152, Issue 5, Supplement 1, Pages S483–S484 (April 2016).
- ***“Pattern of thyroid disease presentation in a cohort of celiac patients diagnosed from 1960-2015”***. Abstract accepted for poster presentation at Digestive Diseases Week (DDW) 2017 conference. Published in *Gastroenterology*; Volume 152, Issue 5, Supplement 1, page S484 (April 2016).

## Changes in Presentation of Celiac Disease in Ireland From the 1960s to 2015

Patricia Dominguez Castro,<sup>\*</sup> G. Harkin,<sup>‡</sup> M. Hussey,<sup>§</sup> B. Christopher,<sup>||</sup> C. Kiat,<sup>‡</sup> J. Liong Chin,<sup>\*</sup> V. Trimble,<sup>\*</sup> D. McNamara,<sup>§</sup> P. MacMathuna,<sup>||</sup> B. Egan,<sup>#</sup> B. Ryan,<sup>§</sup> D. Kevans,<sup>\*</sup> R. Farrell,<sup>||</sup> V. Byrnes,<sup>‡</sup> N. Mahmud,<sup>\*</sup> and R. McManus<sup>\*</sup>

<sup>\*</sup>Trinity Translational Medicine Institute & Department of Clinical Medicine, Trinity Centre for Health Science, St James's Hospital, Dublin, Republic of Ireland; <sup>‡</sup>Department of Clinical Medicine, University College Hospital Galway, Galway, Republic of Ireland; <sup>§</sup>Department of Clinical Medicine, The Adelaide and Meath Hospital, Dublin, Republic of Ireland; <sup>||</sup>Connolly Hospital Blanchardstown, Blanchardstown, Dublin, Republic of Ireland; <sup>#</sup>Gastrointestinal Unit, Mater Misericordiae University Hospital, Eccles Street, Dublin, Republic of Ireland; and <sup>\*</sup>Department of Clinical Medicine, Mayo General Hospital, Castlebar, County Mayo, Republic of Ireland

**BACKGROUND & AIMS:** Celiac disease is an immune-mediated enteropathy characterized with high heterogeneity in presentation among genetically predisposed individuals. In recent years, a change in the phenotypic presentation of celiac disease has been reported. We studied clinical presentation, from 1960 through 2015, in Ireland, which has a high incidence of celiac disease.

**METHODS:** We performed a retrospective analysis of medical charts from patients diagnosed with celiac disease at 5 secondary referral centers in Ireland from 1960 through 2015 ( $n = 749$ ; median age, 56 years; age range, 18–91 years). The cohort was divided into 5 groups based on year of diagnosis ( $\leq 1985$ , 1986–1995, 1996–2005, 2006–2010, or 2011 and later). We collected findings from clinical presentation at diagnosis; serology tests; small intestinal biopsy analyses; and patients' demographic, clinical, and family data. Presentations at diagnosis were classified according to the Oslo criteria as follows: classical (patients presenting with malabsorption), nonclassical (no signs or symptoms of malabsorption at presentation), or subclinical (below the threshold of clinical detection). The primary outcome was change in clinical presentation of celiac disease over time.

**RESULTS:** Of the 749 patients studied, 512 were female and 237 were male (ratio of 2.2:1). Female patients were diagnosed at younger ages than male patients (42 vs 47 years, respectively;  $P = .004$ ), and had more immune-mediated conditions than male patients (35.7% for female patients vs 21.5% for male patients;  $P < .001$ ). For patients diagnosed as adults (after the age of 18 years), the median age of diagnosis increased from 34.0 years during the period  $\leq 1985$  to median ages of 44–46 years after 1985 ( $P < .002$ ). A smaller proportion of patients presented with classical features of celiac disease after 2010 (48.4%) than  $\leq 1985$  (85.2%); the proportion of patients with nonclassical or subclinical celiac disease increased from 14.8%  $\leq 1985$  to 51.6% after 2010 ( $P = .006$  for each). Biopsies categorized as Marsh 3c decreased, from 52.2% in the period 1996–2005 to 22.5% in the period after 2010 ( $P = .003$ ). The prevalence of associated thyroid disease has decreased during the study period, from 36.6%  $\leq 1985$  to 17.1% after 2010 ( $P = .039$ ), whereas body mass index at diagnosis increased from 21.5 kg/m<sup>2</sup>  $\leq 1985$  to 24.8 kg/m<sup>2</sup> after 2010 ( $P < .001$ ).

**CONCLUSIONS:** We found the clinical presentation of celiac disease changed significantly in Ireland from 1960 through 2015. The age of presentation in adulthood increased over this time period, as did the proportions of patients with nonclassical or subclinical disease.

**Keywords:** BMI; Autoimmune Disease; Epidemiology; Environmental Factors; Gluten.

**Abbreviations used in this paper:** BMI, body mass index; CD, celiac disease; ELISA, enzyme-linked immunosorbent assay; GFD, gluten-free diet; IMC, immune-mediated condition; T1DM, type 1 diabetes mellitus; TD, thyroid disease; TTGA, anti-tissue transglutaminase antibodies.

© 2017 by the AGA Institute

1542-3565/\$36.00

<http://dx.doi.org/10.1016/j.cgh.2016.11.018>

Celiac disease (CD) is a chronic small intestinal, immune-mediated enteropathy, precipitated by exposure to dietary gluten in genetically predisposed individuals.<sup>1</sup> Once considered to be a disorder characterized by overt malabsorptive symptoms and confined mainly to childhood, serologic screening studies have proven that CD occurs in both pediatric and adult populations at a similar prevalence of approximately 1%.<sup>2</sup> Recent studies demonstrate that increased CD prevalence has been accompanied by a considerable increase in its incidence rates over time.<sup>3-5</sup>

It is increasingly clear that CD has a wide clinical spectrum, from classical or overt presentation with malabsorptive gastrointestinal symptoms to subclinical forms of the disease.<sup>2-4</sup> Gender differences in disease presentation and prognosis have also been reported.<sup>3,6</sup> Changes in the features of CD have been reported in recent years, with increasing prevalence of nonclassical and subclinical presentation at diagnosis, and rising median age of diagnosis in both the pediatric and adult CD population.<sup>7,8</sup> These changes may be caused by increased awareness paired with improved diagnostic techniques; however, environmental factors have also been suggested to play a role.<sup>9,10</sup>

Patients with CD have a predisposition to develop other immune-mediated conditions (IMC), notably thyroid disease (TD) and type 1 diabetes mellitus (T1DM). Sharing of genetic susceptibility variants presumably contributes to the high prevalence of these coexistent IMCs in CD.<sup>11</sup> However, gene-environment interactions likely also contribute, as highlighted by studies suggesting a protective effect of the gluten-free diet (GFD) on comorbid IMCs.<sup>12</sup> Therefore, a timely CD diagnosis may be important in improving both patient quality of life and disease prognosis.<sup>13</sup> However, diverse studies continue to show high proportions of undiagnosed patients and long delays in diagnosis, sometimes over 10 years from the onset of symptoms.<sup>14</sup> Accurate, up-to-date profiles of clinical presentation and prognosis are desirable to improve understanding and recognition of CD by patients and physicians.

Studies describing the clinical phenotype and disease evolution in large, well-characterized cohorts of CD patients are few.<sup>15</sup> There is a particular paucity of data regarding current patterns of presentation in Ireland, traditionally an area of high diagnosis because of high awareness of the condition since the 1970s.<sup>16-18</sup> The aims of the present study were to carry out a cross-sectional analysis of the clinical characteristics of the disease in a cohort of patients with CD (n = 749) attending secondary referral centers in Ireland, with a view to documenting the evolution of the condition over a 55-year time period in a country with relatively high awareness of the condition.

## Methods

### Study Design and Study Population

A retrospective analysis of medical records was performed using a CD patient cohort (n = 749) attending 5 secondary referral centers in Ireland and diagnosed from 1960–2015. Patients with suspected CD are normally referred to secondary referral centers with expertise in CD diagnosis and management; our patients are therefore likely to be a relatively unselected and representative sample. Most of the patients were identified based on their participation in a previous study on the genetics of CD. All patients are white and of Irish ancestry. A total of 79% of patients were diagnosed in the 5 participating referral centers with the remainder diagnosed at other secondary referral centers. Data were collected from medical records, hospital information systems, and a structured questionnaire completed by the patient and the physician or research nurse for patients currently attending the celiac clinic. All patients gave written informed consent. CD was diagnosed by serology and histology following current guidelines.<sup>19</sup>

Serology included antigliadin antibodies for patients diagnosed during the 1960s to the 1990s; anti-tissue transglutaminase antibodies (tTGA) were introduced in the late 1990s and have since been the first-line screening method for CD diagnosis together with anti-endomysial antibodies. Small intestinal biopsies were taken from the jejunum or the second duodenal portion, and most were classified according to the modified Marsh classification by the pathologist. Where the degree of damage to the bowel mucosa was provided as partial, subtotal, and total villous atrophy, these were reclassified as Marsh 3a, Marsh 3b, and Marsh 3c, respectively.<sup>20</sup> Sixteen patients with abnormal tTGA and/or endomysial antibodies and Marsh 1 biopsy were included based on improvement on the GFD and serology normalization.<sup>19</sup> Two patients with abnormal antigliadin antibodies and Marsh 1 biopsy were included in the analysis on the basis of coexistent dermatitis herpetiformis, family history of CD, and improvement on the GFD with serology normalization.<sup>20</sup> Fourteen seronegative patients with no IgA deficiency but abnormal biopsy at diagnosis were included in the analysis based on evidence of improvement on the GFD and/or biopsy.<sup>21</sup> The study was approved by the Research Ethics Committees of St James's Hospital Dublin and all other contributing centers.

### Data Collection and Definition of Study Variables

The following information was collected: age, gender, age at diagnosis, symptoms at diagnosis, serology and



histology data at diagnosis, ethnicity, and family history of CD and of other IMCs. Patient age was considered to be the age at last follow-up. The following manifestations at diagnosis were collected: diarrhea, steatorrhea, abdominal pain, constipation, vomiting, heartburn or indigestion, bloating, flatulence, weight loss, fatigue or tiredness, and iron deficiency anemia. A category classified as "other manifestations" was also created for signs and symptoms present at diagnosis that had not been included initially.

The clinical presentation at diagnosis was classified according to the Oslo classification as classical, nonclassical, and subclinical using the previously recorded manifestations (Supplementary Table 1).<sup>1</sup> Patients diagnosed <18 years of age were excluded from analyses relating to clinical presentation over time because of the small sample size and information deficits on clinical presentation at diagnosis. Manifestations at diagnosis were available for 602 patients diagnosed in adulthood.

The total prevalence of the following CD extraintestinal manifestations and coexistent IMCs was determined in the study cohort: bone disorders (osteoporosis and/or osteopenia), iron deficiency anemia, vitamin deficiency (B<sub>12</sub> and/or folic acid deficiency), lymphoma, depression, ataxia, neuropathy, dermatitis herpetiformis, T1DM, TD, rheumatoid arthritis, inflammatory bowel disease, pernicious anemia, vitiligo, psoriasis, systemic lupus erythematosus, and Reiter syndrome. A category "other" was created to include any IMCs that had not been included initially.<sup>22,23</sup>

#### Assays Used to Determine Celiac Serology

During the time period covered by this study, tTGA value calculations were determined using a variety of tTGA diagnostic tests with different cutoff points. To compare results from different tests, tTGA values were expressed as the fold increase above the cutoff value by calculating the ratio of the observed tTGA at diagnosis divided by the cutoff value for an equivocal result.<sup>24</sup> The resulting values were corrected for normality by log transformation, and a value of 1 was added to remove negative values.<sup>25</sup> To test the resulting tTGA variable we correlated it with the degree of mucosal damage at diagnosis (Marsh criteria) and observed a significant correlation as reported in previous studies (Supplementary Figure 1).<sup>24</sup> tTGA values at diagnosis were available for 346 patients diagnosed in adulthood; 113 of these were analyzed using Phadia Celikey enzyme-linked immunosorbent assay (ELISA) with anti-tTGA titer of 1.9–5 U/mL for an equivocal result, 111 used Thermo Fisher ELISA assay 5–10 U/mL, 59 used Thermo Fisher Celikey ELISA assay 4–10 U/mL, 38 used Thermo Fisher Celikey ELISA 7–10 U/mL, and 25 used Phadia Celikey ELISA assay 2–8 U/mL.

#### Statistical Analysis

A categorical variable was created for the year of diagnosis as follows: (1) ≤1985, (2) 1986–1995, (3) 1996–2005, (4) 2006–2010, and (5) ≥2011. These periods are similar to Kivela et al<sup>8</sup> and were selected to capture potential changes in CD presentation and associated complications, and the introduction of new diagnostic procedures. Thus before 1985, all biopsies were jejunal capsule biopsies, which were widely replaced by duodenal pinch biopsies from 1986. Endomysial antibodies and tTGAs were introduced in 1994 and 1997, respectively, substituting previously used antigliadin antibodies.<sup>8</sup>

Statistical analysis was carried out using SPSS version 21.0 (SPSS, Inc, Chicago, IL). Discrete variables are presented as percentages and continuous variables as medians and ranges. The relationship between discrete variables was analyzed using chi-square test. Differences between groups were assessed using Mann-Whitney *U* test and Kruskal-Wallis nonparametric test. The Jonckheere-Terpstra test was used to carry out the trend analysis. A threshold *P* value of < .05 was considered significant in all analyses.

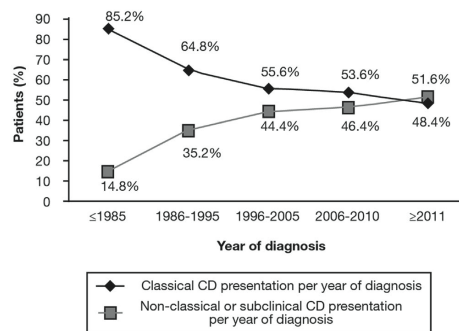
#### Results

##### Characteristics of the Sample

The median age at data collection was 56 years (range, 18–91). A total of 97 patients were diagnosed <18 years of age, and 652 were diagnosed in adulthood. A higher prevalence of female patients was found (female/male ratio, 2.2:1). Classical CD was the most common presentation at diagnosis (55.6%). Diarrhea and fatigue/tiredness were the most commonly reported intestinal and extraintestinal symptoms (42.9%) and (31.4%), respectively (Table 1). A total of 40.9% of patients presented either overweight or obese at diagnosis (Table 1), and showed a tendency to be diagnosed later in life (51 years) compared with those who presented underweight or normal weight (48 years) (*P* = .072). A total of 66.1% of patients had a CD-associated disorder and 31.2% had a coexistent IMC. The most prevalent IMCs were TD (19.9%) and T1DM (3.6%), and the most prevalent CD-associated disorder was bone disorders (40.9%) (Supplementary Table 2).

##### Characteristics of the Sample Segregated by Gender

Female patients were diagnosed earlier in life compared with male patients (*P* = .002) (Table 2). Constipation, bloating, and iron deficiency anemia were more common in female patients than in male patients (*P* = .002, *P* = .009, and *P* = .026, respectively).



**Figure 1.** Changes in CD clinical presentation at the time of diagnosis in adulthood over 5 time periods (n = 602).

commonly reported before 1996, but decreased significantly in the period from 1996 to 2005, with continued decrease thereafter ( $P = .001$ ). Milder damage to the small bowel mucosa ( $P < .001$ ) and a significant decrease in the prevalence of associated TD ( $P = .039$ ) is observed over the study period (Table 3). The age at diagnosis of CD was significantly lower before 1986 compared with later periods ( $P < .002$ ) (Figure 2). The median body mass index (BMI) at diagnosis increased from 21.6 kg/m<sup>2</sup> before 1986, to near 25.0 kg/m<sup>2</sup> in the later periods ( $P < .001$ ) (Figure 3).

## Discussion

This study describes the clinical features of CD in a large cohort, and represents the only study of its kind in Ireland. Covering a period from 1960 to 2015, it provides a valuable 55-year perspective on the clinical features of the condition in a region where it has historically has been well recognized and frequently diagnosed.

Although our sample confirms that CD can be diagnosed over a wide age range, a significant increase in the median age at diagnosis happened between 1986 and 2005 coinciding with improved diagnostic techniques.<sup>2,7,8,15</sup> The median age at diagnosis seems to have plateaued thereafter in the late 40's, which is higher than what has been reported in similar studies internationally, but in line with previous studies in Ireland.<sup>7,15,17,18</sup> A continuous gradual decrease in classical presentation mirrored by an increase in nonclassical/subclinical presentation was observed over time, corroborating previous findings.<sup>7</sup>

Concurrently, a less severe damage to the small bowel mucosa is observed at diagnosis in the later time periods, which, as expected, is reflected in lower tTGA levels.<sup>24</sup> Comparing biopsies and serology at diagnosis, adults have a less severe presentation and a higher prevalence of nonclassical symptoms compared with children.<sup>6,7</sup> Taken together, this may imply that adults with unspecific or nonclassical symptoms are being finally diagnosed with CD later in life because of better

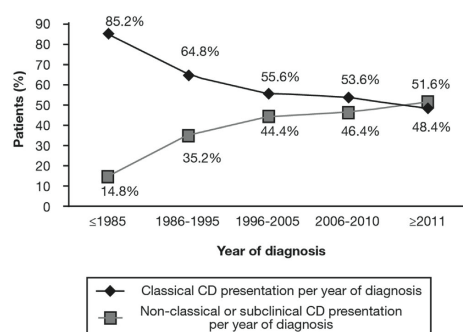
**Table 3.** Clinical Characteristics at the Time of CD Diagnosis in Adulthood and Coexistence of IMC Over 5 Time Periods (n = 652)

| Characteristic <sup>a</sup>                       | ≤1985,<br>n = 41 | 1986-1995,<br>n = 69 | 1996-2005,<br>n = 174 | 2006-2010,<br>n = 239 | ≥2011,<br>n = 128 | P<br>value |
|---------------------------------------------------|------------------|----------------------|-----------------------|-----------------------|-------------------|------------|
| Gender                                            |                  |                      |                       |                       |                   |            |
| Male                                              | 39.0             | 34.8                 | 26.4                  | 32.2                  | 36.4              |            |
| Female                                            | 61.0             | 65.2                 | 73.6                  | 67.8                  | 63.6              | .309       |
| Symptoms at diagnosis (n = 602)                   |                  |                      |                       |                       |                   |            |
| Diarrhea                                          | 70.4             | 46.3                 | 41.9                  | 42.9                  | 36.7              | .031       |
| Weight loss                                       | 51.9             | 44.4                 | 28.1                  | 26.2                  | 20.3              | .001       |
| Constipation                                      | 3.7              | 9.3                  | 13.1                  | 13.7                  | 17.2              | .333       |
| Abdominal pain                                    | 48.1             | 31.5                 | 32.5                  | 36.1                  | 30.5              | .424       |
| Overweight or obese (n = 391)                     | 13.3             | 13.3                 | 34.0                  | 50.0                  | 48.8              | < .001     |
| Presence of coexistent immune-mediated conditions | 46.3             | 33.3                 | 31.6                  | 31.4                  | 26.4              | .209       |
| Thyroid disease                                   | 36.6             | 26.1                 | 20.1                  | 17.6                  | 17.1              | .039       |
| Type 1 diabetes mellitus                          | 2.4              | 1.4                  | 4.6                   | 4.2                   | 3.9               | .802       |
| Psoriasis                                         | 4.9              | 2.9                  | 2.3                   | 2.1                   | 2.3               | .875       |
| Biopsy at diagnosis (n = 348)                     |                  | 1986-1995            | 1996-2005             | 2006-2010             | ≥2011             |            |
| Marsh ≤2                                          | ND <sup>b</sup>  | 8.7                  | 3.7                   | 4.5                   | 9.0               |            |
| Marsh 3a                                          | ND               | 13.0                 | 9.8                   | 27.3                  | 23.6              |            |
| Marsh 3b                                          | ND               | 26.1                 | 42.7                  | 43.5                  | 44.9              |            |
| Marsh 3c                                          | ND               | 52.2                 | 43.9                  | 24.7                  | 22.5              | .003       |
| Tissue transglutaminase <sup>b</sup> (n = 346)    | ND               | ND                   | 2.11                  | 2.07                  | 1.99              | .019       |

ND, no or insufficient data.

<sup>a</sup>Valid percentage provided for each characteristic category.

<sup>b</sup>Tissue transglutaminase antibodies represented as log10 of fold increase above cutoff for equivocal result, constant 1 added to remove negative values.



**Figure 1.** Changes in CD clinical presentation at the time of diagnosis in adulthood over 5 time periods (n = 602).

commonly reported before 1996, but decreased significantly in the period from 1996 to 2005, with continued decrease thereafter ( $P = .001$ ). Milder damage to the small bowel mucosa ( $P < .001$ ) and a significant decrease in the prevalence of associated TD ( $P = .039$ ) is observed over the study period (Table 3). The age at diagnosis of CD was significantly lower before 1986 compared with later periods ( $P < .002$ ) (Figure 2). The median body mass index (BMI) at diagnosis increased from 21.6 kg/m<sup>2</sup> before 1986, to near 25.0 kg/m<sup>2</sup> in the later periods ( $P < .001$ ) (Figure 3).

## Discussion

This study describes the clinical features of CD in a large cohort, and represents the only study of its kind in Ireland. Covering a period from 1960 to 2015, it provides a valuable 55-year perspective on the clinical features of the condition in a region where it has historically has been well recognized and frequently diagnosed.

Although our sample confirms that CD can be diagnosed over a wide age range, a significant increase in the median age at diagnosis happened between 1986 and 2005 coinciding with improved diagnostic techniques.<sup>2,7,8,15</sup> The median age at diagnosis seems to have plateaued thereafter in the late 40's, which is higher than what has been reported in similar studies internationally, but in line with previous studies in Ireland.<sup>7,15,17,18</sup> A continuous gradual decrease in classical presentation mirrored by an increase in nonclassical/subclinical presentation was observed over time, corroborating previous findings.<sup>7</sup>

Concurrently, a less severe damage to the small bowel mucosa is observed at diagnosis in the later time periods, which, as expected, is reflected in lower tTGA levels.<sup>24</sup> Comparing biopsies and serology at diagnosis, adults have a less severe presentation and a higher prevalence of nonclassical symptoms compared with children.<sup>6,7</sup> Taken together, this may imply that adults with unspecific or nonclassical symptoms are being finally diagnosed with CD later in life because of better

**Table 3.** Clinical Characteristics at the Time of CD Diagnosis in Adulthood and Coexistence of IMC Over 5 Time Periods (n = 652)

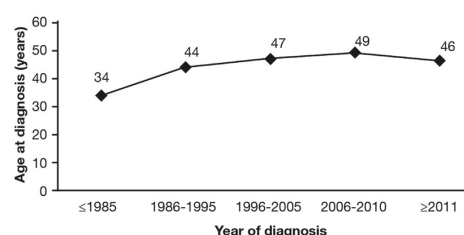
| Characteristic <sup>a</sup>                       | ≤1985,<br>n = 41 | 1986-1995,<br>n = 69 | 1996-2005,<br>n = 174 | 2006-2010,<br>n = 239 | ≥2011,<br>n = 128 | P<br>value |
|---------------------------------------------------|------------------|----------------------|-----------------------|-----------------------|-------------------|------------|
| Gender                                            |                  |                      |                       |                       |                   |            |
| Male                                              | 39.0             | 34.8                 | 26.4                  | 32.2                  | 36.4              |            |
| Female                                            | 61.0             | 65.2                 | 73.6                  | 67.8                  | 63.6              | .309       |
| Symptoms at diagnosis (n = 602)                   |                  |                      |                       |                       |                   |            |
| Diarrhea                                          | 70.4             | 46.3                 | 41.9                  | 42.9                  | 36.7              | .031       |
| Weight loss                                       | 51.9             | 44.4                 | 28.1                  | 26.2                  | 20.3              | .001       |
| Constipation                                      | 3.7              | 9.3                  | 13.1                  | 13.7                  | 17.2              | .333       |
| Abdominal pain                                    | 48.1             | 31.5                 | 32.5                  | 36.1                  | 30.5              | .424       |
| Overweight or obese (n = 391)                     | 13.3             | 13.3                 | 34.0                  | 50.0                  | 48.8              | < .001     |
| Presence of coexistent immune-mediated conditions | 46.3             | 33.3                 | 31.6                  | 31.4                  | 26.4              | .209       |
| Thyroid disease                                   | 36.6             | 26.1                 | 20.1                  | 17.6                  | 17.1              | .039       |
| Type 1 diabetes mellitus                          | 2.4              | 1.4                  | 4.6                   | 4.2                   | 3.9               | .802       |
| Psoriasis                                         | 4.9              | 2.9                  | 2.3                   | 2.1                   | 2.3               | .875       |
| Biopsy at diagnosis (n = 348)                     |                  | 1986-1995            | 1996-2005             | 2006-2010             | ≥2011             |            |
| Marsh ≤2                                          | ND <sup>b</sup>  | 8.7                  | 3.7                   | 4.5                   | 9.0               |            |
| Marsh 3a                                          | ND               | 13.0                 | 9.8                   | 27.3                  | 23.6              |            |
| Marsh 3b                                          | ND               | 26.1                 | 42.7                  | 43.5                  | 44.9              |            |
| Marsh 3c                                          | ND               | 52.2                 | 43.9                  | 24.7                  | 22.5              | .003       |
| Tissue transglutaminase <sup>b</sup> (n = 346)    | ND               | ND                   | 2.11                  | 2.07                  | 1.99              | .019       |

ND, no or insufficient data.

<sup>a</sup>Valid percentage provided for each characteristic category.

<sup>b</sup>Tissue transglutaminase antibodies represented as log10 of fold increase above cutoff for equivocal result, constant 1 added to remove negative values.



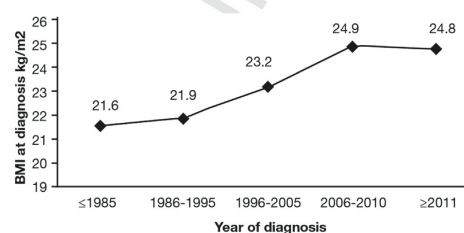


**Figure 2.** Trend analysis of age at diagnosis with CD in adulthood over 5 time periods (n = 652).

diagnostics.<sup>26</sup> This may partly explain the large percentage of our patients diagnosed in adulthood and the increased median age of diagnosis and changed clinical presentation over time. However, the fact that Ireland has traditionally been an area of high awareness and diagnosis raises the intriguing possibility of a biologic shift in CD, perhaps linked to variations in environmental exposures, contributing to an increased cadre of patients with milder disease and older onset.

CD was more prevalent in female patients confirming previous observations.<sup>7,15</sup> Conversely, population-based screening studies suggest a more equal CD gender distribution.<sup>3</sup> However, despite improvements in diagnosis, the female/male ratio has remained stable over time in our cohort. Male patients were diagnosed later in life than female patients, which has been previously reported in Ireland.<sup>17</sup> Potential explanations for this age difference at diagnosis are, women being more likely to require health care during their fertile years, and/or gender-dependent differences in clinical presentation, including the presence of coexistent IMCs.<sup>3,6</sup> Indeed, a greater proportion of women presented with certain intestinal and extraintestinal manifestations, and more coexistent IMCs, supporting the latter contention and suggesting that CD presentation and prognosis was more severe in women.

Our study confirms the association of CD with other IMCs, predominantly TD and T1DM, reported in other studies.<sup>11</sup> TD prevalence has decreased significantly over time in patients diagnosed in adulthood. Some studies



**Figure 3.** Trend analysis of BMI at diagnosis with CD in adulthood over 5 time periods (n = 391).

have suggested tTGAs in patients with CD may promote the development of TD.<sup>27</sup> It could be hypothesized that with better diagnostic tools and screening of at-risk patients, CD has been diagnosed at an earlier stage, thus avoiding long-term exposure to tTGAs and decreasing the prevalence of TD. Meanwhile, T1DM showed a nonsignificant increase in the past decade. Kivela et al<sup>8</sup> recently reported a similar finding in children, suggesting that the increase in T1DM is caused by intensified screening for CD in patients with T1DM, and the decrease in TD could be caused by the disappearance of iodine deficiency in patients with CD. The explanation for the increase in T1DM is also plausible in our cohort; however, iodine deficiency is less likely to explain the reduction in TD.<sup>28</sup> It is also possible that morphing environmental factors are predisposing individuals to develop CD who have less exposure to TD genetic susceptibility factors than heretofore, analogous to our scenario of a biologic shift favoring milder, later-onset disease.

An interesting finding in our study is how, over time, more patients have presented at diagnosis with a normal, overweight, or obese BMI in line with previous reports.<sup>29</sup> A recent large epidemiologic study found the risk of developing dermatitis herpetiformis increased by 14% per BMI unit in obese women; however, the risk of CD decreased 7% per BMI unit, and increased in underweight women. This latter finding may be caused by underdiagnosis of the condition in obese patients, and may explain the observed tendency in our results to a higher age at diagnosis in overweight and obese patients.<sup>30</sup> A limitation to our study is the lack of an unaffected population control sample; thus, the observed changes in BMI of patients with CD over time may be simply reflecting worldwide general population trends.<sup>31</sup> However, discerning the basis of the relationship between BMI and autoimmune conditions, such as CD, is highly desirable.

Our results on the changes in CD clinical presentation in adults over time correlate with those of Kivela et al<sup>8</sup> on the pediatric population. Kivela et al<sup>8</sup> reports that this change toward a milder presentation both clinically and histologically has reached a plateau in Finland because of changes in the diagnostics of the disease, but also, potentially, to modification of environmental factors. Improved diagnostic techniques alone cannot explain the well-reported increase in incidence observed in CD.<sup>3</sup> Moreover, disease incidence can vary considerably between geographically close areas or even between socioeconomic groupings within the same country.<sup>8,10</sup> This points to a true rise in the incidence of CD caused by environmental factors rather than solely improved detection.<sup>3</sup>

The amount of gluten consumed by infants seems to play a role in the process of tTGA seroconversion.<sup>9</sup> A recent study in Indian adults concluded that regional variations in CD prevalence were correlated to wheat intake.<sup>10</sup> Gluten consumption has tripled in the United

States since 1977, which correlates with the time of increased CD incidence. Thus, increased gluten consumption worldwide may possibly underlie a true rise in CD incidence in recent years.<sup>32</sup> In Ireland there are no recent studies on the prevalence and incidence of CD. These studies are desirable to correlate with observations in other countries, especially given the change in CD presentation reported in the present study.

### Strengths and Limitations

This study is the largest on CD clinical presentation in the Republic of Ireland and is also the only study carried out in the country on trends of CD over a lengthy period of time. To the extent of our knowledge, few studies internationally have examined such an extended time frame in the adult CD population. The observed significant changes in clinical presentation in a country with historically high detection rates expands current knowledge about the evolution in CD presentation.

Possible limitations of this study are its retrospective nature and sample convenience, which could lead to bias in data acquisition; however, the participating centers are secondary referral centers, therefore our sample should be broadly representative of patients geographically and in terms of disease severity, mitigating selection bias. Although all efforts were made to recover clinical, histologic, and serologic data for all patients, this information was not available in several cases. The latter limitation has been minimized by the large number of patients included.

The ratios we derived to compare tTGA values, to allow comparison of different time periods and participating centers, may not be robust given the different kinetics of the various assays. Nevertheless, they correlate well with the degree of small bowel damage at diagnosis and are included with this caveat.

### Conclusions

This study reports the clinical features of a large cohort of patients with CD in Ireland both cross-sectionally and over a long time frame. Our results corroborate findings internationally, but also shed light on the changing characteristics of CD in a country where there is a paucity of data, but where high incidence has been reported. This information is needed to increase awareness of CD among physicians and the general population, which may facilitate early diagnosis, improved management, and prevention of associated comorbidities.

The clinical profile of the disease has changed radically over the past 55 years. It seems plausible that these changes are caused by improved detection of nonclassical disease, and potentially shifts in environmental exposures, which are engendering changes in the clinical manifestation of CD.

### Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at [www.cghjournal.org](http://www.cghjournal.org), and at <http://dx.doi.org/10.1016/j.cgh.2016.11.018>

### References

- Ludvigsson JF, Leffler DA, Bai JC, et al. The Oslo definitions for celiac disease and related terms. *Gut* 2013;62:43–52.
- Reilly NR, Green PH. Epidemiology and clinical presentations of celiac disease. *Semin Immunopathol* 2012;34:473–478.
- Ludvigsson JF, Rubio-Tapia A, van Dyke CT, et al. Increasing incidence of celiac disease in a North American population. *Am J Gastroenterol* 2013;108:818–824.
- Rubio-Tapia A, Hill ID, Kelly CP, et al. ACG clinical guidelines: diagnosis and management of celiac disease. *Am J Gastroenterol* 2013;108:656–676, quiz 77.
- Lohi S, Mustalahti K, Kaukinen K, et al. Increasing prevalence of celiac disease over time. *Aliment Pharmacol Ther* 2007;26:1217–1225.
- Vivas S, Ruiz de Morales JM, Fernandez M, et al. Age-related clinical, serological, and histopathological features of celiac disease. *Am J Gastroenterol* 2008;103:2360–2365, quiz 6.
- Volta U, Caio G, Stanghellini V, et al. The changing clinical profile of celiac disease: a 15-year experience (1998–2012) in an Italian referral center. *BMC Gastroenterol* 2014;14:194.
- Kivela L, Kaukinen K, Lahdeaho ML, et al. Presentation of celiac disease in Finnish children is no longer changing: a 50-year perspective. *J Pediatr* 2015;167:1109–1115.
- Szajewska H, Shamir R, Mearin ML, et al. Gluten introduction and the risk of celiac disease. A position paper by the European Society For Paediatric Gastroenterology, Hepatology & Nutrition. *J Pediatr Gastroenterol Nutr* 2016;62:507–513.
- Ramakrishna BS, Makharia GK, Chetri K, et al. Prevalence of adult celiac disease in India: regional variations and associations. *Am J Gastroenterol* 2016;111:115–123.
- Lundin KE, Wijmenga C. Celiac disease and autoimmune disease-genetic overlap and screening. *Nat Rev Gastroenterol Hepatol* 2015;12:507–515.
- Mones RL. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *J Pediatr Gastroenterol Nutr* 2009;48:645–646.
- Snyder MR, Murray JA. Celiac disease: advances in diagnosis. *Expert Rev Clin Immunol* 2016;12:449–463.
- Ditah IC, Marietta E, Nadeau A, et al. Despite an increased use of a gluten free diet, most celiac disease remains undiagnosed in the USA. *Gastroenterology* 2014;146(Suppl 1):S-471.
- Kocsis D, Miheller P, Lorinczy K, et al. Celiac disease in a 15-year period of observation (1997 and 2011) in a Hungarian referral centre. *Eur J Intern Med* 2013;24:461–467.
- Mylotte M, Egan-Mitchell B, McCarthy CF, et al. Incidence of celiac disease in the West of Ireland. *Br Med J* 1973;1:703–705.
- Tajuddin T, Razif S, Dhar R, et al. Clinical presentation of adult celiac disease. *Ir Med J* 2011;104:20–22.
- Saleem A, Connor HJ, Regan PO. Adult celiac disease in Ireland: a case series. *Ir J Med Sci* 2012;181:225–229.
- Ludvigsson JF, Bai JC, Biagi F, et al. Diagnosis and management of adult celiac disease: guidelines from the British Society of Gastroenterology. *Gut* 2014;63:1210–1228.

- 813 20. Dickson BC, Streutker CJ, Chetty R. Celiac disease: an update  
814 for pathologists. *J Clin Pathol* 2006;59:1008–1016.
- 815 21. Ierardi E, Losurdo G, Piscitelli D, et al. Seronegative celiac dis-  
816 ease: where is the specific setting? *Gastroenterol Hepatol Bed*  
817 *Bench* 2015;8:110–106.
- 818 22. Leffler DA, Green PH, Fasano A. Extraintestinal manifestations of  
819 celiac disease. *Nat Rev Gastroenterol Hepatol* 2015;12:561–571.
- 820 23. Gutierrez-Achury J, de Almeida RC, Wijmenga C. Shared ge-  
821 netics in celiac disease and other immune-mediated diseases.  
822 *J Intern Med* 2011;269:591–603.
- 823 24. Singh P, Kurray L, Agnihotri A, et al. Titers of anti-tissue trans-  
824 glutaminase antibody correlate well with severity of villous ab-  
825 normalities in celiac disease. *J Clin Gastroenterol* 2014.
- 826 25. John JA, Draper NR. An alternative family of transformations. *J R*  
827 *Stat Soc Ser C Appl Stat* 1980;29:190–197.
- 828 26. Smukalla S, Lebowitz B, Mears JG, et al. How often do hemato-  
829 logists consider celiac disease in iron-deficiency anemia? Results  
830 of a national survey. *Clin Adv Hematol Oncol* 2014;12:100–105.
- 831 27. Duntas LH. Does celiac disease trigger autoimmune thyroiditis?  
832 *Nat Rev Endocrinol* 2009;5:190–191.
- 833 28. Smyth P, O'Herlihy C. Dietary iodine intake in pregnancy: an  
834 update. *Ir Med J* 2012;105:5–6.
- 835 29. Dickey W, Kearney N. Overweight in celiac disease: prevalence,  
836 clinical characteristics, and effect of a gluten-free diet. *Am J*  
837 *Gastroenterol* 2006;101:2356–2359.
- 838 30. Harpoe MC, Basit S, Andersson M, et al. Body mass index and  
839 risk of autoimmune diseases: a study within the Danish National  
840 Birth Cohort. *Int J Epidemiol* 2014;43:843–855.
- 841
- 842
- 843
- 844
- 845
- 846
- 847
- 848
- 849
- 850
- 851
- 852
- 853
- 854
- 855
- 856
- 857
- 858
- 859
- 860
- 861
- 862
- 863
- 864
- 865
- 866
- 867
- 868
- 869
- 870
- 871 31. Finucane MM, Stevens GA, Cowan MJ, et al. National, regional,  
872 and global trends in body-mass index since 1980: systematic  
873 analysis of health examination surveys and epidemiological  
874 studies with 960 country-years and 9.1 million participants.  
875 *Lancet* 2011;377:557–567.
- 876 32. Kasarda DD. Can an increase in celiac disease be attributed to  
877 an increase in the gluten content of wheat as a consequence of  
878 wheat breeding? *J Agric Food Chem* 2013;61:1155–1159.
- 879
- 880
- 881
- 882
- 883
- 884
- 885
- 886
- 887
- 888
- 889
- 890
- 891
- 892
- 893
- 894
- 895
- 896
- 897
- 898
- 899
- 900
- 901
- 902
- 903
- 904
- 905
- 906
- 907
- 908
- 909
- 910
- 911
- 912
- 913
- 914
- 915
- 916
- 917
- 918
- 919
- 920
- 921
- 922
- 923
- 924
- 925
- 926
- 927
- 928

**Reprint requests**

Address requests for reprints to: Patricia Dominguez Castro, Trinity Trans-  
lational Medicine Institute & Department of Clinical Medicine, Trinity Centre for  
Health Science, St James's Hospital, Dublin 8, Republic of Ireland. e-mail:  
domingup@tcd.ie.

**Acknowledgments**

The authors thank Dr Jeanne Dunne and Professor Con Feighery from the  
immunology department at Saint James Hospital for their input on interpreting  
serologic data. They also thank Dr Lina Zgaga from the Department of Public  
Health and Primary Care in Trinity College Dublin for her comments and advice.  
They thank all patients for providing the information necessary for this study,  
and all the administration staff in the participating hospitals for their help  
retrieving patient data.

**Conflicts of interest**

The authors disclose no conflicts.

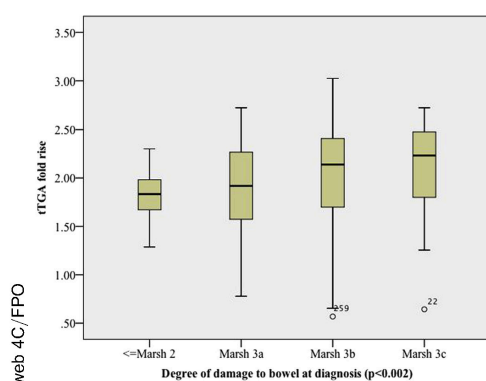
**Funding**

This research was funded by Trinity College Dublin with a Postgraduate Usher  
Fellowship. No other specific grant from any funding agency in the public,  
commercial, or not-for-profit sectors was received. R. McManus is funded by a  
Science Foundation Ireland principal investigator award (09/N.1/B2640).



## Supplementary References

1. Rubio-Tapia A, Ludvigsson JF, Brantner TL, et al. The prevalence of celiac disease in the United States. *Am J Gastroenterol* 2012;107(10):1538–1544, quiz 7, 45.
2. Rubio-Tapia A, Murray JA. Celiac disease. *Curr Opin Gastroenterol* 2010;26(2):116–122.
3. Rampertab SD, Pooran N, Brar P, et al. Trends in the presentation of celiac disease. *Am J Med* 2006;119(4):355.e9–14.
4. Aronsson CA, Lee HS, Liu E, et al. Age at gluten introduction and risk of celiac disease. *Pediatrics* 2015;135(2):239–245.
5. Sanz Y. Microbiome and Gluten. *Ann Nutr Metab* 2015;67(Suppl 2):28–41.
6. Decker E, Engelmänn G, Findeisen A, et al. Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children. *Pediatrics* 2010;125(6):e1433–e1440.
7. Stene LC, Honeyman MC, Hoffenberg EJ, et al. Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study. *Am J Gastroenterol* 2006;101(10):2333–2340.
8. Marild K, Ludvigsson J, Sanz Y, et al. Antibiotic exposure in pregnancy and risk of celiac disease in offspring: a cohort study. *BMC Gastroenterol* 2014;14:75.
9. Fasano A. Systemic autoimmune disorders in celiac disease. *Current opinion in gastroenterology* 2006;22(6):674–679.
10. Gutierrez-Achury J, de Almeida RC, Wijmenga C. Shared genetics in celiac disease and other immune-mediated diseases. *J Intern Med* 2011;269(6):591–603.
11. Cosnes J, Cellier C, Viola S, et al. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *Clin Gastroenterol Hepatol* 2008;6(7):753–758.
12. Norstrom F, Lindholm L, Sandstrom O, et al. Delay to celiac disease diagnosis and its implications for health-related quality of life. *BMC Gastroenterol* 2011;11:118.
13. Gray AM, Papanicolaou IN. Impact of symptoms on quality of life before and after diagnosis of celiac disease: results from a UK population survey. *BMC Health Serv Res* 2010;10:105.
14. Cranney A, Zarkadas M, Graham ID, et al. The Canadian Celiac Health Survey. *Dig Dis Sci* 2007;52(4):1087–1095.
15. Yang A, Chen Y, Scherl E, et al. Inflammatory bowel disease in patients with celiac disease. *Inflammatory bowel diseases* 2005;11(6):528–532.
16. Oberhuber G, Granditsch G, Vogelsang H. The histopathology of celiac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol* 1999;11(10):1185–1194.
17. Rahmati A, Shakeri R, Sohrabi M, et al. Correlation of tissue transglutaminase antibody with duodenal histologic marsh grading. *Middle East journal of digestive diseases* 2014;6(3):131–136.
18. Donaldson MR, Firth SD, Wimpsee H, et al. Correlation of duodenal histology with tissue transglutaminase and endomysial antibody levels in pediatric celiac disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2007;5(5):567–573.
19. Schosler L, Christensen LA, Hvas CL. Symptoms and findings in adult-onset celiac disease in a historical Danish patient cohort. *Scand J Gastroenterol* 2016;51(3):288–294.
20. Vivas S, Ruiz de Morales JM, Fernandez M, et al. Age-related clinical, serological, and histopathological features of celiac disease. *Am J Gastroenterol* 2008;103(9):2360–2365, quiz 6.
21. Naiyer AJ, Shah J, Hernandez L, et al. Tissue transglutaminase antibodies in individuals with celiac disease bind to thyroid follicles and extracellular matrix and may contribute to thyroid dysfunction. *Thyroid* 2008;18(11):1171–1178.
22. Smyth PP, Hetherington AM, O'Carroll D, et al. Iodine intake in Ireland—an update. *Ir J Med Sci* 1988;157(12):377–379.
23. Miranda DM, Massom JN, Catarino RM, et al. Impact of nutritional iodine optimization on rates of thyroid hypoechogenicity and autoimmune thyroiditis: a cross-sectional, comparative study. *Thyroid* 2015;25(1):118–124.
24. Nawoor Z, Burns R, Smith DF, et al. Iodine intake in pregnancy in Ireland—a cause for concern? *Ir J Med Sci* 2006;175(2):21–24.
25. Duntas LH. The Role of Iodine and Selenium in Autoimmune Thyroiditis. *Horm Metab Res* 2015;47(10):721–726.
26. Andren Aronsson C, Lee HS, et al. Effects of Gluten Intake on Risk of Celiac Disease: a case-control study on a Swedish birth cohort. *Clin Gastroenterol Hepatol* 2016;14(3):403–409.



**Supplementary Figure 1.** Median tissue transglutaminase antibodies fold rise in patients with celiac disease having different levels of small bowel damage classified according to Marsh criteria.

**Supplementary Table 1.** The Oslo classification of celiac disease

| Oslo Classification of celiac disease |                                                                                                                                                                              |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Celiac disease presentation           | Characteristics                                                                                                                                                              |
| Classical celiac disease              | Malabsorption present; diarrhoea, steatorrhoea, weight loss or growth failure required at presentation                                                                       |
| Non-classical celiac disease          | No signs and symptoms of malabsorption at presentation (i.e. patients with constipation and abdominal pain). Also monosymptomatic disease other than diarrhoea, steatorrhoea |
| Subclinical celiac disease            | Celiac disease that is below the threshold of clinical detection                                                                                                             |

**Supplementary Table 2.** Prevalence of celiac disease extra intestinal symptoms and co-existent immune-mediated conditions

| Characteristic                                                                    | Celiac Patients                  | %    |
|-----------------------------------------------------------------------------------|----------------------------------|------|
| Prevalence of celiac disease extra-intestinal manifestations <sup>a</sup> , n 749 | Yes                              | 66.1 |
| Celiac disease associated disorders, n 749                                        | Bone Disorders <sup>b</sup>      | 40.6 |
|                                                                                   | Iron Deficiency                  | 35.4 |
|                                                                                   | Vitamin Deficiency <sup>c</sup>  | 16.8 |
|                                                                                   | Depression                       | 9.1  |
|                                                                                   | Dermatitis herpetiformis         | 7.2  |
|                                                                                   | Reflux Oesophagitis              | 4.1  |
|                                                                                   | Ataxia                           | 1.1  |
|                                                                                   | Lymphoma                         | 1.1  |
|                                                                                   | Neuropathy                       | 0.5  |
| Presence of co-existent immune-mediated conditions, n 749                         | Yes                              | 31.2 |
| Co-existent immune-mediated conditions, n 749                                     | Thyroid disease                  | 19.9 |
|                                                                                   | Type 1 diabetes mellitus         | 3.6  |
|                                                                                   | Psoriasis                        | 2.7  |
|                                                                                   | Inflammatory bowel disease       | 2.1  |
|                                                                                   | Rheumatoid arthritis             | 1.6  |
|                                                                                   | Pernicious Anaemia               | 1.1  |
|                                                                                   | Systemic lupus erythematosus     | 0.4  |
|                                                                                   | Other                            | 5.5  |
| Other co-existent immune-mediated conditions, n 749                               | Alopecia areata                  | 1.3  |
|                                                                                   | Sarcoidosis                      | 0.9  |
|                                                                                   | Immunoglobulin A deficiency      | 0.5  |
|                                                                                   | Multiple sclerosis               | 0.4  |
|                                                                                   | Scleroderma                      | 0.4  |
|                                                                                   | Lichen planus                    | 0.4  |
|                                                                                   | Autoimmune hepatitis             | 0.4  |
|                                                                                   | Autoimmune adrenalitis           | 0.3  |
|                                                                                   | Primary biliary cholangitis      | 0.3  |
|                                                                                   | Myasthenia gravis                | 0.1  |
|                                                                                   | Immune-mediated thrombocytopenia | 0.1  |
|                                                                                   | Vitiligo                         | 0.1  |
|                                                                                   | Autoimmune neutropenia           | 0.1  |

<sup>a</sup>Manifestations occurring either at celiac disease diagnosis, or prior or posterior to celiac disease diagnosis.<sup>b</sup>Osteoporosis and/or Osteopenia.<sup>c</sup>B12 or Folic acid deficiency.