

Clinical and genetical aspects of celiac disease

Audur Heida Gudjónsdóttir

Department of Pediatrics
Institute of Clinical Sciences

The Sahlgrenska Academy
at
University of Gothenburg
Sweden



UNIVERSITY OF GOTHENBURG
Göteborg 2008

Cover picture by Magnús Plank 2008

Heredity

I am the family face;
Flesh perishes, I live on,
Projecting trait and trace
Through time to times anon,
And leaping from place to place
Over oblivion.

The years-heired feature that can
In curve and voice and eye
Despise the human span
Of durance — that is I;
The eternal thing in man,
That heeds no call to die.

Thomas Hardy

*(First published in Moments of Vision
and Miscellaneous Verses, Macmillan, 1917)*

To all the celiac children

Abstract

Celiac disease (CD), or gluten-sensitive enteropathy, is one of the most common chronic diseases in childhood but is diagnosed in all ages. CD is a genetically driven immunological intolerance to dietary gluten. The treatment is a gluten-free diet. The diagnostic criteria are the ESPGHAN criteria, which include the histological characteristics of villous atrophy, crypt hyperplasia and increased number of intraepithelial lymphocytes (IEL). The clinical manifestations in CD range from severely affected young children to children and adults with milder symptoms as well as patients with silent CD. There is a strong heredity in CD with the well-known HLA components DQ2 and DQ8. The genetics in CD are believed to confer up to 40% HLA genetics and otherwise non-HLA genetics. The knowledge of the genotype-phenotype association in CD is limited.

The aim of this study has been to estimate the risk of a third sibling being affected in CD sib-pair families, identify the chromosomal region containing susceptibility genes in CD and study the genotype-phenotype association in CD.

Material was collected from 107 families with at least two affected siblings, making a total of 224 CD siblings, as well as their healthy siblings and parents. Screening for CD was performed in these apparently healthy members and the estimated risk for CD in the third sibling and parent was then calculated. Thirteen new CD cases were diagnosed, six siblings and seven parents. The estimated sibling risk was 26.3% and the parent risk was 12.9%. The risk of a sibling of two affected siblings having CD was approximately three times higher compared to siblings of one affected sibling. Considering the high level of knowledge of CD in these families, the number of undiagnosed cases was surprisingly high. We suggested that serological screening should be offered all first-degree relatives of CD patients.

Genome-wide linkage scan was performed in the same material. This work showed significant evidence of linkage to CD with an interesting region on chromosome 5q31-33 and on chromosome 11q. Simplex CD family material was collected for further genetic association studies.

The phenotype-genotype association was examined in two studies. An investigation was made of a possible interaction between the phenotypes and HLA class II risk alleles, the CTLA4 +49 A/G polymorphism, the haplotype MH30*G:-1147*T:+49*A:CT60*G:CT61*A and the 5q31-33 locus, in CD. The patients were grouped according to symptoms at presentation, the age at diagnosis and gender. The heritability of the phenotype was estimated to be 0.45. The AA genotype at the CTLA4 +49A/G polymorphism was associated with clinically silent disease. No other correlations were found between genotypes and clinical presentation, age at diagnosis or gender.

A genotype-phenotype analysis was made of phenotypes in DQ2-negative CD patients in the largest DQ2-negative CD group that has been published compared to DQ2-positive CD controls in a European population. The finding was that the clinical presentation differed significantly between DQ2-negative and DQ2-positive CD patients in Italy and Sweden. In both samples there was an association between DQ2-negative cases and classic symptoms. In the Italian sample there was also an association between silent grade and DQ2-negative cases. Autoimmune disease was significantly overrepresented in DQ8-positive patients.

This thesis shows that the risk for third sibling and parents is, as expected, increased in sib-pair families, as the expected risk of being affected in polygenic diseases is higher in families with multiple cases compared to single-case families. The genome scan indicated significant linkage to 11q and 5q, which makes these regions interesting for further fine mapping of these regions using association analysis. Genotype-phenotype analysis of both HLA and non-HLA locus showed some significant correlation between silent CD and both CTLA4 +49 AA genotype and the DQ2-negatives. In addition, an association was shown between classic symptom grade and DQ2-negative cases.

Key words: celiac disease, sib-pair, screening, genome-wide scan, linkage analysis, heritability, genotypes, DQ2-negative, phenotypes, autoimmune disease

ISBN 978-91-628-7484-1

Göteborg 2008

Contents

Abstract	5
List of publications	9
Abbreviations	11
Glossary	11
Introduction	13
Pathogenesis	15
Genetics in celiac disease	17
Clinical aspects	19
Molecular genetics	25
Complex traits	27
Genetic analysis	27
Genotype-phenotype association	28
Aim of the study	31
Patients and Methods	33
Results	39
Conclusion and discussion	41
Future genetic and genotype-phenotype studies in CD	45
Sammanfattning på svenska	47
Acknowledgements	49
References	51
Paper I-IV	

List of publications

- I. Gudjónsdóttir AH, Nilsson S, Ek J, Kristiansson B, Ascher H.
The risk of celiac disease in 107 families with at least two affected siblings.
J Pediatr Gastroenterol Nutr 2004;38:338-42.
- II. Naluai AT, Nilsson S, Gudjónsdóttir AH, Louka AS, Ascher H, Ek J, Hallberg B, Samuelsson L, Kristiansson B, Martinsson T, Nerman O, Sollid LM, Wahlström J.
Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for CD on chromosomes 5 and 11.
Eur J Hum Genet 2001;9:938-44.
- III. Gudjónsdóttir AH, Nilsson S, Naluai ÁT, Ek J, Amundsen SS, Wahlström J, Ascher H.
Association between genotypes and phenotypes in CD.
In press J Pediatr Gastroenterol Nutr 2008
- IV. Gudjónsdóttir AH, Nilsson S, Hugot J-P, Mustalahti K, Clot F, Coto I, Percopo S, Ascher A.
Clinical features of DQ2-negative compared to DQ2-positive celiac disease.
In manuscript

Abbreviations

AGA	Anti gliadin-antibodies
APC	Antigen presenting cell
ARA	IgA-class Reticulin antibody
CD	Celiac disease
cM	centiMorgan
CTLA4	Cytotoxic T lymphocyte-associated antigen-4
DH	Dermatitis herpetiformis
DNA	Deoxyribonucleic acid
ELISA	Enzyme-Linked ImmunoSorbent Assay
EMA	Anti-Endomysium antibody
ESPGHAN	European Society of Paediatric Gastroenterology, Hepatology and Nutrition
GWA	Genome-wide association study
GWL	Genome-wide linkage study
GFD	Gluten Free Diet
HLA	Human leukocyte antigens
IBD	Identity By Descent
IEL	Intraepithelial T lymphocytes
LD	Linkage disequilibrium
MYO9B	Myosin IXB
NPL	Non parametric linkage
PCR	Polymerase chain reaction
RCD	Refractory sprue or celiac disease
RR	Relative risk
SNP	Single nucleotide polymorphism
TG2	Tissue Transglutaminase type 2
T1D	Type 1 Diabetes mellitus
TDT	Transmission disequilibrium test

Glossary

Adaptive immune response: An immune response mediated by B and T cells after exposure to a specific antigen. Involves memory, self/non-self recognition and specificity.

Allele: Alternative form of a genetic locus; a single allele for each locus is inherited from each parent.

Association: A tendency of two characters (disease, marker, alleles, gene) to occur together at non-random frequencies.

Candidate gene: A gene from the appropriate chromosomal location that is suspected of being the disease gene.

CentiMorgan: The unit of genetic distance. One cM is on average equal to about one mega base-pairs (1cM = 1Mb = 1,000,000 base-pairs).

Epigenetic: Heritable changes in gene expression that do not change the DNA sequence but rather provide an “extra” layer of transcriptional control that regulates how genes are expressed.

Epistasis: Genetic interaction, genes acting on the same or related biological pathway. The gene whose phenotype is expressed is said to be epistatic.

Gene: a unit of heredity, which is equal to a region of DNA.

Genotype: Genotype describes the genetic constitution of an individual, which is the specific allelic combination of the two homologous chromosomes.

Gluten: prolamins or storage proteins in wheat.

Haplotype: A block of alleles that transmit together.

Heritability: The variance in the phenotype caused by genetic factors.

Heterozygote: The presence of different alleles at one or more loci on homologous chromosomes.

Homozygote: The presence of two copies of the same alleles at one or more loci on homologous chromosomes.

Innate immune response: A non-specific immune response to antigens, including anatomic and physiologic barriers, endocytic and phagocytic activity, and inflammatory secretions.

Linkage: A relationship between loci. Two loci are linked if they are located on the same chromosome.

Linkage disequilibrium: A non-random pattern of association between alleles at different loci within a population. Obtained when a particular marker allele is located so close to the disease susceptibility allele that, over generations, the two will be inherited together.

Locus: is a specific position in the genome or on a chromosome.

Microsatellite: Polymorphic loci present in nuclear DNA that consist of repeated units of 1-4 base pairs in length, used as genetic markers.

PCR: A laboratory method used to amplify specific regions of a DNA strand.

Phenotype: The detectable outward manifestations of a specific genotype. Genotype and phenotype are not always directly correlated. Some genes only express a given phenotype in certain environmental conditions. Conversely, some phenotypes could be the result of multiple genotypes.

Polygenic: Polygenic inheritance is when many genes together influence the phenotype.

Sib-pair families: Two or more affected siblings.

Transmission disequilibrium test: If one allele increases risk of disease or trait, this allele will be transmitted to the affected offspring more often than expected by chance alone.

Introduction

CD or gluten-sensitive enteropathy is caused by dietary gluten ingestion in genetically susceptible individuals. Tissue transglutaminase type 2 (TG2) specific auto-antibodies are characteristic of CD, and increased TG2 activity has been observed in the small intestinal biopsies of patients. The immunological reaction that gluten induces causes a chronic immune reaction in the small intestine. CD is probably the best understood HLA disorder, as the environmental factor gluten is known. The non-HLA genetic and other environmental factors are still not fully understood in CD. There is some disagreement as to whether CD is an autoimmune disease or concurrent features typical of allergy and autoimmunity. The definition of autoimmune disease is that the tissue damage must be caused by an adaptive immune response to self-antigens [1-3]. Most authors today consider CD to be an autoimmune disease since TG2 antibody was identified in 1997 [4] as the main autoantigen for the anti-EMA. The approach of investigating genes in diseases is shown in a flow chart in Figure 1.

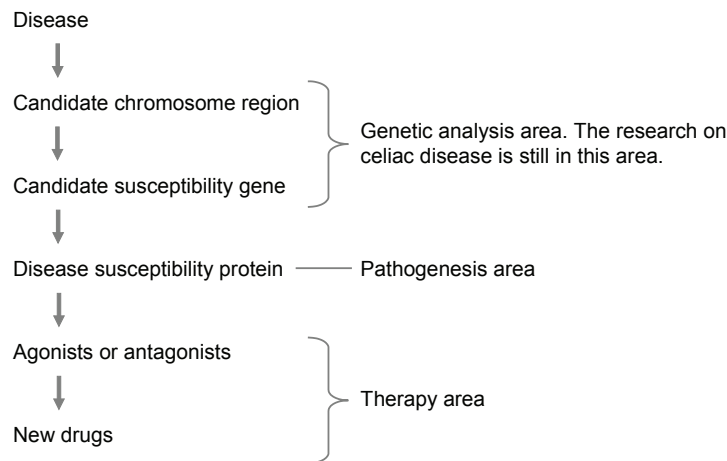


Figure 1. Flow chart showing the approach of investigating genes in diseases.

History – milestones in celiac disease and genetic

Disease history

The modern history of CD is short. Dr Samuel Gee (1839-1911) was an English paediatrician who in 1888 published in his thesis [5] the first complete modern description of the clinical picture of CD, and suggested that diet was important in controlling the disease. He stated: "If the patient can be cured at all, it must be by means of diet." Christian Herter, an American physician, wrote a book in 1908 in which he called children with CD "intestinal infantilism" [6]. He noticed that their

growth was retarded and that fat was better tolerated than carbohydrate. In 1924, Sydney V. Haas, an American paediatrician, reported positive effects of a diet of bananas [7].

The next breakthrough was in 1950 when the Dutch paediatrician Dr Willem Dicke in his thesis showed that children with CD improved when eating a diet without wheat, rye, corn and oats [8]. Together with CM Anderson he later identified gluten as the harmful protein in CD [9, 10]. Studies in 1995 suggested the non-toxic effect of oats [11]. The first codex standard for a gluten-free product came in 1978 and a revised standard in 2001.

Shiner described in 1956 a method of making small intestinal biopsies on adult patients [12].

Serological markers were first described by E Berger in 1958 [13] but were studied and used more after 1970. The AGA antibodies were first used and connective tissue antibodies were later found, such as ARA and EMA (1983). In 1997, W Dietrich discovered the tissue Transglutaminase type 2, the unknown autoantigen of endomycial antibodies [4].

MacDonald first described the hereditary character of CD in 1965 [14].

Genetic history

In 1953, Watson and Crick's paper described the structure of DNA [15]. Further work by Crick and co-workers showed that the genetic code was based on non-overlapping triplets of bases, called codons, allowing Khorana, Holley and Nirenberg to decipher the genetic code and its function in protein synthesis. These findings represent the birth of molecular biology. In 1972, Cohen and Boyer created recombinant DNA, which is biotechnology that became important in the molecular biology. In 1975, the first generation of DNA markers appeared: restriction enzyme length polymorphism (RFLPs). The first gene to be mapped by positional cloning and linkage analysis was Duchenne muscular dystrophy in 1982. In 1983, Kary Mullis invented the PCR technique, which became very important for future genetic research [16]. Microsatellites came in 1989 and made genetic mapping more effective. The sequence of the human genome was published in 2001 [17, 18]. The same year the first non-HLA gene in autoimmune disease was published, NOD2 in Crohn's disease [19].

Inheritance of celiac disease and risk

CD is one of the most common chronic diseases in Swedish children, surpassed only by asthma and allergies. In Sweden, childhood CD was increasingly diagnosed in the 1980s and early 1990s [20-22]. However, after 1996 the incidence suddenly decreased [23]. Other countries have seen increased prevalence and concluded that it is because of increase awareness and screening programmes [24].

Screening studies for CD in Sweden, Norway and other countries have shown prevalence figures varying between 2 and 6 per 1,000, in some studies even as high as 10 per 1,000 [25-31]. In an ongoing study, ETICS (<http://www.umu.se/phmed/epidemi/celiaki/etics/>) in Sweden, a screening study in 12-year-old children, has presented a prevalence as high as 30 per 1000 (95% CI: 25-33), Myléus et al presented

at an International Coeliac Disease Meeting in Maribor in 2007. In North America, CD has been diagnosed increasingly during the last 15 years and a prevalence of 0.75% is found [32].

In family studies (Table 1, Paper I), sibling prevalence more than ten times higher than the prevalence found in population studies has been reported [33, 34]. Reported parent prevalence has varied between 0 and 6.3% [34-41]. Studies of twins have shown a concordance rate in monozygotic twins of at least 86.5% compared to 23% in dizygotic twins [42].

The average prevalence of CD among children with T1D in 26 reports was 4.5% (0.97-16.4%) [43]. A high prevalence of CD is found in some syndromes: in Down's syndrome 3.6-8% in Europe [44-47] and as high as 10.3% in the USA [48, 49], in Turner 2.2-6.4% [50-52] and in Williams's syndrome 9.5% [53].

The CD iceberg

The iceberg model is often used to gain an epidemiological understanding of CD. The visible tip of the iceberg represents the diagnosed cases. The first part under the surface represents the undiagnosed or silent cases with gluten-induced enteropathy, CD patients found by screening. The deepest part represents the genetically predisposed individuals without gluten-induced enteropathy, called latent or potential CD, although the knowledge of when or if they will get CD is unknown.

Pathogenesis

The role of gluten

Gluten prolamins are storage proteins in grain (wheat, barley and rye). Gluten proteins are a mixture of water-insoluble glutenins and the alcohol-soluble gliadins. Human dietary gluten is poorly digested in the upper gastrointestinal tract because of the high proline content. They are resistant to degradation by gastric, pancreatic and intestinal brush-border membrane proteases in the intestine and remain undigested in the intestinal lumen. TG2 is an enzyme in the intestine, found both at the brush border and just below the epithelium. A stress reaction, as a hypothesis via reduced zinc in the intestinal wall and increased Ca⁺ activity [54], activates TG2 which deamidates gliadin peptides. This results in proline-rich peptides containing negatively charged glutamic acid residues, which increase their immunogenicity. Gluten peptides are rich in the amino acid sequence QXP, which is an excellent substance for TG2. Gluten has many known immunogenic peptides identified in α -gliadins, γ -gliadins and the LMW and HMW-glutenins [55]. In CD patients, immune responses to gliadin fractions promote an inflammatory reaction in the upper small intestine and this cause the inflammation and tissue damage (villous atrophy) seen in gluten enteropathy.

Immunopathogenesis

Two immunological pathways are active in CD, the innate immune and the adaptive immune pathway. The gluten peptides are involved in both of them.

The innate immune system, where gluten peptides (α -gliadin) induce IEL (CD8+

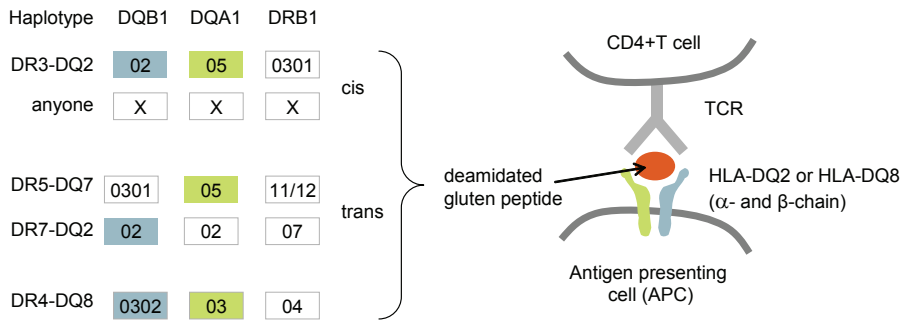


Figure 2. The HLA-DQ2 and DQ8 $\alpha\beta$ heterodimer are on the surface of antigen presenting cells. The binding site has a preference for negatively charged amino acids and thereby bind gluten peptides deamidated by TG2 with increased affinities and activate the CD4⁺ cells.

cis=alleles on the same chromosome

trans=alleles on different chromosomes

cells) mediate enterocytes destruction, by expressing the natural killer receptors D (NKG2D). Gluten can even induce NKG2D expression by stimulating the expression of the cytokine IL-15.

The adaptive immune system, where TG2-modified gliadin residues connect to HLA-DQ2 and DQ8 molecules on APCs and present it to a T cell receptor on CD4⁺ lymphocytes, which results in an immune response with the formation of antibodies against both anti-gliadin and TG2 [1], Figure 2.

The IL15 produced by the APCs and enterocytes is also capable of stimulating IELs and the T-cells of the adaptive immune system and is therefore one of the links between the two pathways [56].

The mechanism of how gluten peptide crosses the epithelial barrier is still unknown.

Environmental factors

Breastfeeding duration, and if ongoing when gluten-containing foods are introduced to infants, may protect against the developing of CD or cause a delay in the onset of symptoms [24, 57-60].

Infections have always been suspected of being a risk factor for developing CD, a prospective study has shown that a high frequency of rotavirus infection may increase the risk of CD in DQ2-positive and DQ8-positive children [61]. The mechanism is not clear but could be due to a combination of increased intestinal permeability and increased presence of TG2 in the mucosa, which may facilitate the deamidation excreta or stress reaction. A study showing increased risk of CD in children younger than two years of age at diagnosis, born in the summer compared to the winter, may support the hypothesis that infections are involved in the pathogenesis [62] as their gluten introduction takes place during the winter.

Genetics in celiac disease

The strongest evidence of genetic factors in CD is the increased risk seen in family-based studies and twin studies as shown above. CD is a complex genetic disease.

HLA

The human leukocyte antigen (HLA) complex is located on the short arm of chromosome 6 (6p21.31). This region is gene-dense, with more than 200 gene loci. Forty per cent of the expressed genes encoded in the region have presumed immune system functions. Three regions are recognized: class II, encoding among others, the DP, DQ and DR molecules, class III, encoding among others the TNF family, and class I. HLA class I molecules are found on the surface of most nucleated cells in the body, while class II are present only on the surface of B cells, T cells and macrophages [15]. There are seven HLA class II DQ variants (DQ2 and DQ4-9). Two of these variants, HLA-DQ2 and DQ8, are associated with CD.

The function of the HLA class II proteins on the APC surface is to bind peptide fragments of processed protein antigens and present them to T-cell receptors on CD4+ lymphocytes of the adaptive immune system, Figure 2. The peptide-binding groove on HLA class II molecules is composed of antiparallel β -sheets as a floor and antiparallel α -helices as walls. Small cavities in the groove bear polymorphic amino acids responsible for the peptide-binding specificity. The binding site has a preference for negatively charged side chains,

HLA association in CD

Genetically, CD is characterised by a strong HLA class II component (CELIAC 1). The HLA component consists of a combination of specific alleles at the DQ locus. The HLA haplotypes and serological typing is shown in Table 1.

Table 1. Haplotypes that predispose to celiac disease, the DQ2 and DQ8 heterodimers.

Haplotype I			Haplotype II			Serological typing notation
DRB1	DQA1	DQB1	DRB1	DQA1	DQB1	
03	05	02	-	-	-	DR3-DQ2
03	05	02	03	05	02	DR3-DQ2/DR3-DQ2
03	05	02	07	0201	02	DR3-DQ2/DR7-DQ2
11/12	05	0301	07	0201	02	DR5-DQ7/DR7-DQ2
04	03	0302	-	-	.	DR4-DQ8

This HLA DQ2 molecule is either encoded in *cis* in individuals who have the DR3-DQ2 haplotype, or in *trans* in individuals who are DR5-DQ7/DR7-DQ2 heterozygous. An increased risk of CD in individuals who are DR3-DQ2 homozygous and DR3-DQ2/DR7-DQ2 heterozygous has been demonstrated [63]. A small proportion of the CD patients have neither DQ2 nor DQ8 and these patients almost

exclusively encode for one half of the DQ2 heterodimer, *i.e.* either *DQA1*05* or *DQB1*02* [64]. In the European population more than 90% of the CD patients carry HLA-DQA1*05-DQB1*02 encoding the DQ2, compared to as many as 20-30% of the healthy population being carriers of DQ2, Figure 3 [15, 65, 66]. Almost all the patients that are DQ2-negative carry the DR4-DQ8.

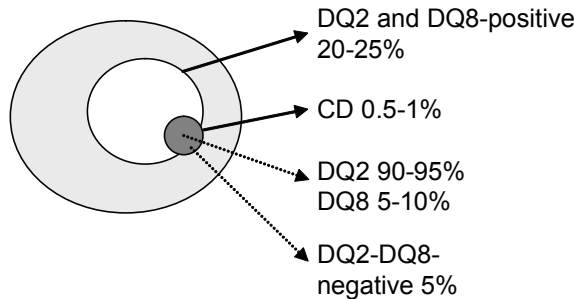


Figure 3. Incidence of HLA DQ2 and DQ8 in the European population and in the CD population.

The possibility of other risk loci in the HLA complex has been studied. The DQ-independent risk effects are relatively difficult to reveal due to the strong linkage disequilibrium in the whole region and this question still remains unsolved [67, 68].

A GWA study showed a strong association with the HLA-DQ2 [69].

Even though the HLA genes are considered necessary for CD some studies have estimated that the HLA DQ genes only contribute with 40% of the genetic risk in CD, indicating the importance of genes outside the HLA complex [70].

Non-HLA candidate genes and regions

Outside the HLA class II region, some other genes and regions of interest have been studied.

The 5q31-33 locus (CELIAC 2), the *CTLA4 +49* gene on 2q34 (CELIAC 3), the *MYO9B* on 19q13.1 locus (CELIAC 4) and the 4q27 (*IL2-IL21* genes) have been linked or associated with CD.

CTLA4 +49 is shown to be associated with CD in several studies and populations [71-74] but not in others [75]. In the GWA study there was a weak association with the CD28-*CTLA4-ICOS* region [69].

As *CTLA4 +49*, a co-stimulatory molecule of the T cells, plays an important role in maintaining immunological tolerance to self-antigens, it is an interesting candidate gene. Some studies have implicated a functional role of *CTLA4 +49*. The G allele of the *CTLA4 +49* A/G dimorphism has been shown to be associated with reduced control of T cell proliferation [76]. The role of *CTLA4 +49* in CD is not clear, but more likely minor.

The 5q31-33 region on chromosome 5 was the strongest region shown in the meta-analysis and pooled analysis of four genome scans [77]. Further association studies have, however, failed to find a strong susceptibility candidate [78]. There are 200

genes in this region, many of immunological importance. Some have been analysed, e.g. IL4, IL5, IL12, IL13 and IL14, without significant association in CD [79, 80]. On *19q13.1* the *MYO9B* gene was shown to be associated with CD in two independent cohorts in the Dutch population [81]. However, this finding has been difficult to replicate in other populations included in the Swedish/Norwegian study or in the GWA study [69, 82-84]. *MYO9B* might play a role by affecting the tight junction and epithelial barrier function of the gut and increase permeability of the gut, allowing the gluten peptides to enter the lamina propria more easily [85].

The strongest association in the recent published GWA study was shown to the *KIAA1109/Tenr/IL2/IL21* region on 4q27 [69]. This finding has been verified by our group in Swedish and Norwegian patients [86]. This region is of immunological interest since IL2 is a cytokine for T-cell activation and proliferation and IL21 is a cytokine that enhances B-, T- and NK-cell proliferation as well as interferon- γ production.

Clinical aspects

Diagnostic criteria

The first diagnostic criteria were discussed at the second Annual Meeting of ESPGHAN (ESPGA at that time) in Interlaken in 1969. The Interlaken or ESPGHAN 70 criteria were published [87] and contain three small intestinal biopsies, characteristic histological appearance of the mucosa, its normalization on a gluten-free diet (GFD) and a histological relapse within two years on reintroduction of gluten. A new round table discussion on the diagnostic criteria took place at the 22nd Annual Meeting of ESPGHAN in Budapest in 1989. The revision of the Interlaken criteria or ESPGHAN 90 criteria was published in 1990 [88]. The ESPGHAN 90 criteria abandoned the previous obligatory gluten challenge and the requirement of three biopsies, in cases with typical first small intestinal histology and a clear remission on GFD and serological normalization. Otherwise the investigation must be continued until the CD diagnosis is confirmed or can be excluded. However, a gluten challenge with a subsequent biopsy has a role in establishing the diagnosis in select clinical settings, such as in patients with a high suspicion of CD who have started with a GFD without biopsy confirmation of the disease, or of a negative serologic test result.

Serological markers

The first serological screening tests were described in the 1960s. Anti-gliadin antibodies (AGA) in serum were the most used antibodies in the earlier research and clinical practice and are still used in clinical practice in combination with the newer antibodies [89].

Later, anti-Endomysium antibody (IgA-EMA), directed against intermyofibril connective tissue of smooth muscle fibres, became available [90]. From the studies IgA-EMA demonstrated sensitivities of 97.4% (95% CI: 0.957-0.985) in adults and 96.1% (95% CI: 0.945-0.973) in children. The specificity of IgA-EMA was 99.6% (95% CI: 0.988-0.999) in adults. In studies of children, the specificity of IgA-EMA was 97.4% (95% CI: 0.963-0.982) [91].

TG2, the main auto-antigen for EMA, was identified in 1997 [4]. TG2 is a Ca^{2+} -dependent enzyme that is responsible for converting glutamine residuals to glutamic acid by deamidation when the amine is replaced by water [92]. An ELISA method using guinea pig IgA-TG2 was used earlier but now human IgA-TG2 is the method of choice. It has been shown to have a sensitivity of 96-100% and a specificity of 96-100% [93, 94] although in a systemic review the sensitivity of human TG2 was similar: in adults 98.1% (95% CI: 0.901-0.997) and in children 95.7% (95% CI: 0.903-0.981) with a specificity of 98.1% (95% CI: 0.958-0.991) in adults and 99% (95% CI: 0.946-0.996) in children [91].

Estimates of the sensitivity of the IgG class antibodies of EMA and TG2 suggest that these tests have poor sensitivities, around 40% (95% CI: 0.363-0.543), although the specificities were high at around 98.8% (95% CI: 0.935-0.998) [91].

Recently, a new ELISA test of antibodies against deamidated gliadin peptides (DGP) has been developed [95]. A few studies have been published, indicating a sensitivity and specificity for IgA-DGP of 83.6-91% and 90.3%-95%, and for IgG-DGP of 84.4-92% and 98.5%-99% respectively. The sensitivity for a combination of IgA-DGP/IgA-TG2 was 100% [96, 97].

However, all tests have overlapping confidential intervals [89, 91] that need to be taken into account when using them. When screening studies of risk groups, priority must be given to sensitivity although when screening unselected population tests with a high specificity should be given priority.

Histopathology

The histological criteria for celiac diagnosis includes villous atrophy, crypt hyperplasia, an increased number of intraepithelial T lymphocytes (IEL) [98], or >20-30 lymphocytes/100 enterocytes, and inflammation of the lamina propria. Different classifications have been used to make the examination as objective as possible. The Alexander classification has four grades based on stereomicroscopic appearance and histological characteristics [99], where grades 3 and 4 fulfil CD criteria. Table 2 shows the modified version. The Marsh classification has types 0-III, where type III fulfils the CD criteria. The subgroups are a, b and c as seen in Table 3 [100, 101].

Table 2. Modified Alexander classification for the histology of the small intestinal biopsies.

Grade	Histological characteristics
1	Normal
2	Normal villous length and crypt depth (villous /crypt ratio ≥ 2). Increased inflammation in lamina propria. Increased IEL.
3	Subtotal villous atrophy, elongation of crypts (villous/crypt ratio < 2). Increased inflammation in the lamina propria. Increased IEL.
4	Total villous atrophy, elongation of crypts. Increased inflammation in the lamina propria. Increased IEL.

Table 3. Marsh classification for histology of the small intestinal biopsies.

Type	Histological characteristics
0	Normal
I	Normal villous architecture. Normal crypt depth. Increased IEL.
II	Normal villous architecture. Crypt hyperplasia. Increased IEL.
III	Villous atrophy. Crypt hyperplasia. Increased IEL.
a.	Partial villous atrophy
b.	Subtotal villous atrophy
c.	Total villous atrophy

Immunohistochemistry examination could be used in uncertain cases. In CD, 20–30% of IELs bear $\gamma\delta+$ T-cell receptor-bearing cells, which comprise less than 10% of the IELs in non-celiac subjects [102].

Clinical manifestation

CD has many clinical manifestations, ranging from severely ill young children to children and adults with milder symptoms as well as asymptomatic patients or silent CD [103]. The clinically silent patients are defined as if they had not complained of illness before screening was performed, but some can in a later examination have a low grade of the illness, mainly diagnosed by screening risk groups or in screening studies. Silent CD is an interesting manifestation and more studies are presenting changing features in CD and more silent patients [104–107].

As a result of increased knowledge of the different clinical manifestations, awareness of the disease and widespread screening, more children and adults are now being diagnosed. Despite the fact that undiagnosed cases are many, population screening studies show that the majority of individuals with CD are undiagnosed [25, 108, 109].

The main symptoms of CD in children [110] are gastrointestinal, such as chronic diarrhoea, steatorrhea and abdominal pain, constipation, vomiting and bloating. Growth retardation is usual in children.

The more “classic” CD, characterised by weight loss due to malabsorption, diarrhoea, fatigue and development stagnation, are more common in younger children. CD in school age presents more often with abdominal pain, abnormal linear growth and lack of puberty [104]. The clinical manifestation in children has changed to milder symptoms and later diagnosis in Sweden [105], Italy [107], the UK [106] and the Netherlands [24].

In adults, the clinical manifestations are usually milder and more systemic. Presentation in the form of diarrhoea has decreased (73% to 43%) and the time for onset of symptoms has decreased (9 years to 4.4 years). Common non-gastrointestinal symptoms are weight loss, fatigue or tiredness. Adults often have atypical symptoms from different organs [111]. CD is twice as common in females as it is in males [23].

Dermatitis herpetiformis (DH) was first described in 1884 by Duhring, although Marks et al [112] described the association with gluten-sensitive enteropathy in 1966. Skin symptoms, with itching blisters, appear on the elbows, knees and buttocks. Studies have shown that 25% of DH patients have only increased intraepithelial lymphocytes or normal mucosa and a lack of antibodies [113, 114]. GFD is the treatment of choice [115]. DH even occurs in children [114] but is probably often missed.

Disorders associated with CD are many, listed in Table 4.

Table 4. Disorders associated with celiac disease.

Oral manifestation	Aphthous ulcers Enamel defects
Bone and connective tissue	Osteopenia or osteoporosis Arthralgia, Arthritis, Myalgia
Skin	Dermatitis herpetiformis Alopecia areata
Haematological	Iron, B12 or folic acid deficiency with/without anaemia Lymphoma
Liver diseases	Primary biliary cirrhosis, Elevated transaminase values Autoimmune hepatitis, Autoimmune cholangitis
Endocrinological	Pubertal delay T1D, Thyroid disease, Addison's disease
Gynaecological	Infertility, Recurrent miscarriage
Neurological and psychological disturbances	Delayed development Epilepsy, Cerebellar ataxia, Peripheral neuropathy, Encephalopathy, myopathy Depression, Concentration difficulties Autism
Other diseases	Inflammatory bowel disease, Pancreas diseases

Multiple haematological manifestations are associated with CD as anaemia with iron deficiency (46%), B12 deficiency (8-41%) or folic acid deficiency, often without anaemia in children, as well as thrombocytosis (60%), thrombocytopenia (rare), leucopenia (rare), coagulopathy (18.5% prolonged prothrombin time), venous thrombo-embolism (rare), hyposplenism (21%), and IgA deficiency (2-3%) [116].

IgA-deficiency patients have around a 10% risk of developing CD. The clinical forms are not different and other autoimmune diseases are also increased [117, 118].

Bone mineral density has been found to be lower in untreated CD patients, and in older children, especially these of short stature [119, 120]. Complete recovery is shown on GFD after one year in children [121] and in most adults [122].

Oral manifestations have been reported. *Dental enamel defects* are the oral lesions most closely related to CD [123, 124]. If systematic, the patients should be screened for CD even in the absence of gastrointestinal symptoms. There are conflicting data on the association between CD and recurrent *aphthous stomatitis* [125, 126]. A correlation of CD to *atrophic glossitis* may occur but is seen more often with other conditions, such as B12, iron, folic acid deficiency [127].

Neurological and psychological disturbances are associated with CD, the most common being ataxia, peripheral neuropathy, encephalopathy and myopathy, but also epilepsy without or with the described cerebral calcification, headache and depression. Gluten-related neurological disease is treated with GFD and has a better prognosis if diagnosed early [128, 129].

Autoimmune diseases have been associated with CD [130]. Ventura et al presented a study showing a prevalence of 34% for autoimmune diseases in adults CD and correlation to the duration of gluten exposure [131], although other studies have not been able to confirm this correlation to the duration of gluten exposure [132-134]. Many autoimmune diseases have a CD preponderance to females. The mechanism of this association of autoimmune disease to CD is still a subject for discussion – whether it is secondary to linkage disequilibrium of genes predisposing for both or if CD leads to the onset of other autoimmune diseases in genetically susceptible individuals.

Type 1 diabetes (T1D) is associated with CD in many studies. The average prevalence of CD among children with T1D varies widely, with an average in different studies of 4.1% (0.97-16.5%) [43]. The risk is regardless of which disease is diagnosed first but more common, in 90% of the cases, is that T1D is diagnosed first [135]. The symptom form of CD is often mild or silent in T1D. Meta-analysis for T1D has shown the well-known increased risk of DR3-DQ2/DR4-DQ8, with predisposing to DR4-DQ8 subtypes [136]. T1D is associated with the gene-encoding CTLA-4 +49 [137].

Autoimmune thyroid diseases are associated with CD with a risk of 2-5%. These conditions share similar HLA haplotypes and are associated with CTLA4 +49 [138].

Autoimmune Addison's disease is increased in CD both in children and adults [139, 140].

Liver diseases, such as primary biliary cirrhosis, autoimmune hepatitis and autoimmune sclerosing cholangitis, are associated with CD although isolated hypertransaminasaemia with non-specific histological changes in a liver biopsy is found

in 9% of CD as well as more severe liver damage [141, 142].

Pancreas diseases, such as pancreatitis and pancreatic exocrine dysfunction, are found to be associated with CD [143].

Rheumatoid arthritis has been found to be associated with CD even if other studies have failed to show an association [144, 145].

CD is associated with an increased risk of developing *malignancy*, especially “enteropathy-associated T-cell lymphoma” (EATL) and other gastrointestinal cancers, such as small bowel adenoma. The overall risk of non-Hodgkin’s lymphoma is much lower than previously thought (relative risk of 2-4%) [146, 147]. In ESPGHAN’s inventory of combined paediatric CD and malignancy only 22 cases were found, including three cases of thyroid carcinoma and five of small intestinal lymphoma (four B-cell lymphoma of the Burkitt type and one sarcoma) [148]. There is compelling evidence to suggest that the GFD protects against the development of CD-associated malignancies, especially if started early in life [149].

Refractory CD or refractory sprue, defined as persistent symptoms and villous atrophy despite scrupulous adherence to a GFD, has a rate of occurrence of approximately 5% in adults. RCD can be categorized into type I or type II. Type I RCD has a more favourable prognosis compared with type II. Type II RCD carries a poor prognosis and is more likely to progress to life-threatening malnutrition or intestinal T-cell lymphoma [150].

Treatment

Gluten free diet

GFD is the current treatment for CD. The Codex Alimentarius Commission was created in 1963 by the FAO and WHO to develop food standards, guidelines and related texts, such as codes of practice, under the Joint FAO/WHO Food Standards Programme. Gluten-free products consist of ingredients that do not contain any prolamin from wheat, rye or barley, with a gluten level not exceeding 20 ppm or consisting of ingredients that contain any prolamin from wheat, rye or barley, with a gluten level not exceeding 200 ppm.

The knowledge of the non-toxic effect of oats came with the first studies in 1995 [11] and follow-up studies, both clinical and immunological, have shown that eating oats is not harmful to adults [151-153]. After studies on children [154] oats were allowed in the GFD for CD children in Sweden in 2004 [155].

Future treatments

Living under the regime of a life-long gluten-free diet is burdensome as wheat and related cereals are very commonly used in the food industry. Moreover, the presence of gluten in food is not always obvious. Patients ask for alternative treatment. Future treatments demand to be as safe and effective as GFD. Alternative therapies are therefore of interest. Ongoing studies deal with different types of “blockers” of TG2, the DQ2-mediated antigen presentation or IL15 [156] and the efficiency of gluten degradation by post-proline cutting enzyme [157]. Attempts to generate wheat varieties have not been successful [158].

Molecular genetics

The human genome

In 1980, the human genome project started. The work increased in 1995 and it was completed in 2001 [17, 18].

Our genetic information is stored in *chromosomes*, which are made up of DNA (deoxyribonucleic acid) [10] and genes, are special units of chromosomal DNA. A human has 23 pairs of chromosomes that vary widely in size and shape. Chromosome 1 is the largest and is over three times bigger than chromosome 22. The 23rd pair of chromosomes determines our sex: females XX and males XY. Each chromosome is a very long molecule and so it needs to be wrapped tightly around proteins for efficient packaging of the DNA double helix. Near the centre of each chromosome is its centromere, a narrow region that divides the chromosome into a long arm (q) and a short arm (p). The chromosomes can be divided further using special stains that produce stripes known as a banding pattern. Each chromosome has a distinct banding pattern, and each band is numbered to help identify a particular region of a chromosome. This method of mapping a gene to a particular band of the chromosome is called cytogenetic mapping. The CTLA4/CD28 region, for example, is found on 2q33 and the HLA is found on 6p21.31. [159].

The DNA is about 2 metres long, and 3 cm or 1.5% of the human genome consists of protein-coding exons. However, DNA sequences that do not code protein may still encode functional non-coding RNA molecules, which are involved in the regulation of gene expression. Genomic DNA is located in the cell nucleus of eukaryotes, as well as small amounts in mitochondria (37 genes). A *gene* is a unit of heredity and is a region of DNA, and the complete set or parts of this information in an organism is called its *genotype*. Humans have paired homologous chromosomes in their somatic cells and these contain two copies of each gene. A person who has two copies of the gene that are identical, i.e. have the same *allele*, is described as being *homozygous* for that gene. A person who has two different alleles of the gene is described as being *heterozygous*.

The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide. These bases are classified into two types: adenine and guanine are fused five- and six-membered heterocyclic compounds called purines, while cytosine and thymine are six-membered rings called pyrimidines [159]. Of the three billion base pairs we have 1/1000 SNPs or three million SNPs that have been located. In genetic studies we look at different variances. Only a small part of the DNA has a known function although our knowledge has increased more rapidly during the past year than ever before. Consortium researchers have confirmed the existence of 19,599 protein-coding genes in the human genome and identified another 2,188 DNA segments that are predicted to be protein-coding genes [160].

Recombination

Chromosomal crossover is the process by which two chromosomes, paired up during prophase I of meiosis of a germ cell, an egg or a sperm, exchange some portion of their DNA. Crossover usually occurs when matching regions or matching chromosomes break and then reconnect to the other chromosome. The result of this process is an exchange of genes, called genetic recombination, Figure 4.

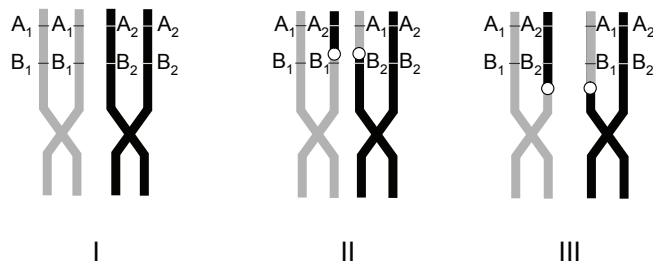


Figure 4. Recombination. Marker A and B can make crossover, in sibling II there is recombination and in sibling III, non-recombination.

This process leads to offspring having different combinations of genes from their parents. Recombination between two loci on chromosome has occurred if one is of maternal and the other of paternal origin. Genetic recombination is necessary for linkage analysis to be able to locate genes on different parts of a chromosome.

The probability of recombination is called recombination fraction; it is a measure of the genetic distance between two loci, measured in centiMorgan (cM). One cM is equal to a 1% chance that a marker at one genetic locus on a chromosome will be separated from a marker at a second locus due to crossing over in a single generation.

Markers

A genetic marker is a known DNA sequence. It can be described as a variation, which may arise due to mutation or alteration in the genomic loci, which can be observed. A genetic marker may be a short DNA sequence, such as a single base-pair change as SNP (single nucleotide polymorphism), or a long one, such as microsatellites that are 1-4 nucleotide long segment repeats. Microsatellites vary between humans and occur throughout the genome. SNP genotyping is the process of determining the single nucleotide polymorphisms of an individual. Tag SNP is a representative single nucleotide polymorphism in a region of the genome with high linkage disequilibrium to other SNPs in the region.

PCR can amplify the markers and fluorescing molecules can make these visible.

The Hap Map project, <http://www.hapmap.org>, now makes it possible to select a limited number of SNPs or CNVs (copy number variant) to capture most of the variation in a given segment of the genome [161].

Complex traits

When the genome was mapped the possibility of carrying out a genome-wide scan for diseases that do not follow the Mendelian segregation or pedigree patterns became a reality. This requires interdisciplinary co-operation between clinicians, geneticists and statisticians.

CD is a disease with a complex inheritance pattern, complex trait or polygenic trait, where it is difficult to see the pedigree patterns. It is then difficult to know anything in detail about disease allele frequencies and penetrances. The phenotypes, clinical manifestation, are many. A complex trait disease is likely to be associated with the effects of multiple genes in combination with lifestyle and environmental factors. The correlation between genotype and disease phenotype can be weak.

Up until last year the history of human genes showed that a few genes of complex traits have been found [162]. Now it seems to be changing. New techniques have made it possible to perform genome-wide association (GWA) studies using SNPs with a 1000-fold greater density compared to the older genome-wide linkage studies. GWA studies use arrays that can examine some 500,000 SNPs at a time, in very large study materials. By tallying which SNPs co-occur with symptoms, the risk associated with each SNP can be determined. This new technique has meant that last year researchers linked variants of more than 50 genes to an increased risk of a dozen diseases. Identifying the relevant genes has been difficult, partly because each causal gene only makes a small contribution to the overall heritability [160]. The new studies have raised the hope that genes influencing complex trait diseases are about to be discovered, such as for diabetes, heart diseases, many autoimmune diseases and infection diseases such as AIDS. The first GWA study on CD has been published, showing evidence that the IL2 and IL21 regions is involved in CD [69].

Genetic analysis

Linkage describes the relationship between different loci on the chromosomes. It is a method used to map genes as searching through the entire genome base by base is not possible in practice. The principle of linkage is the proximity of two or more markers on a chromosome; the closer the markers, the lower the probability that they will be separated during meiosis, and hence the greater the probability that they will be inherited together. That means that if a marker is located beside an inherited disease gene it will more often than not be inherited together with the gene. Genome-wide linkage studies can identify chromosome regions that bring susceptibility genes by examining segregation at several 100 to 1000 loci across the genome.

Association between genetic polymorphisms on the other hand describes the relationship between alleles at different loci and is a statistical statement about the co-occurrence of alleles or phenotypes. Two alleles are genetically associated if they appear together more frequently than we would expect from the allele frequencies. The most common reason for genetic association is Linkage Disequilibrium (LD), which is the

non-random association of alleles at two or more loci. LD describes a situation in which some combinations of alleles or genetic variants occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies. Non-random associations between alleles at different loci are measured by the degree of LD. LD is generally caused by genetic linkage and the rate of recombination, the rate of mutation, the random drift, or non-random mating, and the population structure. It is not the same as linkage, which describes the association of two or more loci on a chromosome with limited recombination between them.

HLA-DQ2, for example, is found in 20-25% of the general Scandinavian population but in over 90% of people with CD and is therefore strongly associated with CD.

An association could have many possible causes, not all of them genetic.

Linkage, on the other hand, is a specific genetic relationship between loci (not alleles or phenotypes). Linkage does not in itself produce any association in the general population. The B locus, for example, is linked to the C locus. Within a family where a C mutation is segregating, we would expect affected people to have the same allele of B, but over the whole population the distribution of B alleles is just the same in people with and without C. Linkage thus creates associations within families, but not among unrelated people. However, if two supposedly unrelated people with disease D have actually inherited it from a distant common ancestor, they may well also tend to share particular ancestral alleles at loci closely linked to D. Where the family and the population merge, linkage and association merge.

Summary: Linkage studies localize chromosomal regions containing disease genes by investing co-inheritance of genetic markers and disease in the families. Association studies assess if a specific allele is more or less frequent in affected individuals. A marker associated with a disease is either the causative gene itself or it is in LD with the causative gene [163].

Genotype-phenotype association

Humans give their children their environment as well as their genotypes.

A genotype is often the largest influencing factor in the development of its phenotype although it is not the only one. Monozygous twins share the same genotype but they have never exactly the same phenotype.

Phenotype describes mostly as alternative A in Figure 5. But our knowledge is changing and maybe alternative B is closer to the reality.

Epigenetic is heritable changes in gene expression that do not change the DNA sequence but rather provide an “extra” layer of transcriptional control that regulates how genes are expressed. Our understanding of the mechanism of the interplay between epigenetic gene expression and the environment is still finite. Epigenetic research aims to understand heritable gene regulation that is not directly encoded in the DNA sequence. Epigenetic mechanisms, such as DNA methylation and histone modifications, modulate the packaging of the DNA in the nucleus and thereby influence gene expression. Patterns of epigenetic information are faithfully propagated

A: Phenotype = Genotype + Environment + Genotype*Environment (interaction)

OR

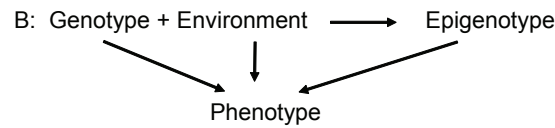


Figure 5. The phenotype of human is influenced directly by its genotype and its environment. In addition, these two factors indirectly influence phenotype via the epigenotype [164].

over multiple-cell divisions, which make epigenetic regulation a key mechanism for cellular differentiation and cell fate decisions. In addition, incomplete erasure of epigenetic information can lead to complex patterns of non-Mendelian inheritance. Stochastic (when two persons do not react the same to treatment for the same diagnosis) and environment-induced epigenetic defects are known to play a major role in cancer and ageing, and they may also contribute to mental disorders and autoimmune diseases [164].

The knowledge of the interaction between genotype and phenotype in CD is limited. Several studies have failed to establish associations between different HLA-genotypes and phenotypes. In an Italian study of 145 CD patients, including 27 silent patients, no correlation was found between HLA-genotypes and phenotypes [165]. In this study there were ten DQ8-positives, two of them asymptomatic, and five DQ2-negative/DQ8-negatives, all of them symptomatic. In other 28 symptom-discordant sib-pairs no association between symptoms and DR-DQ haplotypes was found. Of the patients, 25% were DQ2 homozygous. The symptomatic CD seemed to have earlier onset than silent CD [166]. In two studies however, a gene dose effect of DQB1*02 on the phenotype was found. Homozygous patients had a moderate overrepresentation of classic presentation, female gender, lower age at diagnosis and a shorter delay between onset of symptoms and diagnosis. In this study four patients were DQ8-positive patients, all had classic symptoms, and of four DQ2-DQ8-negative patients one had classic symptoms, two had atypical symptoms and one was asymptomatic [167]. In the other study DQB1*02 homozygous were correlated to severity of the villous atrophy, lower age at diagnosis and lower hemoglobin values at diagnosis [168].

In the clinical presentation of 25 DQ2-negative patients in a study of Finnish and Spanish patients [169], all were assessed as typical of CD. Of the 25 DQ2-negative patients there were two patients without any part of DQ2 or DQ8. They did not differ clinically from the other patients.

One study showed no association between IL12B and IRF1 genes, in the 5q31-33 region, and enteropathy grading according to the Marsh criteria [170].

The role of epigenetic in CD is unknown although Megiorni et al showed a major distortion in the DR3-DQ2 transmission from fathers to daughters and suggested a possible different effect of parent-specific epigenetic modifications in the two genders [171].

The role played by environmental factors in CD is still not clear – if, for example, breastfeeding prevents or simply delays symptoms or the role of infections. See above.

Aim of the study

The aim of the thesis was to:

Screen for CD in apparently healthy members of nuclear families with two affected siblings and to estimate the risk of CD in the remaining siblings and parents.

Perform a genome-wide scan in a population of Scandinavian families with CD, to identify non-HLA candidate chromosomal regions showing linkage to CD.

Investigate the heritability of the phenotype in CD and the influence on the phenotype of different genes associated with the disease.

Compare the clinical presentation between DQ2-negative and DQ2-positive CD patients in a European population.

Patients and Methods

Patients

Multiplex families (Paper I, II, III and IV)

Families with at least two affected siblings were collected from Sweden and southern Norway in order to perform a genome-wide scan for celiac genes and genotype-phenotype association studies (Papers I-IV).

The aim was to collect blood samples and clinical data from at least one hundred affected sib-pairs, their parents and their healthy siblings if any.

Three lines of action were used for recruiting families: 1. Advertisement in the journal of The Coeliac Society of Sweden. 2. Inquiries to all Swedish paediatric departments. 3. Invitation to all CD sib-pair families of the registries of our own paediatric departments. By collecting the families in these different ways more sib-pair families were recruited and the range of clinical presentation in the material was wider.

A semi-structured telephone interview was performed when entering the study. The interview included questions concerning the diagnosis, the gluten consumption, general health and possible symptoms of parents and siblings without CD diagnosis.

In order to avoid inclusion of falsely diagnosed cases, strict inclusion and exclusion criteria were applied. Medical records of individuals with diagnosed CD were collected and scrutinized. Diagnostic data and clinical manifestations were recorded. Only families where all CD siblings were diagnosed according to the ESPGHAN 90 criteria [88] were included. Children under the age of two years at the time of diagnosis were included in the study only if they fulfilled the original ESPGHAN 70 [87] with three biopsies on different diets. Individuals on GFD without CD diagnosis were excluded.

Blood samples for DNA extraction and serological screening were collected from all CD patients, the healthy parents and healthy siblings available.

Of the 152 families recruited, 113 families met the inclusion criteria. Six families dropped out and the remaining 107 families with two to four affected siblings met the diagnostic criteria and were included in the study (Paper I, Figure 1 and Table 2). Of 102 additional siblings, eight were not included in the study: seven were living abroad and one was excluded because he was on a gluten-free diet without diagnosed CD.

In the screening study (Paper I), families without healthy siblings available for participation in the study were excluded as well as individuals without CD diagnosis not eating a gluten-containing diet. Altogether, 65 families (56 with two affected, 8 with three affected and one with four affected) with 94 siblings were included in the calculations of the risk of CD in a third sibling. A total of 187 of the 192 non-affected parents were available for screening. Twenty-two parents had a previously diagnosed CD, one family had no parents available and three families had only one parent available.

The median age at diagnosis in the 224 affected siblings (76 males and 148 females)

was 2.2 years (range 6 months – 58 years). The median age at diagnosis in the 22 affected parents (10 males and 12 females) was 39 years (range 19–78 years).

Simplex families (Paper IV)

For confirmation studies a simplex material consisting of 135 families with one affected child was collected from south-west Sweden. An invitation was sent to CD families of the registries at our own paediatric department in the Gothenburg region. The same strict inclusion and exclusion criteria were applied as in the multiplex material. Medical records of individuals with diagnosed CD were collected and reviewed. Diagnostic data and clinical manifestations were recorded. Blood samples for DNA extraction were collected from all CD patients and the healthy parents available as well as samples for serologic screening.

At the start of the study there were two CD parents (both females, one with DH) although seven new CD parents were found through screening (three males and four females).

The mean age at diagnosis in the 135 Swedish affected CD (44 males and 90 females) was 1.9 years and the median age 1.4 years (range 7 months – 12 years).

DQ2-negative cases and DQ2-positive controls (Paper IV)

The patients in this study were among those collected by the different European partners in the EU financed research consortium on CD, The European Cluster on Coeliac Disease (QLKT – 1999-00037) [64]. The cases were DQ2-negative from France, Italy, Finland and Sweden and the DQ2-positive controls from the same materials were matched only by country. A total of 85 DQ2-negative cases (38 DQ8-positive and 47 DQ2-negative/DQ8-negative) and 102 DQ2-positive controls were collected. This material is described in detail in Paper IV. The same diagnostic criteria as described previously in multiplex and simplex materials were used for the Swedish patients and at least the ESPGHAN 90 [88] for the patients from Finland, France and Italy.

A questionnaire was used to collect clinical data retrospectively from the medical files. Distribution of age at diagnosis and gender are presented in Table 2, Paper IV.

Ethics

The ethics committee of respective centre approved the study.

Methods

Serologic markers

IgA-EMA was the antibody of choice used as a screening method for CD in both the multiplex and simplex materials [89]. IgA-EMA was used in clinical practice at the time of the study [172]. Sera with IgA-EMA antibodies detected at a dilution of 1:10 or more were considered positive. All IgA-EMA positive siblings and parents

without previously diagnosed CD underwent small intestinal biopsy. IgA deficiency was excluded in negative individuals using routine laboratory methods.

Small intestinal biopsy

The biopsies were done according to the routine of the home clinic. The biopsy routines at clinics in Sweden and Norway are quite similar. Children were mainly investigated using capsule biopsy [173] and adults sing endoscopic biopsies. The histology reports were evaluated and classified by me according to a modified Alexander scale (Table 2) [99, 174]. When the written reports were difficult to assess or evaluate the histological slides were re-evaluated by a pathologist specialised in paediatric gastroenterology (Walter Ryd, Department of Pathology, Sahlgrenska University Hospital, Göteborg, Sweden). Alexander grade 3 or more was considered compatible with CD, used for the multiplex and simplex family materials.

In Paper IV MARSH 3 (Table 3) [100, 101] was considered indicative of CD.

Statistics

Linkage analysis uses a non-parametric approach. A study of allele-sharing between relatives. Relatives concordant for disease should have increased allele-sharing close to a susceptibility locus, while relatives who are discordant should have decreased allele sharing. No special model of inheritance is assumed and it is therefore a suitable method for complex diseases such as CD. One common non-parametric approach is the affected sib-pair method, which counts the alleles that a sib-pair has inherited in common, IBD at a marker locus. According to the Mendel's first law of segregation, the siblings would have, at any non-linked locus, a 25% probability of inheriting 0 alleles IBD, a 50% probability of inheriting one allele IBD (1 maternal or paternal shared allele) and a 25% probability of inheriting two alleles IBD (both the maternal and paternal alleles shared).

In the calculation of the risk of CD in the members of nuclear families with two affected siblings, we denoted K_{2s} as the risk that an additional sib to an affected sib-pair is affected and K_{2o} as the risk of a parent of an affected offspring-pair being affected (page 340, Paper I).

Bootstrapping is a computer method [175] that we used for estimating confidence intervals for the risk in Paper I. New samples of the same size are drawn with replacement from the original sample. The idea behind bootstrap is to produce a variety of values whose variability reflects what would be obtained if samples were repeatedly taken from the whole population.

	Median
Original sample: 11, 22, 33, 44, 55, 66, 77, 88, 99	55
Bootstrap 1: 22, 33, 55, 55, 33, 99, 88, 44, 11	44
Bootstrap 2: 44, 55, 22, 11, 11, 66, 99, 88, 88	55

This is repeated 1000 times. The medians are then ordered by increasing value. The 25th and 975th values out of 1,000 give the lower and upper estimates of the 95% confidence interval.

The **GENEHUNTER** computer program [176] was used for linkage and linkage disequilibrium analyses of the results of the fragment analysis of the microsatellites in Paper II.

Allegro is another computer program for multipoint linkage analysis, which can analyse pedigrees of larger size [177]. To analyse possible associations between IBD sharing at CTLA4 and 5q31-33 on the one hand and the phenotype on the other, we used the non-parametric linkage (NPL) statistics. Families were divided into two groups according to symptom gradation, one containing symptom-concordant sib-pairs and one containing symptom-discordant sib-pairs (Table II, Paper III). For each group, NPL was calculated using **Allegro** 1.2, thus obtaining two statistics, NPL_{conc} and NPL_{disc} .

Permutation analysis with 10,000 iterations was further used to assess the significance of the difference between NPL_{conc} and NPL_{disc} . Permutation analysis shows the range of statistics possible under the null hypothesis – that there is no association between marker genotype and phenotype. The idea is to keep the original statistic T_{obs} , e.g. a correlation between two variables, and then by repeatedly shuffling either variable destroy all but random associations between them and calculate a statistic T_i for each permutation i . Finally, the statistic T_{obs} is compared to all T_i ($i = 1$ to n) n typically large (often 10,000 or more). All the T_i form an empirical distribution from which an empirical p-value can be estimated.

The heritability of the symptom grade was established using the software **SOLAR** (Sequential Oligogenic Linkage Analysis Routines) a software that has the general pedigree variance component and IBD estimation methods [178].

The one-sided **TDI** was used to analyse transmission of A and G alleles in the high and low HLA risk groups in Paper II.

Spearman's rank correlation coefficient does not require the assumption that the relationship between the variables is linear with normally distributed residual, nor does it require the variables to be measured on interval scales; it can be used for variables measured at the ordinal level as in our case age at diagnosis.

Fisher's exact test or Chi-square test was used for comparing two groups.

Mantel-Haenszel test was used for comparing data from several 2x2 tables.

Statistical stratification

Only one sibling from each sib-pair family was used (Papers III and IV) to avoid bias, as the siblings' phenotype can be familiar for genetic as well as non-genetic reasons.

The phenotype distribution as well as the genotype distribution differs between countries in Paper IV. Therefore, the groups were compared first with Fisher's exact test or chi-square tests within each country and secondly, if the odds ratios were found to be homogenous between countries, a Mantel-Haenszel statistics was used.

Genetic analysis

Analysis strategy

A two-step strategy was used for the analysis. First we selected 70 families, which we considered to be the most informative, i.e. those with a large number of children and affected individuals (Group A). The remaining 36 families (Group B) were added to group A in a second step and were genotyped over selected regions based on one of three different criteria. First criterion: chromosomes implicated by the previously published CD genome-wide screens: chromosomes 5, 11 and 15 [179-181]. Second criterion: novel chromosomes identified from Group A, which showed NPL-values of above 2.0. Third criterion: chromosome 2 and 20 because of the location of two candidate genes, suggested as being involved in CD: CTLA4 +49 and TG2. (Figure 1, Paper II). In total, 137 markers across nine chromosomes: 2, 5, 6, 9, 11, 14, 17, 20 and X were analysed for all 106 nuclear families (Groups A and B). The remaining 261 markers on the other chromosomes were only analysed for the initial 70 families (Group A).

Microsatellite genotyping

In Paper II, 106 sib-pair were analysed with total of 398 microsatellites. Weber screening set version six from Research Genetics, containing 390 microsatellite markers with an average distance between the markers of 10 cM (<ftp://ftp.resgen.com/pub/mappairs/humanset>), was used. Additional markers were used in two regions, typed over a 2 - 5 cM region: in the HLA class II region on chromosome 6 and in the CTLA4/CD28 region on chromosome 2q [72, 182]. We amplified the microsatellite marker regions using an ABI 877PCR robot under standard conditions. PCR products were separated by electrophoresis on an ABI 377XL sequencer and a 5% denaturing polyacrylamide gel, Figure 6. Genotyping was performed using GENESCAN ANALYSIS 2.1 and GENOTYPER 2.0 software (<http://www.appliedbiosystems.com>).

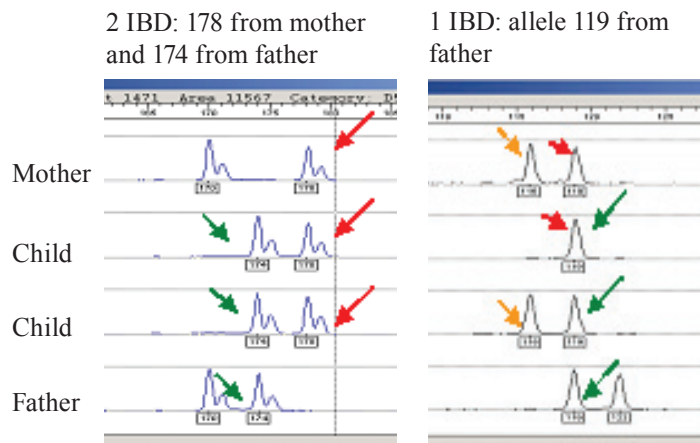


Figure 6. Two microsatellite markers fluorescently labelled.

The linkage analysis was done with the non-parametric statistic NPLall using the software GENEHUNTER version 2.0 for autosomals and version 1.3 for the X chromosome. NPLall was used in multipoint mode and uses the disease status from all family members and takes into account, for example, if one parent is affected and produces higher scores if an affected parent transmits a chromosome IBD.

In our studies the following genotypes have been found to have an association and/or linkage to CD and were used in Paper III: association to CTLA4 +49 A/G by TDT and linkage using NPL analysis [72]. Analysed SNPs on MH30, -1147, CT60 and CT61 observed strong LD. A haplotype of this region marked by the alleles -1147*T: +49*A:CT60*G:CT61*A was associated significantly with CD [183]. IBDs from linkage on 5q31-33 [184]. The HLA genotyping was done previously [68].

For the case-control association study in Paper IV, the previously performed HLA genotyping was used [64].

Phenotypes

I did the clinical classification, retrospectively from the medical records (Papers III and IV) or from a questionnaire by the person responsible for each country (Paper IV). The patients were classified into three symptom grades: Grade 1: Patients with “classic” celiac symptoms such as diarrhoea, vomiting, abdominal distension malabsorption and growth failure. Grade 2: Patients with milder presentation of symptoms typical of CD. Grade 3: Originally we had few patients with atypical symptoms (three patients in Paper III and five patients in Paper IV) but for statistical analysis purposes they were grouped as Grade 2. Patients with clinically silent CD are defined as asymptomatic individuals with positive antibodies and gluten enteropathy.

Age at diagnosis, age at onset of symptoms, gender, being a proband, presence of gastrointestinal symptoms and presence of autoimmune disorder or DH were recorded.

Genotype-phenotype association (Paper III-IV)

In Paper III the genotyping was classified into four groups: 1. HLA-risk groups: high when a patient had either of the two genotypes, DR3-DQ2/DR3-DQ2 or DR3-DQ2/DR7-DQ2, intermittent when DR3-DQ2/X (X is any other genotype) or DR5-DQ7/DR7-DQ2 and low when the patient had any other HLA genotype. 2. The CTLA4 +49 A/G allele (AA, AG, GG). 3. The haplotype MH30*G:-1147*T:+49*A:CT60*G:CT61*A. 4. The 5q31-33 region, IBD status at locus D5S436.

In Paper IV, the HLA-DQ2-negative cases were sub-grouped as DQ8-positive and DQ2-negative/DQ8-negative patients (Figure 1, Paper IV) and compared to DQ2-positive controls.

Results

Paper I

Fourteen apparently healthy relatives were IgA-EMA positive, six siblings and eight parents. All underwent a small intestinal biopsy, which confirmed the CD diagnosis in all except one parent (Table 4, Paper I). The risk for the remaining siblings and parents in sib-pair families was increased, as expected according to Falconer's theory [159, 185]. We could confirm a threefold increased risk compared to families with only one child with CD. The sibling risk was 26.3% (95% CI, 13.8-38.7) and the parent risk 12.9% (95% CI, 9.0-17.1).

Table 5. Estimated risk for CD in population and in first-degree relatives in single and sib-pair families.

	Population risk	Siblings	Parents
Populations studies	0.2 – 1.0 %		
Single CD case families		2.6 – 12.2 %	0 – 6.3 %
Sib-pair CD families		26.3 %	12.9 %

An unexpected male preponderance was found among the new CD cases (10 males: 3 females) compared to all CD cases (86 males: 160 females) at the start of the study.

Paper II

The genome-wide scan of 106 celiac families and 398 microsatellite markers. In total, chromosomes 2, 5, 6, 9, 11, 14, 15, 20 and the X chromosomes were analyzed in all families and the remaining chromosomes were analysed in 70 families. Apart from HLA on chromosome 6 (NPL 4.40) our best results were found in eight chromosome regions, 2q11-13, 3p24, 5q31-33, 9p21, 11p15, 11q23-25, 17q22 and Xp11 (Table 2, Paper II). Taking into account that chromosomes 5 and 11 had been reported previously, our results strengthened the notion that these regions could harbour susceptibility genes for CD.

Paper III

The heritability of the phenotype was statistically significant and was estimated at 0.45 ($p = 0.02$). Association analysis (Table III, Paper III) showed a statistically significant dependence between CTLA4 +49 A/G genotypes and the different symptom grades ($p = 0.014$), with more AA genotypes than expected in the clinically silent group.

No significant association between symptom grade and the other genotypes was found.

The stratified linkage analysis showed a difference in allele-sharing at CTLA4 +49A/G among symptom-concordant sib-ship families (NPLconc = 2.0) compared to the

symptom-discordant sib-ship families (NPLdisc = -0.7) ($p = 0.04$). The 5q31-33 locus, however, had no significant difference ($p = 0.13$) in allele-sharing between the two groups.

No significant correlation was found between age at diagnosis in the probands and the genotypes or between sex and the genotypes. Nor was any significant association between sex and symptom grade found.

Paper IV

First we compared the DQ2-negative cases and DQ2-positive controls (Table 2, Paper IV). The clinical grades were significantly different between the cases and the controls in Italy ($p = 0.006$) and Sweden ($p = 0.014$). In both countries there was an association between classic presentation (*Grade 1*) and DQ2-negative cases, and in the Italian group there was an additional association between DQ2-negative cases and silent presentation. No significant differences between cases and controls were found for age at diagnosis, age of onset of symptoms, gender or gastrointestinal symptoms. The DQ2-negative cases in the Italian group were diagnosed significantly more often by screening ($p = 0.006$). In the Finnish cases a milder intestinal histology was found ($p = 0.001$). There were no significant differences between cases and controls for DH or other autoimmune diseases. When conditions for analysing all countries as one group were achieved, a Mantel-Haenszel test (HM) was performed. There were no significant differences between cases and controls for these other phenotypes.

In the next step the cases in the DQ8-positive and DQ2-negative/DQ8-negative subgroups were compared in the same way (Table 4, Paper IV).

Autoimmune diseases were significantly ($p = 0.007$) overrepresented in the DQ8-positive subgroup compared to DQ2-negative/DQ8-negative. There were no significant differences between DQ8-positive and DQ2/DQ8-negative subgroups for the other phenotypes, including clinical symptom grades.

In the DQ2-negative/DQ8-negative subgroup all patients except three encoded half of the DQ2 heterodimer, either *DQA1*05* or *DQB1*02*. These three individuals were from the Italian group and satisfied the ESPGHAN diagnostic criteria for CD [88]. All of them were diagnosed in childhood. One patient was six years old at diagnosis. He was diagnosed in a family screening and had a clinically silent CD. The other two were both two years of age at diagnosis and had classic symptoms such as diarrhoea, failure to thrive and vomiting. One of them had a sibling with CD. These cases are previously described in detail [64].

Autoimmune disease was diagnosed in six DQ2-negative patients. All of them were DQ8-positive. Three had type 1 diabetes mellitus (T1D), one thyroid disease, one rheumatic disease and one ulcerative colitis. Thus six out of the 38 (16%) DQ8-positive cases had another autoimmune disease compared to none of the 47 DQ8-negative cases ($p = 0.026$). Furthermore, five of the six patients were homozygous for DQ8; the patient with ulcerative colitis was the only heterozygous. In the DQ2-positive CD controls only two (one T1D, one rheumatic disease) out of 102 (2%) had another autoimmune disease, which was significantly less compared to the DQ8-positive cases ($p = 0.008$).

Conclusion and discussion

This thesis was a co-operative endeavour between physicians, geneticists and statisticians to discover the non-HLA genes in CD as well as to increase the knowledge of HLA genetics. Furthermore, performing genotype-phenotype analysis is an attempt to understand the complexity of the phenotypes. This work required collaboration with other centres in Sweden and other European countries.

Some families have stronger genetic susceptibility for CD as in families with two or more affected siblings. These affected sib-pair families are important for genetic studies to find common susceptibility genes.

The strategy for collecting the best material for genetic analysis and genotype-phenotype analysis in CD is as follows: Large sib-pair material, reliable diagnosis in the cases, age at diagnosis, gender and clinical symptoms grades.

The advantage of the Swedish/Norwegian material is that the risk of patients with a false diagnosis is minimal as review of all investigation was done with as much thought as possible. The other patients are part of genetic study materials and are diagnosed according to the ESPGHAN 90 criteria. All healthy relatives were set as unknown because, as in a disease such as CD, one screening cannot eliminate the possibility of being affected later.

Almost all parents' DNA was available and so the risk of genotype error was minimal as we could check the allele or haplotype. The patient materials are family-based, which can be powerful when combining both genetic linkage and association, and they are also regional population-based.

All the IgA-EMA positive individuals underwent a small intestinal biopsy to confirm the CD diagnosis and minimise the risk of false-positive cases. Latent or potential CD is also an unknown factor here. We are aware of siblings that had their CD diagnosis a few years after the study came to an end and our results are thus probably slightly underestimated. Considering the high level of knowledge of CD in these families, the number of undiagnosed cases is surprisingly high and many had symptomatic CD. We suggested that serological screening should be offered to all first-degree relatives of CD patients. Clinical practice is often otherwise. The time for screening children is always a debate as screening only once in younger children is not enough. The natural course of undetected CD is still a matter for debate – whether the risk of other autoimmune diseases, anaemia, osteoporosis and malignancy is increased to the same extent for patients who are clinically silent as for symptomatic patients. Should those that have positive antibodies without gluten enteropathy receive treatment with GFD or only be offered follow-up biopsy. One parent in our study with positive IgA-EMA and increased IEL cell infiltration in the intestinal biopsy was symptomatic and commenced GFD with good clinical response.

The increased risk for remaining siblings and parents shown in our study is in accordance with another study, which showed a sibling risk of 21.3% (95% CI, 11.0-31.6) where they even calculated an offspring risk of 14.7% (95% CI, 6.2-23.2), second-degree relatives 19.5% (95% CI, 15.1-23.9) and first cousins 17% (95% CI, 6.8-27.7) [186].

For genetic analysis, it could be a good strategy to first analyse genetic information in region- or ethnic-based study material and then make confirmation analysis in other materials. This material, one of the largest sib-pair materials of CD published, is relatively homogeneous with regard to regional ethnicity with mostly Swedish Caucasian families and some south Norwegian Caucasian families. In a complex disease such as CD, all susceptibility genes are not likely to cause disease in all families. Many independent sib-pair families are preferable in GWL studies of complex disease as there is a good chance of finding those common susceptibility genes that cause disease in the majority of cases. The different phenotypes can include age at diagnosis, gender, clinical manifestation, associated diseases and HLA-DQ type. In this material there are patients with differences with regard to age at diagnosis as well as different clinical phenotypes and we also had access to the parents' genetic information. The two regions on chromosome 5q31-33 and 11q23-25 that our GWL study indicated as true susceptibility regions have been reported in other sib-pair GWL studies, which makes these regions interesting for the further study of susceptibility genes in CD. The GWL study was also included in a meta analysis together with three other European partners (Finland, Italy and UK), where chromosome 5q31-33 was the only significant locus, apart from HLA [77]. Further studies in our research group have been published and some are ongoing on 5q31-3 [78, 184].

The genotype-phenotype study included the susceptible gene and regions that had shown association and or linkage in our material. Paper III is the first study to publish a genotype-phenotype analysis for CTLA4 +49 A/G in CD. Our findings indicate that this gene may influence the phenotype of CD. Surprisingly, we found that the AA genotype was correlated to the silent phenotype. As the genetic association was to the A allele [72] one might think that the AA genotype would be associated more to symptomatic disease. The role of CTLA4 +49 in CD is not known. A few studies have conducted a genotype-phenotype analysis for CTLA4 +49 A/G in other autoimmune diseases. In MS, a genotype-phenotype study showed that AA and AG genotypes are associated with MS progression [187] but in rheumatoid arthritis no CTLA4 +49 association to the phenotypes was found [188].

No association was found with the different HLA risk groups, the MH30*G:-1147*T:+49*A:CT60*G:CT61*A haplotype or the 5q31-33 region.

Knowledge of genotype-phenotype can help to stratify the material by sub-phenotype, which selects for those specific individuals who carry a rare gene variant and increase the chance of detecting linkage or association due to rare variants within a given sample set.

HLA is a necessary and single most important genetic factor in CD. The HLA-DQ2 is responsive in 90% of the patients and together with DQ8 predisposes to disease development with a known mechanism (Figure 2). The DQ8-positive and the DQ2-negative/DQ8-negative patients are 5% each of the CD population. We had the opportunity to compare the clinical presentation in DQ2-negative cases with DQ2-positive controls in a large DQ2-negative sample. Silent presentation was correlated to DQ2-negative status but only in the Italian group. The other countries had too few patients with silent presentation to detect such a pattern. However, analysing only symptomatic patients, DQ2-negativity was associated with classic symptoms or

Grade 1, seen both in the Italian and the Swedish groups. Our results differ from a study of 25 DQ2-negative cases in Finland and Spain, where no clinical differences compared to the DQ2-positive controls were found [12].

In clinical practice, HLA, the use of serology tests that test for DQ2 and DQ8-heterodimers, has become more common. However, they need to be used with caution as our results showed that CD could not be excluded by only testing whether or not patients are DQ2 and DQ8-positive. There were more silent patients in the DQ2-negative cases with an even division between DQ8-positive and DQ2-negative/DQ8-negative CD patients. However, most of the DQ2-negative/DQ8-negative patients have one half of the DQ2 heterodimer, except for the three patients that had the non-part of the heterodimer but had classic CD symptoms.

Our effort was to collect two controls for every DQ2-negative and country, which was not accomplished. There were less DQ2-positive controls from Italy than we had planned. We therefore needed to analyse each country separately in case of genetic or clinical differences between the countries.

Other autoimmune diseases were looked at in Papers III and IV and despite the fact that the samples have small amounts there were some interesting findings. T1D was the disease that occurred most frequently. Interestingly, there were no patients with autoimmune disease carrying homozygous for CTLA4 +49 A-allele or who shared 2 haplotype IBD with their co-sibling in the 5q31-33 region. Autoimmune diseases were significantly overrepresented in the DQ8-positive subgroup compared to the DQ2-negative/DQ8-negative subgroup. Genetic analyses are ongoing in most autoimmune diseases associated with CD and many are looking at genes in the same region. It is most likely that there are some common genes in these diseases, probably with some genes with immunological function.

The significance of interaction or epistasis of different genes is not possible to determine from this thesis as no gene-to-gene interaction analysis has been done. Theoretically, there are probably gene interactions and the genotypes could have a different significance in different patients e.g. HLA high-risk patients may not need any or few other candidate genes whereas HLA low-risk patients need other genes or several genes to get CD. The genetic effects may be understood before we understand how different environments may interact with specific genetic components that lead to the disease.

Future genetic and genotype-phenotype studies in CD

This thesis has increased the knowledge of the genetic and the genotype-phenotype association in CD. The results support the theory that there could be different genetic susceptibility in different phenotypes. The main findings were the association shown between CTLA4 +49 A/G genotypes and the different symptom grades with more AA genotypes than were expected in the clinically silent group. DQ2-negative cases were significantly different compared to DQ2-positive controls with regard to clinical grades. Autoimmune diseases were significantly overrepresented in the DQ8-positive subgroup compared to DQ2-negative/DQ8-negative.

The research into the genetics of common or complex diseases is now in an exciting phase. There are new technologies in genetic research, such as GWA studies that have been listed as “the breakthrough of the year” 2007 in *Science* [160]. More than 50 new disease-susceptibility genes have been found in different common diseases in just one year. This will hopefully offer a completely new view of CD and other immune and autoimmune diseases. Discover new potential candidate genes that could increase the potential for genotype-phenotype analysis. Increase our understanding of the role that genes play in disease pathogenesis and design molecular mechanism-based therapies that will improve clinical outcome or quality of life.

One GWA study on CD has been published [69]. The HLA locus came out with very highly significant P values as expected. The second strongest association was found for a few SNPs located close to the IL2 and IL21 genes on chromosome 4q27. A follow-up study by the same group of >1,000 markers showing association signals in the initial celiac GWA study has identified an additional seven loci associated with CD. The regions are 1q31, 2q11-12, 3p21, 3q25-26, 3q28, 4q27, 6q25 and 12q24. Immune-related genes map to six of these regions, underscoring the possibility that CD has an immunological basis [189].

Fine-mapping the new regions will be required to find potential candidate genes and provides better conditions for further functional and clinical studies. Moreover, there are probably modifying effects between the genetic (epistasis) and environmental factors and interactions between the genetic and environmental factors and epigenetic factors. It may also be that in different patients with identical phenotypes, only some genetic/environmental factors are shared and only some of the pathways leading to disease development are identical. This makes genotype-phenotype analysis difficult.

When gene-gene (epistasis) analysis with true candidate genes and optimal patient material, with all the phenotypes and perhaps larger subgroups as clinical symptom grades, HLA-subgroups, CD patients with other autoimmune diseases etc, we might succeed in our ambition to walk along the research path (Figure 1) towards disease susceptibility proteins that will increase our understanding of CD and hopefully other autoimmune diseases. By associating phenotypes to specific profiles of gene/protein expression pathological mechanisms can be unveiled, predictive instruments

in future diagnosis can be found and candidates for specific therapies can be selected.

In summary, this thesis shows that the risk for third siblings and parents is as expected increased in sib-pair families, as the expected risk of being affected in polygenic diseases is higher in multiple-case families compared to single-case families. The genome scan indicated significant linkage to 11q and 5q, which makes these regions interesting for further fine-mapping of these regions using association analysis. Genotype-phenotype analysis of both HLA and non-HLA locus showed some significant correlation between silent CD and both CTLA4 +49 AA genotype and the DQ2-negatives. In addition, an association between classical symptoms and DQ2-negative cases has been shown, when only the symptomatic patients were analysed.

Sammanfattning på svenska

Sjukdomen celiaki (CD) eller glutenintolerans är en kronisk inflammatorisk sjukdom av autoimmun typ och är en av de vanligaste sjukdomarna i barnaåren, men diagnostiseras i alla åldrar. Förekomsten av CD i Sverige är runt 1 % men kan vara betydligt högre. Risken för ett syskon (syskonrisken) att bli sjuk är 10 % om man har ett sjukt syskon. CD är mångfaktoriell sjukdom med samspel av troligtvis flera gener och omgivningsfaktorer. CD är unik sjukdom genom att man känner den utlösande faktorn, äggviteämnet i gluten, som leder till inflammation och skada på tunntarmens ludd. För att kunna ställa rätt diagnos krävs tunntarmsbiopsi, men screening med serologisk antikroppstest görs ofta först. Behandling med gluten fri kost utan vete, råg och korn, leder till utläkning. Diagnos kriterier för CD, eller ESPGHAN criteria, grundar sig på att tunntarmsslemhinnans morfologi i biopsi preparat karaktäriseras av villusatrofi, krypthyperplasi och ökat antal av intraepitheliala lymphocyter (IEL). Sjukdomsbilden eller fenotypen varierar från svårt sjuka barn till barn och vuxna med mindre symptom och till patienter utan symptom eller symptomfri (silent) CD.

Genetiken har stor betydelse vid sjukdomen. HLA är en typ av antigenpresenterande molekyler som existerar på cellytan hos alla celler i människokroppen. Dess funktion är att presentera utvalda peptider för immunförsvarets celler. HLA klass II på kromosom 6 är väl känd vid CD. HLA-DQ2 är den allra vanligaste molekylen men en mindre del har HLA-DQ8. Andra HLA typer kan förekomma och kallas här DQ2-negativa/DQ8-negativa. HLA genen kan bara förklara runt 40 % av sjukdomens genetik så resten måste förklaras av non-HLA gener. Dessa kandidatgener är ännu inte kända men eventuella kandidatgen områden har identifierats. Genotyp-fenotyp association vid CD är begränsad.

Målsättningen med detta arbete var att beräkna syskonrisken hos det tredje syskonet och föräldrar till två sjuka syskon, genomföra systematisk screening för att kartlägga misstänkta sjukdomsregioner och undersöka genotyp-fenotyp association vid CD.

Material insamlades från Sverige och Norge med 107 familjer med minst två sjuka barn, totalt 224 CD syskon, samt deras friska syskon och de flesta av deras föräldrar. Screening för CD genomfördes hos friska syskon och deras föräldrar. Tretton nya fall av CD diagnostiserades, 6 syskon och 7 föräldrar. Den beräknade syskonrisken var 26.3% och föräldrarisken var 12.9% eller nästan tre gånger högre än om ett syskon är sjukt. Det höga antal ny diagnostiserade fall är förvånande med tanke på den höga kunskapsnivån om sjukdomen som finns i dessa familjer. Vi föreslår att man erbjuder alla första grad släktingar serologisk screening för CD.

Systematisk screening av alla kromosomer gjordes med kopplingsanalys och genetiska varianter för att identifiera kromosomregioner som nedärvs tillsammans med sjukdomen. Förutom HLA, visade den signifikant koppling till sjukdomen på kromosom 11 och 5q31-33. Område 5q31-33 har visat koppling i flera andra studier.

Ytterligare material med 136 svenska familjer med ett sjukt barn samlades in för genetiska associationsanalyser.

Två genotyp-fenotyp association studier genomfördes. Den första på de kandidatgener eller genområden som hade visat association och/eller koppling i vårt mate-

rial. Dessa var HLA klass II risk grupper, CTLA4 +49 A/G polymorfier, MH30*G:-1147*T:+49*A:CT60*G: CT61*A haplotypen och 5q31-33 området.

Patienterna delades upp i olika sjukdomsbilder vid insjuknande, ålder vid diagnos och kön. Heritability för fenotypen var kalkylerad till 0.45. Det är den del av fenotypen som förklaras av genotypen.

Genotypen AA i CTLA4 +49A/G polymorfier visade association till den symptomfria sjukdomsbilden. Ingen association fanns mellan de andra genotyperna och symptom bilden, ålder vid diagnos eller kön.

Genotyp-fenotyp analys gjordes mellan fenotyperna hos DQ2-negativa CD patienter, i den största DQ2-negativa CD grupp som har publicerats, jämfört med DQ2-positiv CD kontroller i en Europeisk population. Den kliniska symptom bilden skilde sig signifikant mellan DQ2-negativa and DQ2-positiva CD patienter i Italien and Sverige. I båda länderna fanns associationen mellan DQ2-negativa sjukdomsfallen och den klassiska symptom bilden. Hos patienter i Italien fanns desutom association mellan den symptomfria sjukdomsbilden och de DQ2-negativa sjukdomsfallen. Autoimmuna sjukdomar var signifikant överrepresenterad i DQ8-positiva patienter.

Ökad kunskap om genetiska faktorer och omgivningsfaktorer samt interaktionen mellan dessa leder till ökad kunskap om patogenesen t.ex. de immunologiska mekanismerna vid CD och även andra autoimmuna sjukdomar. Detta kan på sikt leda till ökade möjligheter till primärprevention och nya behandlingsformer.

Acknowledgements

The celiac disease patients and their families are condition for the clinical and genetic research in this thesis. Therefore I want to express my sincere thanks to all the families that participated in my studies, both in Sweden and Norway but even those I have never spoken to in Finland, France and Italy.

These studies have required interdisciplinary cooperation between clinicians, geneticists and statisticians. The teamwork has made the research more profitable and pleasant.

I want to express my warmest thanks to all my collaborators:

Henry Ascher, my supervisor, for his guidance into the world of celiac disease research.

Bengt Kristiansson, my former supervisor, for his guidance in the beginning of this project.

Britt-Marie Käck, my research nurse, without all your enthusiasm and patience we would not have managed to collect all the samples from near and far.

Åsa Torinsson Naluai, my co-author, who introduced me into the world of genetic.

Staffan Nilsson, my co-author, for your introduction to the genetic statistic.

Svetlana Adamovic, my co-author, for your support and collaboration.

Jan Wahlström, Lena Samuelsson, Tommy Martinsson, my co-authors, at the Dep. Of Clinic Genetics, for their collaboration and **Birgitta Hallberg**, for helping with the samples.

Johan Ek, my colleague in Drammen, Norway, for your collaboration identifying sib-pair families.

Ludvig Sollid, Andrew Louka and Silja Amundsen, in the Oslo group, Norway, for your collaboration.

Kirsi Mustalahti, in Finland, **J-P Hugot and Fabienne Clot**, in France, **Iolanda Coto and Selvaggia Percopo**, in Italy, my co-authors, for your collaboration in the European study.

Walter Ryd, at the Dep. Pathology, Sahlgrenska, for you help with checking the biopsies.

Ola Hjalmarson, professor at the Dep. of Pediatrics, for your good advices and friendly support.

Marie Krantz, Robert Saalman, Carola Kullberg-Lindh and Karin Hallberg, my colleagues at the division of Paediatric Gastroenterology and Nutrition, and **all the staff at ward 334**, for your friendly support and for taking care of my patient when I was doing research.

Ann-Sofie Petersson, at the photograph studio, for helping me with the layout of this thesis.

My friends in Göteborg, for your friendship and support through the years.

My friends in Iceland, for you have always been there for me when I'm at Home.

My family, in Iceland and Uppsala, for you have always been there for me both in good as well as bad times.

This thesis has been supported by grants from:

The Queen Silvia's Jubilee Fund, Göteborg Medical Society, Swedish Medical Society, The European Commission (QLRT 1999-00037), Foundation for Strategic Research, Swedish Research Council (2003-5560), The Sven Jerring Fund, The research funds of the Children's Hospital in Göteborg and The Wilhelm and Martina Lundgren Research Foundation.

References

1. Jabri, B. and L.M. Sollid, *Mechanisms of disease: immunopathogenesis of celiac disease*. Nat Clin Pract Gastroenterol Hepatol, 2006. **3**(9): p. 516-25.
2. Jabri, B., D.D. Kasarda, and P.H. Green, *Innate and adaptive immunity: the yin and yang of celiac disease*. Immunol Rev, 2005. **206**: p. 219-31.
3. Sollid, L.M. and B. Jabri, *Is celiac disease an autoimmune disorder?* Curr Opin Immunol, 2005. **17**(6): p. 595-600.
4. Dieterich, W., et al., *Identification of tissue transglutaminase as the autoantigen of celiac disease [see comments]*. Nat Med, 1997. **3**(7): p. 797-801.
5. Gee, S.J., *On the celiac affection* St. Bartholomew's Rep, 1888. **24**(17).
6. Herter, C.A., *On infantilism from chronic intestinal infection* New York: MacMillan, 1908. **14**.
7. Haas, S.V., *Value of banana in treatment of celiac disease*. Am J Dis Child, 1924. **24**: p. 421-37.
8. Dicke, W.K., *Investigation of the harmful effects of certain type of cereals on patients with celiac disease (Thesis)*. The Netherlands: University of Utrecht, 1950.
9. Anderson, C.M., A.C. Frazer, and et.al., *Coeliac disease. Gastrointestinal studies and the effect of dietary wheat flour*. Lancet, 1952. **i**: p. 836-42.
10. Dicke, W.K., H.A. Weijers, and J.H. Van De Kamer, *Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease*. Acta Paediatr, 1953. **42**(1): p. 34-42.
11. Janatuinen, E.K., et al., *A comparison of diets with and without oats in adults with celiac disease*. N Engl J Med, 1995. **333**(16): p. 1033-7.
12. Shiner, M., *Duodenal biopsy*. Lancet, 1956(i): p. 17-19.
13. Berger, E., *Zur allergischen Pathogenese der Cöliakie*. Bibliotheca Paediatrica., 1958. **67** (suppl.1): p. 1-55
14. Macdonald, W.C., W.O. Dobbins, 3rd, and C.E. Rubin, *Studies of the Familial Nature of Celiac Sprue Using Biopsy of the Small Intestine*. N Engl J Med, 1965. **272**: p. 448-56.
15. Louka, A.S. and L.M. Sollid, *HLA in coeliac disease: Unravelling the complex genetics of a complex disorder*. Tissue Antigens, 2003. **61**(2): p. 105-17.
16. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.
17. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
18. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.
19. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 599-603.
20. Ascher, H., I. Krantz, and B. Kristiansson, *Increasing incidence of coeliac disease in Sweden*. Arch Dis Child, 1991. **66**(5): p. 608-11.
21. Cavell, B., et al., *Increasing incidence of childhood coeliac disease in Sweden. Results of a national study*. Acta Paediatr, 1992. **81**(8): p. 589-92.
22. Ivarsson, A., et al., *Epidemic of coeliac disease in Swedish children*. Acta Paediatr, 2000. **89**(2): p. 165-71.

23. Ivarsson, A., et al., *The Swedish coeliac disease epidemic with a prevailing twofold higher risk in girls compared to boys may reflect gender specific risk factors*. Eur J Epidemiol, 2003. **18**(7): p. 677-84.
24. Steens, R.F., et al., *A national prospective study on childhood celiac disease in the Netherlands 1993-2000: an increasing recognition and a changing clinical picture*. J Pediatr, 2005. **147**(2): p. 239-43.
25. Ivarsson, A., et al., *High prevalence of undiagnosed coeliac disease in adults: a Swedish population-based study*. J Intern Med, 1999. **245**(1): p. 63-8.
26. Hovdenak, N., et al., *High prevalence of asymptomatic coeliac disease in Norway: a study of blood donors*. Eur J Gastroenterol Hepatol, 1999. **11**(2): p. 185-7.
27. Not, T., et al., *Celiac disease risk in the USA: high prevalence of antiendomysium antibodies in healthy blood donors*. Scand J Gastroenterol, 1998. **33**(5): p. 494-8.
28. Catassi, C., et al., *The coeliac iceberg in Italy. A multicentre antigliadin antibodies screening for coeliac disease in school-age subjects*. Acta Paediatr Suppl, 1996. **412**: p. 29-35.
29. Maki, M. and K. Holm, *Incidence and prevalence of coeliac disease in Tampere. Coeliac disease is not disappearing*. Acta Paediatr Scand, 1990. **79**(10): p. 980-2.
30. Csizmadia, C.G., et al., *An iceberg of childhood coeliac disease in the Netherlands*. Lancet, 1999. **353**(9155): p. 813-4.
31. Carlsson, A.K., et al., *Serological screening for celiac disease in healthy 2.5-year-old children in Sweden*. Pediatrics, 2001. **107**(1): p. 42-5.
32. Fasano, A., et al., *Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study*. Arch Intern Med, 2003. **163**(3): p. 286-92.
33. Ascher, H., et al., *Influence of infant feeding and gluten intake on coeliac disease*. Arch Dis Child, 1997. **76**(2): p. 113-7.
34. Ascher, H., *Childhood Coeliac Disease in Sweden, epidemiological changes and possible associated factors.*, in *Department of Paediatrics*. 1996, Göteborg University: Göteborg.
35. Mylotte, M., et al., *Family studies in coeliac disease*. Q J Med, 1974. **43**(171): p. 359-69.
36. Shipman, R.T., et al., *A family study of coeliac disease*. Aust N Z J Med, 1975. **5**(3): p. 250-5.
37. Stern, M., et al., *Serum antibodies against gliadin and reticulin in a family study of coeliac disease*. Eur J Pediatr, 1980. **135**(1): p. 31-6.
38. Rolles, C., et al., *The familial incidence of asymptomatic coeliac disease*. In: McConell R, editor. *The genetic of coeliac disease*. Lancaster: MTP press., 1981: p. 235-44.
39. Sagaro, E. and N. Jimenez, *Family studies of coeliac disease in Cuba*. Arch Dis Child, 1981. **56**(2): p. 132-3.
40. Stenhammar, L., A. Brandt, and J. Wagermark, *A family study of coeliac disease*. Acta Paediatr Scand, 1982. **71**(4): p. 625-8.
41. Maki, M., et al., *Serological markers and HLA genes among healthy first-degree relatives of patients with coeliac disease*. Lancet, 1991. **338**(8779): p. 1350-3.
42. Greco, L., et al., *The first large population based twin study of coeliac disease*. Gut, 2002. **50**(5): p. 624-628.
43. Holmes, G.K., *Screening for coeliac disease in type 1 diabetes*. Arch Dis Child, 2002. **87**(6): p. 495-8.
44. Csizmadia, C.G., et al., *Accuracy and cost-effectiveness of a new strategy to screen for celiac disease in children with Down syndrome*. J Pediatr, 2000. **137**(6): p. 756-61.
45. Carnicer, J., et al., *Prevalence of coeliac disease in Down's syndrome*. Eur J Gastroenterol Hepatol, 2001. **13**(3): p. 263-7.

46. Hansson, T., et al., *Celiac disease in relation to immunologic serum markers, trace elements, and HLA-DR and DQ antigens in Swedish children with Down syndrome*. J Pediatr Gastroenterol Nutr, 1999. **29**(3): p. 286-92.
47. Bonamico, M., et al., *Prevalence and clinical picture of celiac disease in Italian Down syndrome patients: a multicenter study*. J Pediatr Gastroenterol Nutr, 2001. **33**(2): p. 139-43.
48. Mackey, J., et al., *Frequency of celiac disease in individuals with Down syndrome in the United States*. Clin Pediatr (Phila), 2001. **40**(5): p. 249-52.
49. Book, L., et al., *Prevalence and clinical characteristics of celiac disease in Down syndrome in a US study*. Am J Med Genet, 2001. **98**(1): p. 70-4.
50. Ivarsson, S.A., et al., *Prevalence of coeliac disease in Turner syndrome*. Acta Paediatr, 1999. **88**(9): p. 933-6.
51. Gillett, P.M., et al., *Increased prevalence of celiac disease in girls with Turner syndrome detected using antibodies to endomysium and tissue transglutaminase*. Can J Gastroenterol, 2000. **14**(11): p. 915-8.
52. Bonamico, M., et al., *Prevalence and clinical picture of celiac disease in Turner syndrome*. J Clin Endocrinol Metab, 2002. **87**(12): p. 5495-8.
53. Giannotti, A., et al., *Coeliac disease in Williams syndrome*. J Med Genet, 2001. **38**(11): p. 767-8.
54. Stenberg, P., E.B. Roth, and K. Sjoberg, *Transglutaminase and the pathogenesis of coeliac disease*. Eur J Intern Med, 2008. **19**(2): p. 83-91.
55. Molberg, O., et al., *Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease*. Nat Med, 1998. **4**(6): p. 713-7.
56. Monsuur, A.J. and C. Wijmenga, *Understanding the molecular basis of celiac disease: what genetic studies reveal*. Ann Med, 2006. **38**(8): p. 578-91.
57. Falth-Magnusson, K., et al., *Infant feeding history shows distinct differences between Swedish celiac and reference children*. Pediatr Allergy Immunol, 1996. **7**(1): p. 1-5.
58. Ivarsson, A., et al., *Breast-feeding protects against celiac disease*. Am J Clin Nutr, 2002. **75**(5): p. 914-21.
59. Norris, J.M., et al., *Risk of celiac disease autoimmunity and timing of gluten introduction in the diet of infants at increased risk of disease*. Jama, 2005. **293**(19): p. 2343-51.
60. Akobeng, A.K., et al., *Effect of breast feeding on risk of coeliac disease: a systematic review and meta-analysis of observational studies*. Arch Dis Child, 2006. **91**(1): p. 39-43.
61. Stene, L.C., et al., *Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study*. Am J Gastroenterol, 2006. **101**(10): p. 2333-40.
62. Ivarsson, A., et al., *Children born in the summer have increased risk for coeliac disease*. J Epidemiol Community Health, 2003. **57**(1): p. 36-9.
63. Sollid, L.M., et al., *Genes and environment in celiac disease*. Acta Odontol Scand, 2001. **59**(3): p. 183-6.
64. Karell, K., et al., *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, 2003. **64**(4): p. 469-77.
65. Sollid, L.M., et al., *Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer*. J Exp Med, 1989. **169**(1): p. 345-50.
66. Sollid, L.M., *Molecular basis of celiac disease*. Annu Rev Immunol, 2000. **18**: p. 53-81.
67. Bolognesi, E., et al., *Additional factor in some HLA DR3/DQ2 haplotypes confers a four-fold increased genetic risk of celiac disease*. Tissue Antigens, 2003. **61**(4): p. 308-16.

68. Louka, A.S., et al., *HLA in coeliac disease families: a novel test of risk modification by the 'other' haplotype when at least one DQA1*05-DQB1*02 haplotype is carried*. Tissue Antigens, 2002. **60**(2): p. 147-54.
69. van Heel, D.A., et al., *A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21*. Nat Genet, 2007. **39**(7): p. 827-9.
70. Petronzelli, F., et al., *Genetic contribution of the HLA region to the familial clustering of coeliac disease*. Ann Hum Genet, 1997. **61**(Pt 4): p. 307-17.
71. Djilali-Saiah, I., et al., *CTLA-4 gene polymorphism is associated with predisposition to coeliac disease*. Gut, 1998. **43**(2): p. 187-9.
72. Naluai, A.T., et al., *The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders*. Tissue Antigens, 2000. **56**(4): p. 350-5.
73. Haimila, K., et al., *Genetic association of coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33*. Genes Immun, 2004. **5**(2): p. 85-92.
74. Mora, B., et al., *CTLA-4 +49 A/G dimorphism in Italian patients with celiac disease*. Hum Immunol, 2003. **64**(2): p. 297-301.
75. Clot, F., et al., *Linkage and association study of the CTLA-4 region in coeliac disease for Italian and Tunisian populations*. Tissue Antigens, 1999. **54**(5): p. 527-30.
76. Kouki, T., et al., *CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease*. J Immunol, 2000. **165**(11): p. 6606-11.
77. Babron, M.C., et al., *Meta and pooled analysis of European coeliac disease data*. Eur J Hum Genet, 2003. **11**(11): p. 828-34.
78. Amundsen, S.S., et al., *A comprehensive screen for SNP associations on chromosome region 5q31-33 in Swedish/Norwegian celiac disease families*. Eur J Hum Genet, 2007. **15**(9): p. 980-7.
79. Louka, A.S., et al., *The IL12B gene does not confer susceptibility to coeliac disease*. Tissue Antigens, 2002. **59**(1): p. 70-2.
80. Ryan, A.W., et al., *Haplotype variation at the IBD5/SLC22A4 locus (5q31) in coeliac disease in the Irish population*. Tissue Antigens, 2004. **64**(2): p. 195-8.
81. Van Belzen, M.J., et al., *A major non-HLA locus in celiac disease maps to chromosome 19*. Gastroenterology, 2003. **125**(4): p. 1032-41.
82. Hunt, K.A., et al., *Lack of association of MYO9B genetic variants with coeliac disease in a British cohort*. Gut, 2006. **55**(7): p. 969-72.
83. Giordano, M., et al., *A family-based study does not confirm the association of MYO9B with celiac disease in the Italian population*. Genes Immun, 2006. **7**(7): p. 606-8.
84. Amundsen, S.S., et al., *Association analysis of MYO9B gene polymorphisms with celiac disease in a Swedish/Norwegian cohort*. Hum Immunol, 2006. **67**(4-5): p. 341-5.
85. Matter, K. and M.S. Balda, *Signalling to and from tight junctions*. Nat Rev Mol Cell Biol, 2003. **4**(3): p. 225-36.
86. Adamovic, S., et al., *Association study of IL2/IL21 and FcγRIIa: significant association with the IL2/IL21 region in Scandinavian coeliac disease families*. Genes and Immunity, 2008.
87. Meeuwisse, G., *Diagnostic criteria in Coeliac disease*. Acta Paediatric Scand, 1970. **59**: p. 461-3.
88. Walker-Smith, J., *Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition*. Arch Dis Child, 1990. **65**(8): p. 909-11.
89. Ascher, H., et al., *Value of serologic markers for clinical diagnosis and population studies of coeliac disease*. Scand J Gastroenterol, 1996. **31**(1): p. 61-7.

90. Chorzelski, T.P., et al., *IgA class endomysium antibodies in dermatitis herpetiformis and coeliac disease*. Ann N Y Acad Sci, 1983. **420**: p. 325-34.
91. Rostom, A., et al., *The diagnostic accuracy of serologic tests for celiac disease: a systematic review*. Gastroenterology, 2005. **128**(4 Suppl 1): p. S38-46.
92. Alaedini, A. and P.H. Green, *Autoantibodies in celiac disease*. Autoimmunity, 2008. **41**(1): p. 19-26.
93. Wong, R.C., et al., *A comparison of 13 guinea pig and human anti-tissue transglutaminase antibody ELISA kits*. J Clin Pathol, 2002. **55**(7): p. 488-94.
94. Wolters, V., et al., *Human tissue transglutaminase enzyme linked immunosorbent assay outperforms both the guinea pig based tissue transglutaminase assay and anti-endomysium antibodies when screening for coeliac disease*. Eur J Pediatr, 2002. **161**(5): p. 284-7.
95. Schwertz, E., et al., *Serologic assay based on gliadin-related nonapeptides as a highly sensitive and specific diagnostic aid in celiac disease*. Clin Chem, 2004. **50**(12): p. 2370-5.
96. Agardh, D., *Antibodies against synthetic deamidated gliadin peptides and tissue transglutaminase for the identification of childhood celiac disease*. Clin Gastroenterol Hepatol, 2007. **5**(11): p. 1276-81.
97. Volta, U., et al., *Usefulness of Antibodies to Deamidated Gliadin Peptides in Celiac Disease Diagnosis and Follow-up*. Dig Dis Sci, 2007.
98. Ferguson, A. and D. Murray, *Quantitation of intraepithelial lymphocytes in human jejunum*. Gut, 1971. **12**(12): p. 988-94.
99. Alexander, J., *The small intestine in dermatitis herpetiformis*. Dermatitis herpetiformis. Major problems in dermatology; 4., ed. J. Alexander. Vol. 4. 1975, London: W.B. Saunders Co. Ltd. 236-80.
100. Marsh, M.N., *Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue')*. Gastroenterology, 1992. **102**(1): p. 330-54.
101. Marsh, M.N., *The immunopathology of the small intestinal reaction in gluten-sensitivity*. Immunol Invest, 1989. **18**(1-4): p. 509-31.
102. Halstensen, T.S., H. Scott, and P. Brandtzaeg, *Intraepithelial T cells of the TcR gamma/delta+ CD8- and V delta 1/J delta 1+ phenotypes are increased in coeliac disease*. Scand J Immunol, 1989. **30**(6): p. 665-72.
103. Hill, I.D., et al., *Celiac disease: Working Group Report of the First World Congress of Pediatric Gastroenterology, Hepatology, and Nutrition*. J Pediatr Gastroenterol Nutr, 2002. **35 Suppl 2**: p. S78-88.
104. Ascher, H., et al., *Different features of coeliac disease in two neighbouring countries*. Arch Dis Child, 1993. **69**(3): p. 375-80.
105. Ludvigsson, J.F., et al., *Symptoms and signs have changed in Swedish children with coeliac disease*. J Pediatr Gastroenterol Nutr, 2004. **38**(2): p. 181-6.
106. Ravikumara, M., D.P. Tuthill, and H.R. Jenkins, *The changing clinical presentation of coeliac disease*. Arch Dis Child, 2006. **91**(12): p. 969-71.
107. Garampazzi, A., et al., *Clinical pattern of celiac disease is still changing*. J Pediatr Gastroenterol Nutr, 2007. **45**(5): p. 611-4.
108. Maki, M., et al., *Prevalence of Celiac disease among children in Finland*. N Engl J Med, 2003. **348**(25): p. 2517-24.
109. Tommasini, A., et al., *Mass screening for coeliac disease using antihuman transglutaminase antibody assay*. Arch Dis Child, 2004. **89**(6): p. 512-5.
110. Fasano, A., *Clinical presentation of celiac disease in the pediatric population*. Gastroenterology, 2005. **128**(4 Suppl 1): p. S68-73.

111. Green, P.H., *The many faces of celiac disease: clinical presentation of celiac disease in the adult population*. Gastroenterology, 2005. **128**(4 Suppl 1): p. S74-8.
112. Marks, J., S. Shuster, and A.J. Watson, *Small-bowel changes in dermatitis herpetiformis*. Lancet, 1966. **2**(7476): p. 1280-2.
113. Collin, P. and T. Reunala, *Recognition and management of the cutaneous manifestations of celiac disease: a guide for dermatologists*. Am J Clin Dermatol, 2003. **4**(1): p. 13-20.
114. Alonso-Llamazares, J., L.E. Gibson, and R.S. Rogers, 3rd, *Clinical, pathologic, and immunopathologic features of dermatitis herpetiformis: review of the Mayo Clinic experience*. Int J Dermatol, 2007. **46**(9): p. 910-9.
115. Fry, L., et al., *Gluten-free diet and reintroduction of gluten in dermatitis herpetiformis*. Arch Dermatol, 1969. **100**(2): p. 129-35.
116. Halfdanarson, T.R., M.R. Litzow, and J.A. Murray, *Hematologic manifestations of celiac disease*. Blood, 2007. **109**(2): p. 412-21.
117. Collin, P., et al., *Selective IgA deficiency and coeliac disease*. Scand J Gastroenterol, 1992. **27**(5): p. 367-71.
118. Latiff, A.H. and M.A. Kerr, *The clinical significance of immunoglobulin A deficiency*. Ann Clin Biochem, 2007. **44**(Pt 2): p. 131-9.
119. Tau, C., et al., *Bone mineral density in children with celiac disease. Effect of a Gluten-free diet*. Eur J Clin Nutr, 2006. **60**(3): p. 358-63.
120. Pludowski, P., et al., *Skeletal and muscular status in juveniles with GFD treated clinical and newly diagnosed atypical celiac disease--preliminary data*. J Clin Densitom, 2007. **10**(1): p. 76-85.
121. Barera, G., et al., *Body composition in children with celiac disease and the effects of a gluten-free diet: a prospective case-control study*. Am J Clin Nutr, 2000. **72**(1): p. 71-5.
122. Valdimarsson, T., et al., *Reversal of osteopenia with diet in adult coeliac disease*. Gut, 1996. **38**(3): p. 322-7.
123. Farmakis, E., J.W. Puntis, and K.J. Toumba, *Enamel defects in children with coeliac disease*. Eur J Paediatr Dent, 2005. **6**(3): p. 129-32.
124. Bossu, M., et al., *Enamel hypoplasia in coeliac children: a potential clinical marker of early diagnosis*. Eur J Paediatr Dent, 2007. **8**(1): p. 31-7.
125. Olszewska, M., J. Sulej, and B. Kotowski, *Frequency and prognostic value of IgA and IgG endomysial antibodies in recurrent aphthous stomatitis*. Acta Derm Venereol, 2006. **86**(4): p. 332-4.
126. Robinson, N.A. and S.R. Porter, *Low frequency of anti-endomysial antibodies in recurrent aphthous stomatitis*. Ann Acad Med Singapore, 2004. **33**(4 Suppl): p. 43-7.
127. Pastore, L., et al., *Oral Manifestations of Celiac Disease*. J Clin Gastroenterol, 2008.
128. Hadjivassiliou, M., R.A. Grunewald, and G.A. Davies-Jones, *Gluten sensitivity as a neurological illness*. J Neurol Neurosurg Psychiatry, 2002. **72**(5): p. 560-3.
129. Bushara, K.O., *Neurologic presentation of celiac disease*. Gastroenterology, 2005. **128**(4 Suppl 1): p. S92-7.
130. Petaros, P., et al., *Prevalence of autoimmune disorders in relatives of patients with celiac disease*. Dig Dis Sci, 2002. **47**(7): p. 1427-31.
131. Ventura, A., G. Magazzu, and L. Greco, *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, 1999. **117**(2): p. 297-303.
132. Sategna Guidetti, C., et al., *Duration of gluten exposure in adult coeliac disease does not correlate with the risk for autoimmune disorders*. Gut, 2001. **49**(4): p. 502-5.

133. Cataldo, F. and V. Marino, *Increased prevalence of autoimmune diseases in first-degree relatives of patients with celiac disease*. J Pediatr Gastroenterol Nutr, 2003. **36**(4): p. 470-3.
134. Viljamaa, M., et al., *Coeliac disease, autoimmune diseases and gluten exposure*. Scand J Gastroenterol, 2005. **40**(4): p. 437-43.
135. Ludvigsson, J.F., et al., *Celiac disease and risk of subsequent type 1 diabetes: a general population cohort study of children and adolescents*. Diabetes Care, 2006. **29**(11): p. 2483-8.
136. Thomson, G., et al., *Relative predispositional effects of HLA class II DRB1-DQB1 haplotypes and genotypes on type 1 diabetes: a meta-analysis*. Tissue Antigens, 2007. **70**(2): p. 110-27.
137. Bartsocas, C.S. and A. Gerasimidi-Vazeou, *Genetics of type 1 diabetes mellitus*. Pediatr Endocrinol Rev, 2006. **3 Suppl 3**: p. 508-13.
138. Larizza, D., et al., *Celiac disease in children with autoimmune thyroid disease*. J Pediatr, 2001. **139**(5): p. 738-40.
139. Betterle, C., et al., *Celiac disease in North Italian patients with autoimmune Addison's disease*. Eur J Endocrinol, 2006. **154**(2): p. 275-9.
140. Elfstrom, P., et al., *Risk of primary adrenal insufficiency in patients with celiac disease*. J Clin Endocrinol Metab, 2007. **92**(9): p. 3595-8.
141. Abdo, A., J. Meddings, and M. Swain, *Liver abnormalities in celiac disease*. Clin Gastroenterol Hepatol, 2004. **2**(2): p. 107-12.
142. Rubio-Tapia, A. and J.A. Murray, *The liver in celiac disease*. Hepatology, 2007. **46**(5): p. 1650-8.
143. Freeman, H.J., *Hepatobiliary and pancreatic disorders in celiac disease*. World J Gastroenterol, 2006. **12**(10): p. 1503-8.
144. Nishihara, R.M., et al., *Rheumatoid Arthritis and anti-endomysial antibodies*. Acta Rheumatol Port, 2007. **32**(2): p. 163-67.
145. Stagi, S., et al., *Thyroid function, autoimmune thyroiditis and coeliac disease in juvenile connective tissue diseases*. Clin Exp Rheumatol, 2005. **23**(2): p. 277.
146. Askling, J., et al., *Cancer incidence in a population-based cohort of individuals hospitalized with celiac disease or dermatitis herpetiformis*. Gastroenterology, 2002. **123**(5): p. 1428-35.
147. Catassi, C., I. Bearzi, and G.K. Holmes, *Association of celiac disease and intestinal lymphomas and other cancers*. Gastroenterology, 2005. **128**(4 Suppl 1): p. S79-86.
148. Schweizer, J.J., A. Oren, and M.L. Mearin, *Cancer in children with celiac disease: a survey of the European Society of Paediatric Gastroenterology, Hepatology and Nutrition*. J Pediatr Gastroenterol Nutr, 2001. **33**(1): p. 97-100.
149. Peters, U., et al., *Causes of death in patients with celiac disease in a population-based Swedish cohort*. Arch Intern Med, 2003. **163**(13): p. 1566-72.
150. Al-Toma, A., W.H. Verbeek, and C.J. Mulder, *Update on the management of refractory coeliac disease*. J Gastrointest Liver Dis, 2007. **16**(1): p. 57-63.
151. Janatuinen, E.K., et al., *No harm from five year ingestion of oats in coeliac disease*. Gut, 2002. **50**(3): p. 332-5.
152. Kempainen, T., et al., *No observed local immunological response at cell level after five years of oats in adult coeliac disease*. Scand J Gastroenterol, 2007. **42**(1): p. 54-9.
153. Storsrud, S., et al., *Adult coeliac patients do tolerate large amounts of oats*. Eur J Clin Nutr, 2003. **57**(1): p. 163-9.
154. Hogberg, L., et al., *Oats to children with newly diagnosed coeliac disease: a randomised double blind study*. Gut, 2004. **53**(5): p. 649-54.

155. Stenhammar, L., L. Hogberg, and R. Saalman, [*The Pediatric Association recommends: Oats can be implemented in the gluten-free diet*]. *Lakartidningen*, 2004. **101**(18): p. 1610-1.
156. Xia, J., et al., *Cyclic and dimeric gluten peptide analogues inhibiting DQ2-mediated antigen presentation in celiac disease*. *Bioorg Med Chem*, 2007. **15**(20): p. 6565-73.
157. Mitea, C., et al., *Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease*. *Gut*, 2008. **57**(1): p. 25-32.
158. Molberg, O., et al., *Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease*. *Gastroenterology*, 2005. **128**(2): p. 393-401.
159. Strachan, T. and A. Read, *Human Molecular Genetic 3*. 3rd ed. 2004, Oxford, UK.: BIOS Scientific Publishers Ltd.
160. Pennisi, E., *Breakthrough of the year. Human genetic variation*. *Science*, 2007. **318**(5858): p. 1842-3.
161. Frazer, K.A., et al., *A second generation human haplotype map of over 3.1 million SNPs*. *Nature*, 2007. **449**(7164): p. 851-61.
162. Glazier, A.M., J.H. Nadeau, and T.J. Aitman, *Finding genes that underlie complex traits*. *Science*, 2002. **298**(5602): p. 2345-9.
163. Sollid, L.M., *Hunting for celiac disease genes*. *Gastroenterology*, 2008. **134**(3): p. 869-71.
164. van Vliet, J., N.A. Oates, and E. Whitelaw, *Epigenetic mechanisms in the context of complex diseases*. *Cell Mol Life Sci*, 2007. **64**(12): p. 1531-8.
165. Greco, L., et al., *Lack of correlation between genotype and phenotype in celiac disease*. *J Pediatr Gastroenterol Nutr*, 1998. **26**(3): p. 286-90.
166. Mustalahti, K., et al., *Genetic dissection between silent and clinically diagnosed symptomatic forms of coeliac disease in multiplex families*. *Dig Liver Dis*, 2002. **34**(12): p. 842-5.
167. Zubillaga, P., et al., *HLA-DQA1 and HLA-DQB1 genetic markers and clinical presentation in celiac disease*. *J Pediatr Gastroenterol Nutr*, 2002. **34**(5): p. 548-54.
168. Nistico, L., et al., *Concordance, disease progression, and heritability of coeliac disease in Italian twins*. *Gut*, 2006. **55**(6): p. 803-8.
169. Polvi, A., et al., *HLA-DQ2-negative celiac disease in Finland and Spain*. *Hum Immunol*, 1998. **59**(3): p. 169-75.
170. Seegers, D., et al., *IL12B and IRF1 gene polymorphisms and susceptibility to celiac disease*. *Eur J Immunogenet*, 2003. **30**(6): p. 421-5.
171. Megiorni, F., et al., *HLA-DQ and Susceptibility to Celiac Disease: Evidence for Gender Differences and Parent-of-Origin Effects*. *Am J Gastroenterol*, 2008.
172. Ladinser, B., E. Rossipal, and K. Pittschieler, *Endomysium antibodies in coeliac disease: an improved method*. *Gut*, 1994. **35**(6): p. 776-8.
173. Stenhammar, L., et al., *Small bowel biopsy in Swedish paediatric clinics*. *Acta Paediatr*, 2002. **91**(10): p. 1126-9.
174. Fallstrom, S.P., B. Kristiansson, and W. Ryd, *Histological studies of small-intestinal biopsies from infants with low rate of weight gain*. *Acta Pathol Microbiol Scand [A]*, 1981. **89**(6): p. 431-8.
175. Campbell, M.J., *Bootstrapping. Statistics at square two*. *BMJ Books*. 2001: p. 115-119.
176. Kruglyak, L., et al., *Parametric and nonparametric linkage analysis: a unified multipoint approach*. *Am J Hum Genet*, 1996. **58**(6): p. 1347-63.
177. Gudbjartsson, D.F., et al., *Allegro, a new computer program for multipoint linkage analysis*. *Nat Genet*, 2000. **25**(1): p. 12-3.
178. Almasy, L. and J. Blangero, *Multipoint quantitative-trait linkage analysis in general pedigrees*. *Am J Hum Genet*, 1998. **62**(5): p. 1198-211.

179. Greco, L., et al., *Genome search in celiac disease*. Am J Hum Genet, 1998. **62**(3): p. 669-75.
180. Zhong, F., et al., *An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland*. Nat Genet, 1996. **14**(3): p. 329-33.
181. Houlston, R.S., et al., *Linkage analysis of candidate regions for coeliac disease genes*. Hum Mol Genet, 1997. **6**(8): p. 1335-9.
182. Holopainen, P., et al., *CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study*. Tissue Antigens, 1999. **53**(5): p. 470-5.
183. Amundsen, S.S., et al., *Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease*. Tissue Antigens, 2004. **64**(5): p. 593-9.
184. Adamovic, S., et al., *Fine mapping study in Scandinavian families suggests association between coeliac disease and haplotypes in chromosome region 5q32*. Tissue Antigens, 2008. **71**(1): p. 27-34.
185. Falconer, D. and T. Mackay, *Quantitative Genetics*. 1996, Essex: Longman Group Ltd.
186. Book, L., J.J. Zane, and S.L. Neuhausen, *Prevalence of celiac disease among relatives of sib pairs with celiac disease in U.S. families*. Am J Gastroenterol, 2003. **98**(2): p. 377-81.
187. Bilinska, M., et al., *Progression of multiple sclerosis is associated with exon 1 CTLA-4 gene polymorphism*. Acta Neurol Scand, 2004. **110**(1): p. 67-71.
188. Karlson, E.W., et al., *Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study*. Ann Rheum Dis, 2008. **67**(3): p. 358-63.
189. Hunt, K.A., et al., *Newly identified genetic risk variants for celiac disease related to the immune response*. Nat Genet, 2008. **40**(4): p. 395-402.