Preface and outline of the thesis

Before the Human Genome Project was completed in 2003, scientific and medical attention for genetic disorders was mainly focused on understanding rare single-gene disorders, such as Huntington’s disease, Duchene muscular dystrophy, and cystic fibrosis, as well as on chromosomal abnormalities. In recent years, with the further completion of the international HapMap Project and the development of new methods for genotyping individual DNA samples for more than 500,000 markers, the attention of genomics and genetic researchers has shifted toward understanding the basis of common multifactorial disorders, such as celiac disease, type 2 diabetes, coronary heart disease and cancer.

Celiac disease (CD) is one of the most common inflammatory disorders of the small intestine caused by permanent gluten intolerance in susceptible individuals. It affects around 1% of Western populations but remains largely unrecognized. The only treatment is a lifelong gluten-free diet. The age at onset ranges from infancy to late adulthood, and the clinical presentation of the disease is highly variable ranging from common gastrointestinal symptoms like diarrhea and abdominal pain to a more systemic non-reversible symptoms like anemia, osteoporosis, and infertility. Untreated CD are at increased risk to develop other autoimmune diseases like type 1 diabetes and autoimmune thyroiditis in addition to having higher rate of mortality compared to the general population.

CD is a multifactorial disorder: several genetic factors combined with an environmental trigger are necessary for the disease to develop. The major, well-studied genetic risk factor for CD is the human leukocyte antigen (HLA), more specifically the HLA-DQ locus, coding for the HLA-DQ2 and HLA-DQ8 molecules. These heterodimers contribute to 35-40% of CD etiology. Other non-HLA genes have been reported to be associated to CD but they have only a modest effect.

An important benefit from the study of the genetics of human disease is to be able to predict the risk that individuals may have of succumbing to a particular disease. Genetic testing of monogenic diseases – where there is a strong correlation between risk genotype and disease – has been employed successfully in a diverse range of applications from prenatal and newborn screening, to carrier testing and medical diagnosis. With the success of genome-wide association studies and the promises of whole-genome sequencing, attention has now shifted to translating this new wave of basic genetic knowledge into personalized medicine.

In this thesis, my overall aim is to discuss the identification of genetic risk variants for CD, replicate them in new populations, and develop a risk model which can improve diagnosis of CD. In the introduction, chapter 1, I describe the history of CD, its wide spectrum of clinical features, and its current diagnosis, pathogenesis and genetic background. More than 95% of patients carry HLA-DQ2 and/or DQ8 molecules, however, 30-40% of the general population also carry these molecules. Thus HLA is necessary but not
sufficient to develop the disease. It has a sensitivity of over 96% in most populations implying that individuals without HLA-DQ2 and/or DQ8 are unlikely to develop the disease. A majority of members of the European Society of Pediatric Gastroenterology and Nutrition (ESPGHAN) have demanded modification of the current CD diagnostic criteria in order to include HLA testing as an additional screening parameter. In chapter 2, I validate a novel HLA-genotyping method in three European populations. This method was developed in a Dutch population and used six HLA-tagging single nucleotide polymorphisms (SNPs); it is suitable for high-throughput approaches.

The first genome-wide association study (GWAS) on CD and its follow-up identified 8 non-HLA loci that contribute significantly to CD risk. In chapter 3, I replicate these findings in 538 cases and 593 controls from Italy and show that CD risk loci are differently associated in different populations. For example, CCR3 and IL18RAP are associated in UK and Dutch populations, but not in Italian and Irish cohorts. Different genes may be implicated in different populations due to human migration and genetic drift. A second GWAS and a fine-mapping project identified a total of 57 non-HLA SNPs to contribute to CD development.

Advances in technology and increased knowledge on biology and genetic risk factors for common complex diseases have led to the creation of genetic risk models that can be used to target diagnostic, preventive, and therapeutic interventions based on a person’s genetic risk, or to complement existing risk models based on non-genetic factors, like family history or the presence of other diseases. As CD is a major socio-economic burden on patients, their families and society, improved diagnosis and early prevention would ease these negative effects. Chapter 4 shows how testing multiple genetic loci simultaneously, which collectively result in superior prediction of CD, might be used as a diagnostic or screening tool to prevent long-term and irreversible complications. Chapter 5 describes the genetic risk profile which I developed for CD based on HLA and non-HLA risk alleles, using cases and controls from our first GWAS. The study showed that using 10 non-HLA risk alleles can improve identification of high-risk individuals. To improve and validate the genetic risk model, I have increased the number of SNPs in the model to 26 and 57 variants and tested the model with 26 variants on two different cohorts (a nested case-control and a prospective cohorts) in chapter 6.

Genomic profiling for CD might not yet be applicable in clinical practice, but with some improvement, it might become part of the future diagnosis and treatment of CD. In chapter 7, I discuss who can benefit from this genetic profiling and how to improve the accuracy of the risk prediction. As a first step, this genetic profiling can mainly improve the diagnosis of CD in individuals at high-risk, such as first-degree relatives and individuals with other immune-mediated diseases. Maybe one day it will also be a good screening test for selecting newborns who could benefit from early intervention to prevent CD. This model can be improved by including more susceptibility variants, which can be rare, population-specific, have a parental origin effect, causative for an endo-phenotype of CD, or pathway-specific.
In conclusion, my thesis shows that risk profiling for CD may well have an application for identifying individuals at high risk for CD. Genomic profiling might lead to a future where there is personalized medicine for CD patients, where individuals could be categorized as having a low, intermediate or high risk of developing CD and could then benefit from early intervention to prevent CD or receive different treatments specific to their genetic background.

References
