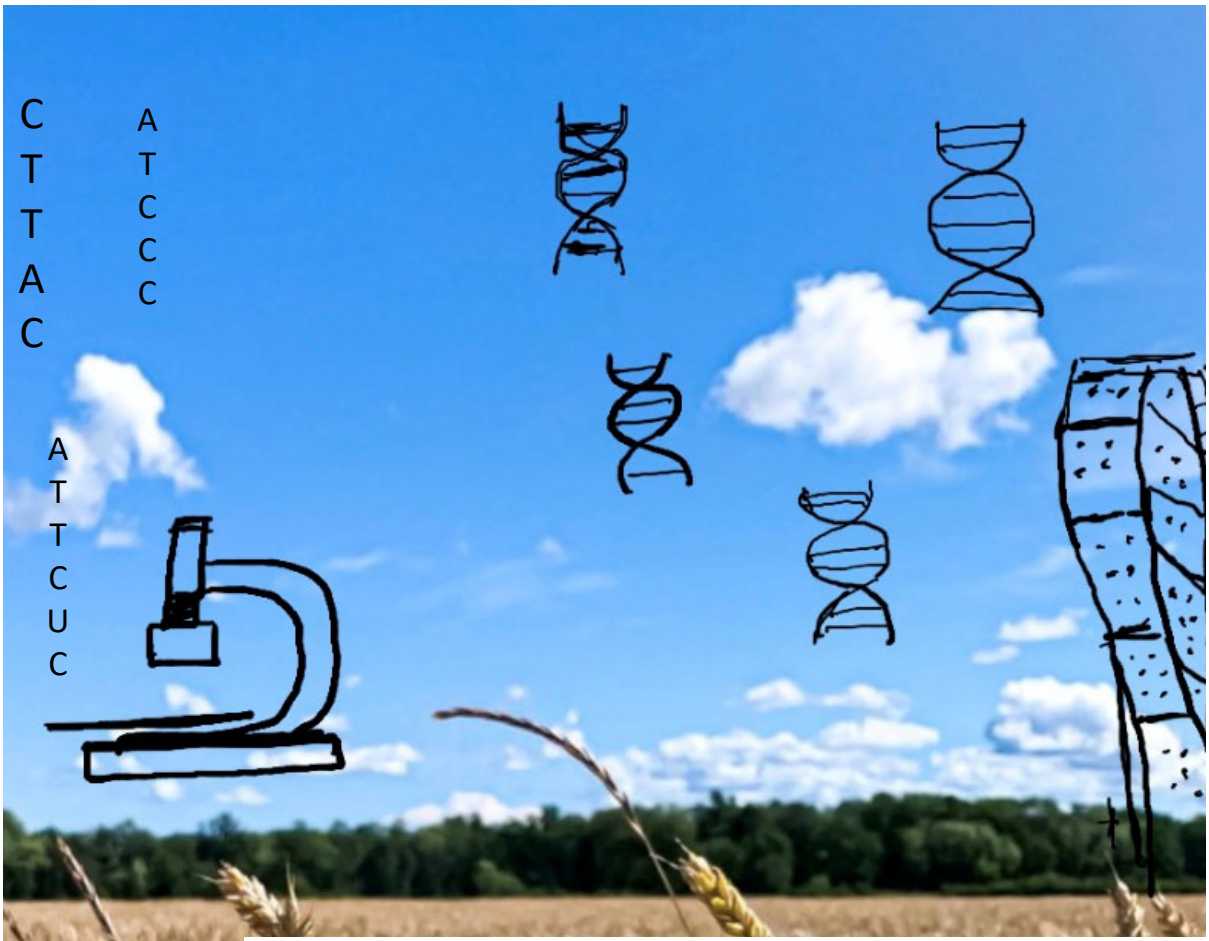
Immunogenetics of Type 1 diabetes and Celiac disease

Alshiekh, Shehab Abdulaziz

2020

Document Version:
Publisher's PDF, also known as Version of record



Immunogenetics of Type 1 diabetes and Celiac disease

SHEHAB ALSHIEKH
DEPARTMENT OF CLINICAL SCIENCES, MALMÖ | LUND UNIVERSITY

Immunogenetics of Type 1 diabetes and Celiac disease

Immunogenetics of Type 1 diabetes and Celiac disease

Shehab Alshiekh

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty Medicine at Lund University, Sweden.
To be defended publicly in Medelhavet lecture hall,
Inga Marie Nilssons gata 53, Malmö
On 25th September 2020 at 13.00 p.m.
for the degree of
Doctor of Philosophy in Medicine

Faculty opponent
Professor Riccardo Troncone
University of Naples Federico II, Naples, Italy

Organization **LUND UNIVERSITY** **Department of Clinical Sciences, Malmö** **Paediatric Endocrinology**	**Document name: Doctoral Dissertation**
	Date of issue : Septemper 25th, 2020
Author: Shehab Alshiekh	Sponsoring organization

Title and subtitle: Immunogentics of Type 1 Diabetes and Celiac disease

Abstract

The primary purpose of understanding disease etiology is to explain how a specific phenotype is determined by genotype. In pursue of this aim, exploring the diversity in DNA sequence variants that affect biomedical traits, especially those related to the onset and progression of genetically determined human disease. The human leukocyte antigens (HLA) are highly polymorphic cell surface proteins encoded in the major histocompatibility complex (MHC) region on chromosome 6. The HLA molecules are integral regulators for susceptibility to several autoimmune and inflammatory diseases, including type 1 diabetes (T1D) and celiac disease (CD), which share high-risk HLA haplotypes. Through next-generation sequencing (NGS), an integrated genotyping system of HLA loci was developed to genotype alleles of the MHC region. The full depth of allele association was used to target the novel mechanisms of HLA-associated risk alleles in T1D and CD.

The research presented in this thesis aimed to use high-resolution genotyping with NGS of HLA loci and study extended associations in patients with T1D and CD as well as in a group of patients affected by both diseases (T1D w/CD).

The main findings of importance were: -

- ***HLA-DRB3, DRB4,* and *DRB5* affect the risk of islet autoimmunity and progression to the clinical onset of T1D and should be considered when examining the role of HLA-DR genetic risk.**
- **Two distinct CD risk *DR3-DQA1*05:01-DQB*02:01* haplotypes distinguished by either HLA-*DRB3*01:01:02* and *DRB3*02:02:01* alleles in the *DRB3*01:01:02- DQA1*05:01-DQB1*02:01* extended haplotype distinguished the risk of CD, indicating that different *DRB1*03:01-DQB1*02:01* haplotypes confer different risks for CD among patients of Scandinavian background.**
- **HLA-*DRB4*01:03:01, DRB3*01:01:02*, and *DRB3*02:02:01* are associated with T1D and CD of which *DRB4*01:03:01* confers the strongest risk allele for T1D w/CD.**
- **HLA-*A*68:01:02* was identified as an additional allele positively associated between T1D w/CD and T1D.**

In conclusion, by utilizing high-resolution sequencing technologies for extended genotyping of HLA class I and II genetic determinants, the full spectrum of alleles and haplotypes variation associated with T1D and CD were explored. This basic knowledge should prove helpful contribution in building comprehensive inventories of genotype-phenotype relationships and resolving some of the HLA roles in the heritability risk for either T1D or CD, as well as in genetic models for the risk of developing both diseases.

Key word: celiac disease, type 1 diabetes, children, genetics, human leukocyte antigen, next generation sequencing		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language:English
ISSN and key title: 1652-8220		**ISBN:** 978-91-7619-963-3
Recipient's notes	Number of pages:92	Price
	Security classification	

'Signature  Date 2020-08-25

Immunogenetics of Type 1 diabetes and Celiac disease

Shehab Alshiekh

LUND
UNIVERSITY

Supervisors

Associate professor *Daniel Agardh*
Professor *Åke Lernmark*
Associate professor *Helena Elding Larsson*
2020

Faculty of Medicine
Department of Clinical Sciences, Malmö
Lund University

ISBN 978-91-7619-963-3
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

To my Family

Table of contents

List of publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals. The papers are appended at the end of the thesis.

I. Lue Ping Zhao, **Shehab Alshiekh**, Michael Zhao, Annelie Carlsson, Helena Elding Larsson, Gun Forsander, Sten A. Ivarsson, Johnny Ludvigsson, Ingrid Kockum, Claude Marcus, Martina Persson, Ulf Samuelsson, Eva Örtqvist, Chul-Woo Pyo, Wyatt C. Nelson, Daniel E. Geraghty, Åke Lernmark. Next-Generation Sequencing Reveal that HLA-DRB3, -DRB4, and -DRB5 May Be Associated with Islet Autoantibodies and Risk for Childhood Type 1 Diabetes. *Diabetes.* 2016, 65 (3) 710-718.

II. **Shehab Alshiekh**, Lue Ping Zhao, Åke Lernmark, Daniel E. Geraghty, Åsa T. Naluai and Daniel Agardh. Different DRB1*03:01-DQB1*02:01 haplotypes confer different risk for celiac disease. *HLA.* 2017; 90:95–101.

III. **Shehab Alshiekh**, Marlena Maziarz, Daniel E. Geraghty, Helena Elding Larsson and Daniel Agardh. High-resolution genotyping suggests that children with type 1 diabetes and celiac disease share three HLA class II loci in DRB3, DRB4, and DRB5 genes. Resubmitted after revision to *HLA.*

IV. **Shehab Alshiekh**, Daniel E. Geraghty and Daniel Agardh. High-resolution genotyping of HLA class I loci in children with type 1 diabetes and celiac disease. Submitted to *Pediatric Diabetes*.

Abbreviations

AB	Antibody
AGA	Anti-gliadin antibodies
BDD	Better Diabetes Diagnosis
CD	Celiac Disease
DQ2	Haplotype HLA-*DQA1*05:01-DQB1*02:01*
DQ8	Haplotype HLA-*DQA1*03:01-DQB1*03:02*
eQTL	Local Expression Quantitative Trait
LD	Linkage Disequilibrium
HLA	Human Leukocyte Antigen
HSCT	Hematopoietic Stem Cell Transplantations
IAA	Insulin Autoantibodies
IEL	Intra Epithelial Lymphocytosis
GADA	Glutamic Acid Decarboxylase Autoantibodies
GWAS	Genome-Wide Associations Studies
GFD	Gluten-Free Diet
MHC	Major Histocompatibility Complex
OR	Odds ratio
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RPE	Relative predispositional effects
RR	Relative Ratio
SSOP	Sequence-Specific Oligonucleotide Probe Hybridization
SSP	Sequence-Specific Priming
SBT	Sequence-Based Typing
SNPs	Single Nucleotide Polymorphisms
TCR	T cell receptor
TEDDY	The Environmental Determinants of Diabetes in the Young
TG2	transglutaminase 2
T1DGC	Type 1 Diabetes Genetics Consortium
TID	Type 1 diabetes
T1D w/CD	Type 1 diabetes with Celiac disease
tTG	tissue Transglutaminase
UTR	Untranslated Region
ZnT8A	Zinc Transporter 8 Autoantibodies

Abstract

The primary purpose of understanding disease etiology is to explain how a specific phenotype is determined by genotype. In pursue of this aim, exploring the diversity in DNA sequence variants that affect biomedical traits, especially those related to the onset and progression of genetically determined human disease. The human leukocyte antigens (HLA) are highly polymorphic cell surface proteins encoded in the major histocompatibility complex (MHC) region on chromosome 6. The HLA molecules are integral regulators for susceptibility to several autoimmune and inflammatory diseases, including type 1 diabetes (T1D) and celiac disease (CD), which share high-risk HLA haplotypes. Through next-generation sequencing (NGS), an integrated genotyping system of HLA loci was developed to genotype alleles of the MHC region. The full depth of allele association was used to target the novel mechanisms of HLA-associated risk alleles in T1D and CD.

The research presented in this thesis aimed to use high-resolution genotyping with NGS of HLA loci and study extended associations in patients with T1D and CD as well as in a group of patients affected by both diseases (T1D w/CD).

The main findings of importance were: -

- *HLA-DRB3*, *DRB4*, and *DRB5* affect the risk of islet autoimmunity and progression to the clinical onset of T1D and should be considered when examining the role of HLA-DR genetic risk.
- Two distinct CD risk *DR3-DQA1*05:01-DQB*02:01* haplotypes distinguished by either *HLA-DRB3*01:01:02* and DRB3*02:02:01 alleles in the *DRB3*01:01:02- DQA1*05:01-DQB1*02:01* extended haplotype distinguished the risk of CD, indicating that different *DRB1*03:01-DQB1*02:01* haplotypes confer different risks for CD among patients of Scandinavian background.
- *HLA-DRB4*01:03:01, DRB3*01:01:02*, and *DRB3*02:02:01* are associated with T1D and CD of which *DRB4*01:03:01* confers the strongest risk allele for T1D w/CD.
- *HLA-A*68:01:02* was identified as an additional allele positively associated between T1D w/CD and T1D.

In conclusion, by utilizing high-resolution sequencing technologies for extended genotyping of HLA class I and II genetic determinants, the full spectrum of alleles and haplotypes variation associated with T1D and CD were explored. This basic knowledge should prove helpful contribution in building comprehensive inventories of genotype-phenotype relationships and resolving some of the HLA roles in the heritability risk for either T1D or CD, as well as in genetic models for the risk of developing both diseases.

Introduction

Type 1 diabetes (T1D), known also as autoimmune diabetes, is a chronic endocrine disease characterized by a pancreatic β-cells loss that leads to insulin deficiency resulting in hyperglycemia complications (1-3). Celiac disease (CD) is a chronic immune-mediated enteropathy induced by ingested dietary gluten, which is present in grains, including wheat, rye, and barley (4, 5). The susceptibility to T1D and CD, in which the immune system plays a significant role, has a strong genetic background, as demonstrated by diseases concordance in monozygotic twin pairs (6, 7). The rate of familial disease clustering in T1D patients is 6% in siblings compared to 0.4% clustering rate in the general population, and 50% concordance rate in monozygotic twins (6, 8). The average pooled prevalence of CD among first-degree relatives (FDR) exceeds that of the general population by 7.5%–15% (9, 10) and concordance around 80% in monozygotic twins and less than 20% among dizygotic twins (7, 11). Genetics is the cornerstone of disease etiology in CD (12) and T1D(13), of which specific human leukocyte antigen (HLA) genes are located in the major histocompatibility complex (MHC) on the short arm of chromosome 6 (6p21.3) confer the strongest risk association. This introduction summarizes the current state of knowledge of HLA susceptibility in T1D and CD.

History of HLA-associated diseases

The discovery of the association between the *HLA-B* gene and Hodgkin lymphoma was the first link between HLA genes and human disease (14). The association between HLA risk genes and autoimmune diseases was first described more than 50 years ago (15). HLA-*B*27* (called HL-Aw27 previously) link with ankylosing spondylitis was the first reported discovery between HLA risk and autoimmune diseases (16). The association of HLA with T1D was first reported in the 1970s (17, 18). Meanwhile, the association between the HLA region and CD development was also discovered (19-21).

Similarly, HLA association was also observed for several other autoimmune and inflammatory diseases, including systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and multiple sclerosis (MS) as well infectious disease (for example, *DRB1*1302* is protective against chronic hepatitis B virus (22)) and even neuropsychiatric disorders underscoring the central importance of HLA risk to

physiology, protective immunity and deleterious, disease-causing autoimmune reactivity (7). To date, genes in the HLA region are still the primary genetic risk determinants for T1D and CD, accounting for the major share of the genetic susceptibility to diseases, estimated from studies of affected sibling pairs (23). However, even with a large body of defined genetic associations between HLA and many autoimmune diseases have been identified, most genetic risk associations in the HLA complex remains unexplored.

MHC region gene structures

The HLA region encoding for the MHC antigen-presenting receptor molecules is generally organized into three subclasses **(Fig.1, left**):

- HLA class I region, including the classical, highly polymorphic *HLA-A, HLA-B,* and *HLA-C* heavy chain paralogues, besides the non-classical *HLA-E, HLA-F,* and *HLA-G* genes.
- HLA class II region, including *HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB2, HLA-DRB3, HLA-DRB4,* and *HLA-DRB5* as well as fewer variable genes.
- Finally, the HLA class III region, which contains genes implicated in inflammatory responses, leukocyte maturation, and the complement cascade.

  Classical MHC molecules are cell-surface membrane glycoproteins that function in serological specificity by binding and presenting protein-peptide antigens recognized by the T cell receptor (TCR). MHC molecules are generally categorized into two principal classes, with three main kinds of antigens in each class. MHC class I molecules A, B, and C comprise a protein polypeptide chain that forms a heterodimer shape consisting of a heavy chain adjunct with a relatively invariant β-2 microglobulin protein. Class I molecules act a principal function by presenting peptides processed by the endoplasmic reticulum to the immune cells. They are expressed in nearly all nucleated cells. The *HLA-DR, -DQ* and *-DP* genes encode MHC class II molecules *and* consist of a heterodimer consisted of two polypeptides: an α and a β chain (alpha and beta), both anchored in the cell membrane of antigen-presenting cells (B lymphocytes, dendritic cells, macrophages). The overall structure of MHC molecules is relatively similar, as that is illustrated schematically in (**Fig.1, right**).

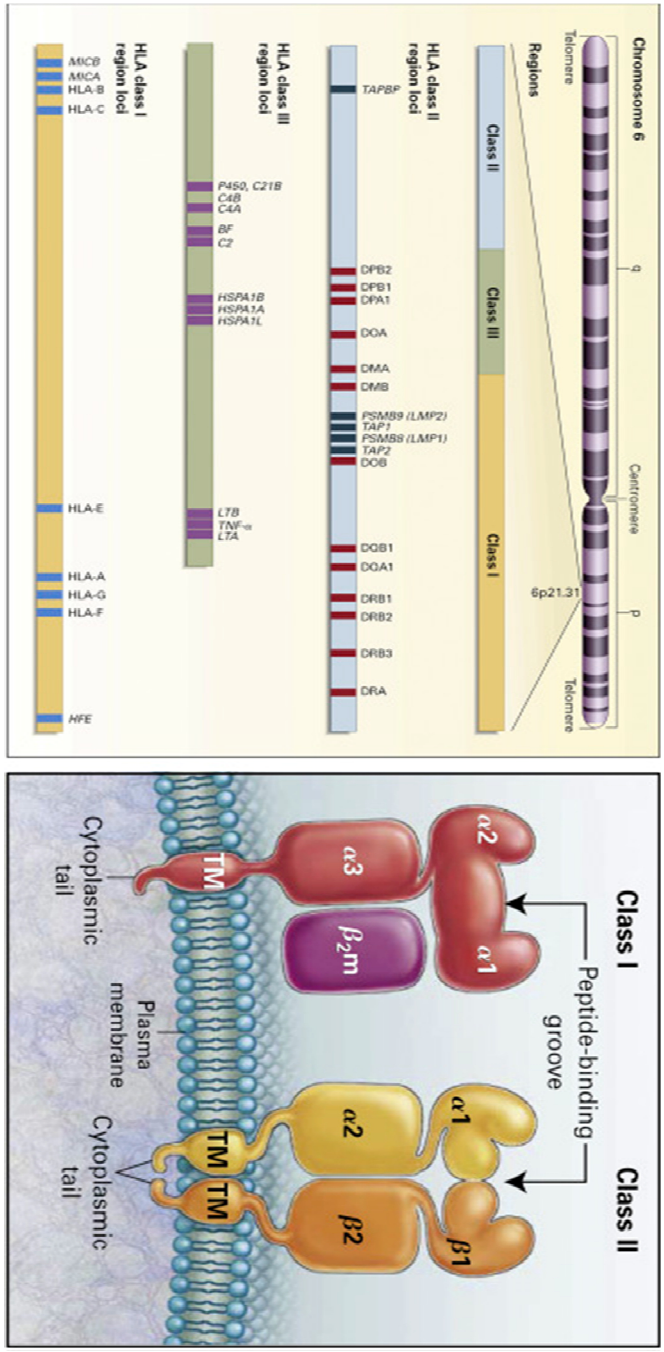


Figure 1. A simplified location organization of the HLA complex on chromosome 6 (left figure). A schematic structure of MHC class I and class II molecules (right fgure). (Adapted from Klein, J., & Sato, A. (2000). The HLA system. N Engl J Med, 343(10), 702–709)

The initial structural explanation of the *HLA-A2* protein structure in 1987 showed that the extracellular portion of the protein transform by folding into a β-pleated sheet on which two α-helices create a groove or a channel into which peptides bind (24, 25). The shape and charges within the peptide-binding groove dictate the peptides' repertoire that can bind to a specific HLA antigen. The TCR identify the combination of HLA protein and antigen peptide, creating what is known as the "tri-molecular complex" that initiates the immune response, as shown in (**Fig. 2**).

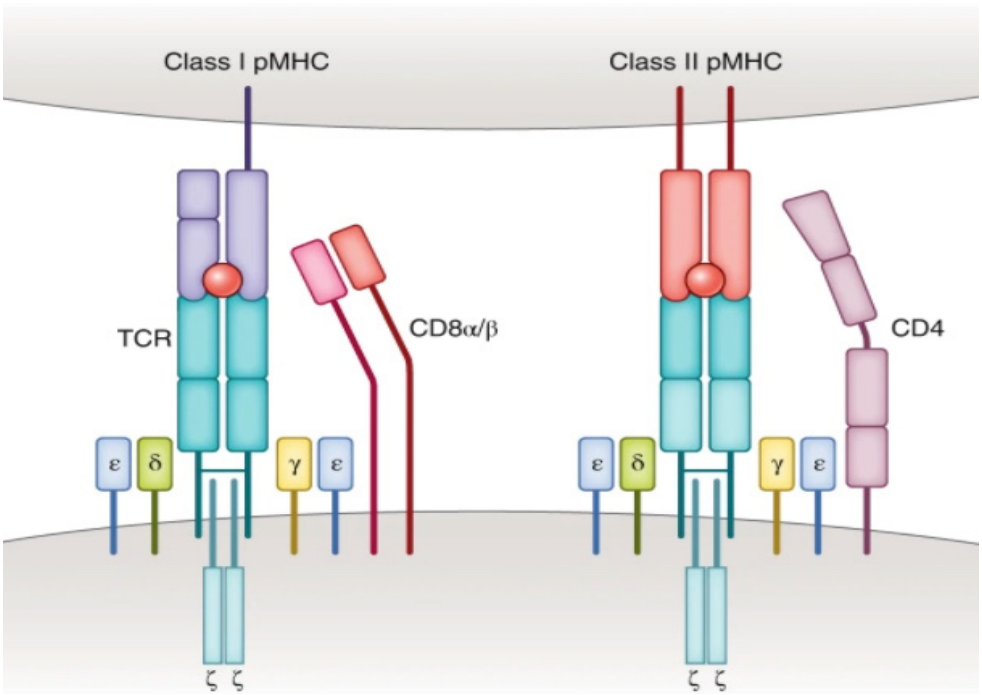


Figure 2. Molecular basis of TCR–MHC recognition.
The classical TCRαβ complex formed of cell surface receptors CD3ζ, CD3γ, CD3δ and CD3ε chains. The MHC class I molecule consists of an α chain and a β2-microglobulin chain. The MHC class II molecule is composed of two non-covalently associated α and β chains. (Adapted fromJoglekar, A.V., Li, G. T cell antigen discovery. Nat Methods (2020)).

HLA gene mappings

The HLA genomic loci map is located on chromosome 6 (6p21) short arm (26) (**Fig. 3**). The order of the exons and introns in HLA-encoding genes is analogous, in which the signal peptide of the surface receptor is encoded in exon 1 containing 5′ untranslated sequence (UTR), and immunoglobulin-like domain of MHC extracellular portion of the MHC molecule is encoded by exons 2 and 3 (exon 4 for class I genes).

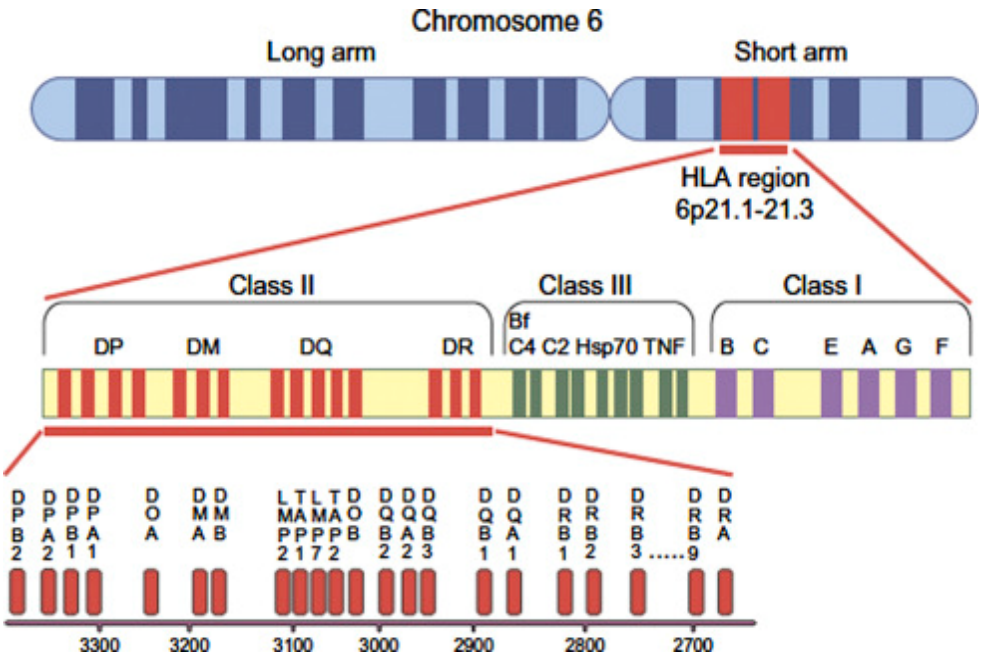


Fig.3 The HLA region location on the short arm of chromosome 6.
The HLA class II genes at the *DRB3*, *DRB4,* and *DRB5* along with *DRB1** as well as *DQA1*-B1** are tightly organized and LD strongly affecting risk associations in HLA-linked diseases.(Adapted from Autoimmune (type 1) diabetes. / Lindbladh, Ida; Svärd, Agnes Andersson; Lernmark, Åke.The Autoimmune Diseases. ed. / Noel R. Rose; Ian R. Mackay. 6. ed. Academic Press, 2019. p. 769-787) .

Exons 2 and 3 in HLA class I genes (e.g., HLA-B) and exon 2 from each of the HLA class II two genes encoding an HLA class II antigen molecule (e.g., HLA-DQA1 and DQB1), define the cell surface receptor peptide-binding groove and considered the essential determinants of antigen specificity. HLA genes exons encoding the peptide-binding groove are the most polymorphic sites in the HLA region. The reason for the polymorphism in these exons is speculated to result from natural selection. Old conventional HLA genotyping methods minimally extends the genotyping of those core exons in comparison to new methods. A single HLA class I gene (*HLA-A, B,* and *C*, respectively) encode for each HLA class I (A, B, and C) protein. Two different genes are responsible for creating the heterodimer molecule of HLA class II antigens. Accordingly, the translation of two different genes, e.g., in the DQ antigen, *HLA-DQA1* encodes the α chain, and *HLA-DQB1* encodes the β chain; in the DP antigen, *HLA-DPA1* encodes the α chain, and *HLA-DPB1* encodes the β chain, and in the DR antigen, *HLA-DRA1* encodes the α chain, and *HLA-DRB1* encodes the β chain. DRβ chains can be encoded by additional HLA loci found on different chromosomes. *HLA-DRB1* is located on all copies of chromosome 6.

Some DRB genes are regarded as pseudogenes, but not for *HLA-DRB3, DRB4*, and *DRB5* as they all produce a functional DRβ chain that can dimerize with the

translated protein of the *DRA1* gene. The HLA-*DRB3, DRB4*, and *DRB5* genes are commonly designated as a single *DRB345* allele because they segregate as a single unit locus, but they are separate loci and not different alleles of the same locus. Because of the pattern in which HLA haplotypes evolved, a given class II haplotype can only have zero or one of these secondary, expressed *DRB* alleles, which is highly dependent on the identity of the *DRB1* allele on that haplotype. Thus, the data for these loci are often assigned to three different alleles of a single locus as *DRB345*.

The MHC locus exhibits two unique features: extreme polymorphism and strong linkage disequilibrium within a range possibility. These peculiar characteristics have made the localization of the specific genes and alleles responsible for disease association signals in the region difficult. Genes on a given chromosome are said to be linked if alleles at respective genes show a non-random pattern of association between alleles at different loci within a population; those alleles are considered to be in linkage disequilibrium (LD) (27). Genetic analysis of HLA loci mapped the *HLA-C* and *B* genes to be situated within a 90-kb region at chromosome 6p21.33 (28). Allele combinations of these specific two genes are often preserved because of LD, likely be derived from a shared ancestral chromosome segment. The LD between *HLA-B* and *C* is initially called the *HLA-B~C* haplotype block (29). As in the *HLA-B~C* block, HLA class II genes *HLA-DRB3, DRB4, and DRB5*, *DRB1, DQA1,* and *DQB1* genes within the HLA class II region are located in a 150 - 210-kb range at chromosome 6p21.32 (28). Consequently, alleles of these genes are also in strong LD and constitute what has denominated as the *HLA-DR~DQ* block (30).

HLA gene polymorphism and nomenclature

HLA genes are the most polymorphic genome identified in the human genome(26, 31) and are associated with the highest number of human diseases in the genome(32). In comparison to most human genes that have only one form or have a few variant sequences, the classical HLA genes can have thousands of variable alleles, as seen in their extreme polymorphism (**Table 1**). There are currently more than 27,980 HLA alleles described by the HLA nomenclature and listed in the IPD-IMGT/HLA published database (Release 3.41.0, 2020-07-13) (33, 34). As the updated database shows, most polymorphism for MHC class I is seen in the *HLA-B*, in contrast to the *HLA-A* and *HLA-C* genes. This variable pattern of polymorphism also manifested in MHC class II *DR* encoding genes, where the *DRB1* gene is hugely polymorphic, while the *DRA1* gene is considerably invariant. The heterogeneity of the DR antigens on the cell surface is increased in haplotypes by pairing a second *DRB* with the *DRA1* gene. The polymorphism in HLA class II DQ genes encoding the two antigen polypeptides results in differentiating the possible types of cell-surface DQ receptor molecules to four by combinational heterogeneity. For DP gene, almost all of the variability is encoded in the *DPB1*

gene compared to *DPA1* gene limited polymorphism, and generally, only a small number of *DPA1* alleles are seen in each population.

Table 1. Classical HLA loci extreme polymorphism.
Number of identified alleles for each classical HLA gene as it published on IPD-IMGT/HLA database(2020 release).

HLA Locus		Number of identified alleles
HLA class I loci	*A*	6,192
	B	7,431
	C	6,067
HLA class II loci	*DRA*	29
	DRB1	2,737
	DRB3	345
	DRB4	166
	DRB5	130
	DQA1	260
	DQB1	1,857
	DPA1	202
	DPB1	1,584

Frequent updating in HLA nomenclature used was required because of the rapid surge in the number of identified HLA alleles. The current nomenclature was first designated in April 2010 with frequent updates, as illustrated in **Fig. 4**. The nomenclature system uses numeric fields divided by colons to represent four levels of resolution (33). Briefly, the 1st field (formerly 2-digit level) is the serological definition of related allele groups. The 2nd field (4-digit) differ in their unique HLA protein sequence. Finally, the 3rd and 4th fields describe alleles harbouring synonymous silent polymorphisms, including exonic coding and non-coding variations.

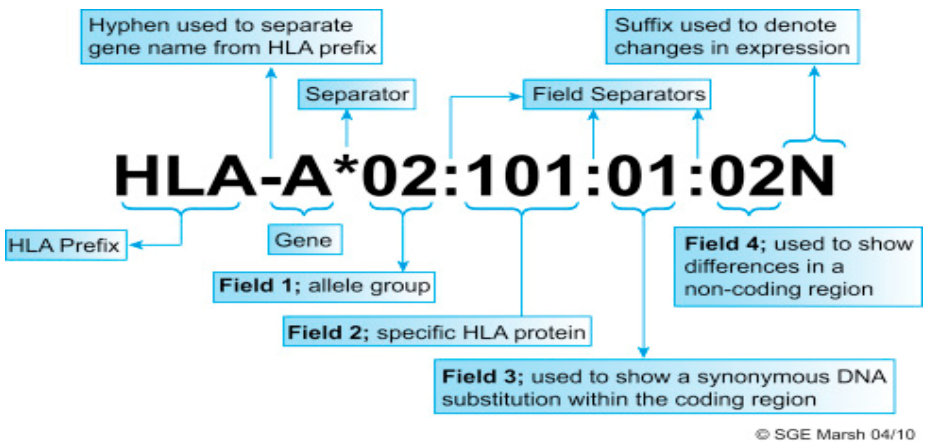


Figure 4. Human leukocyte Antigen (HLA) nomenclature
(From Nunes, E., Heslop, H., Fernandez-Vina, M., et al. (2011). Definitions of histocompatibility typing terms. *Blood, 118*, e180–e183)

Early *DPB1* alleles nomenclature numbered the alleles in order of their discovery, regardless of serologic reactivity, because few serologic reagents were developed before the availability of DNA-based genotyping. With more genotyping of more *DPB1* alleles, the nomenclature system was required to add the colon field separators in the later update (35).

Data interpretation can be challenging to interpret among studies performed with different genotyping methods due to many identified alleles. This reason impedes the ability to produce allele-level genotype recognition with all the highest-resolution sequencing. For instance, high-resolution genotyping can distinguish specific allele designations on the fourth field, including sequencing of intronic and untranslated sequences. Fortunately, the relevant evidence for function resides is shown in the first two serology fields (four-digits), which define the amino acid sequence of the encoded HLA protein for most disease genetic association studies. Generally, the silent polymorphisms and polymorphisms in intronic and untranslated regions do not affect the function, and the first two-field resolution is sufficient.

HLA genotyping progression

HLA genotyping technology is developing at an accelerated pace. Measurement of sensitization to tissue histocompatibility differences began early in the 1960s. It was first noticed when patients receiving several transfusions of ABO compatible blood experienced an unexpected immune reaction where antibodies were being produced against donors' white blood cells and that these antibodies triggered another immune reaction in half the samples of other white blood cells donors. Initially called the factor responsible for the immune reaction the MAC, then was recognized as the first of a group of human leukocyte antigens, or HLAs. The first histocompatibility testing was conducted by applying cell-based techniques using multiparous women serum for fetus blood group testing. Extensive standardization among the different HLA typing laboratories of testing reagents and sequencing protocols was justified, and HLA genotyping assays evolution resulted in low-resolution sequencing with many spaces. The use of DNA-based genotyping technology during the 1980s enhanced the sequence reading and the determination of the individual HLA alleles present in study subjects which resulted in an increased number of discovered alleles that led to the first complete sequence-based gene map that analysed 224 gene loci and studied 128 (57%) loci expression in 1999(36) and continuously going on. Initial DNA technology applied restriction fragment length polymorphism (RFLP) using restriction enzymes and Southern blotting (37). Polymerase Chain Reaction (PCR) techniques allowed the development of numerous PCR-based genotyping methods, including the use of sequence-specific oligonucleotide probe hybridization (SSOP) in analysing

amplified β-globin and HLA-DQ (38), typing of HLA-DR by sequence-specific priming (SSP) (39), and sequence-based typing (SBT)of HLA-class II (40).

The DNA sequencing technology that has revolutionized the genomics research currently is the next-generation sequencing (NGS)-based method on the Illumina platform, in which DNA is fragmented and quickly sequenced to produce short sequencing reads with low-cost. A platform that leverages a sequence-by-synthesis approach to arrange the order of nucleotides in a DNA strand(41). Illumina's DNA sequencing technology constructs highly accurate (higher than 99.9%) sequencing reads, which are reasonable to produce on a massive scale. These advantages have prompted the ascent of the Illumina platform to become the current gold standard of clinical and research sequencing that led to innumerable scientific discoveries over the past decade that have enhanced our understanding of evolution, adaptation and disease pathogenesis through the discovery of pathogenic variants (42). The specificities of each of the most common sequencing platforms selected according to their strengths and weaknesses for HLA sequencing for a given study, depends on their intended use. For instance, a simple clinical test can be more cost-effective than the full sequencing of *HLA-B*27* to diagnose ankylosing spondylitis disease. However, typing and eventually matching donors and recipients in solid organ and hematopoietic stem cell transplantations (HSCT) are particularly genotype-specific, underlining the importance of DNA genotyping for at least two-field resolution of HLA genes for organ-rejections operations decisions in case of organs transplantation surgeries. **Table 2** illustrates the resolution levels of various HLA genotyping methods.

Table 2.
Resolution levels of HLA genotyping methods shown in increasing level of genotyping resolution (35).

Resolution Level	Genotyping Method	Sequencing Fields
Low	SNP imputation	Varies
Low	Serology	1
Low	SSP	Up to 3
Low	SSO	3
Low	SBT	3
High	NGS (exon-based)	3
High	NGS (with introns and UTR)	4

Next-generation sequencing (NGS) is currently applied for HLA genotyping (43-45). HLA genotyping using (NGS) is becoming a successful approach in research and clinical laboratories. Recent advances in immunogenetics genotyping unlock the gateway to implementing HLA typing in risk stratification of diseases in the clinic (46) in comparison to the costly slow conventional methods such as Sanger sequencing or Sequence-Specific Oligonucleotides probe hybridization (SSOP)/Sequence-Specific Priming (SSP) (47, 48). Besides that, NGS has allowed quintessential insights in "single-molecule sequencing" of the growing number of

polymorphic HLA alleles by typing all exons and introns through combining clonal (single-molecule) sequencing and a high level of parallelism for a potential full understanding of regulation and expression of these genes(49). All DNA sequencing innovations have raised the level of resolution of the genotyped data as shown in generational progression (**Fig. 6**).

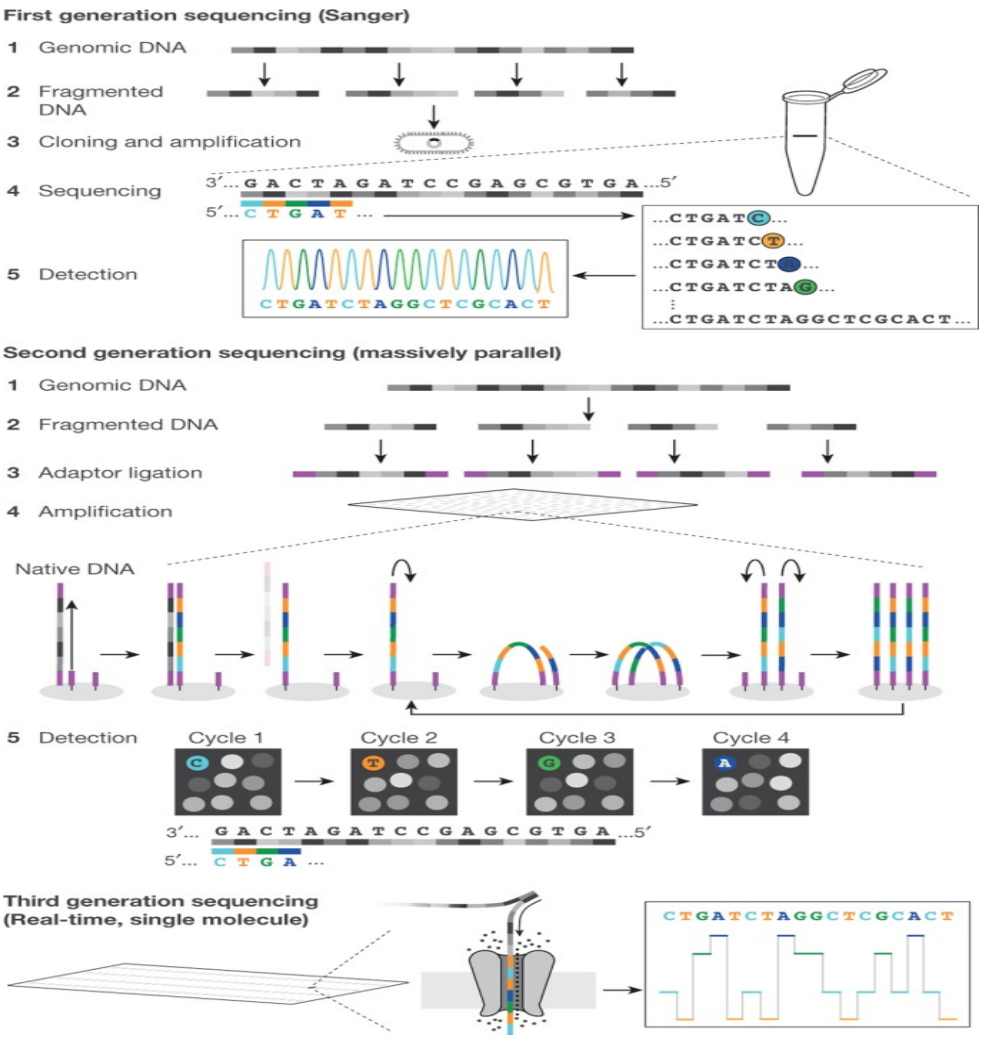


Figure 6. Different DNA sequencing technologies evolution.
Schematic illustrations of first, second and third generation DNA sequencing. Second generation sequencing is also referred as next-generation sequencing (NGS) (adapted from J Shendure et al. Nature 1–9 (2017))

NGS systems generate substantial numbers of "clonal" sequence reads derived from subject DNA molecules, in a massively parallel fashion. The key advantages of NGS clonal kind enable each sequence read assigned to a single allele, resulting in an in-depth analysis of HLA types by phasing linked polymorphisms and resolve cis-trans ambiguities more than those obtained from more popularly used Sanger-sequencing based typing (SBT) methods (50, 51). NGS platforms can return full-length (four-field) alleles and detect novel alleles by generating at once many more sequences reads than SBT instruments, yielding non-core exons, introns, and untranslated regions to be sequenced in addition to core exons (52). By applying next-generation sequencing (NGS), an integrated genotyping system of HLA genes exons 1–4 was developed to extend genotyping of all *DRB1, DRB3, DRB4,* and *DRB5* alleles (53, 54).

The advantage of performing full haplotypes phasing – instead of individual genes brings important critical information to overcome the complexity inflicted by the unique LD pattern and genetic properties of the HLA region. With a facilitated imputation of HLA alleles, as compared to GWAS data, NGS performs more functionally relevant association studies at the amino acid level (55). Consequently, the causative variants/loci of many HLA-linked diseases will be discovered by complete sequencing approaches that are accessible by NGS.

The HLA complex is a frequent hit in genome-wide association studies (GWAS) and has been linked with many immune-related diseases than any other region of the human genome (56). The sequencing resolution up to allele-level will investigate the influences of individual allele, haplotype, or genotype on disease association studies. Low-resolution HLA genotyping generates short sequences with high amounts of data points in each class, providing noteworthy statistical significance. However, an insufficient number of data points for each class, especially when analyzing genotype associations, can result in low statistical significance conclusions to reveal practical disease association effects. Alleles with varying effects size on disease susceptibility can conceal the precise effects of individual alleles. Study design should include a method that presents an adequate resolution to test the scientific hypotheses with an effect size ranking statistical method. The cost is an important part in determining the use of the genotyping method as higher resolution charge higher costs. HLA genotype resolution studies frequently vary among published literature, and the analysis of comparisons among genetic studies should consider that in mind.

Through NGS, the researchers can obtain high-resolution typing data (all four fields) and thereby improves the full characterization of population HLA diversity, evolution, and demographics by several means. First, it significantly extends the genetic difference between populations. For this purpose, information about synonymous substitutions is of particular interest. These variants are prone to neutral evolution and maybe illuminating for the identification of demographic events in a specific population at an increased risk for a disease like Scandinavian

ethnicities with known increased risk for T1D and CD. Second, although a deep part of the HLA molecular distinction is concentrated in the exons encoding the peptide-binding clefts, polymorphic sites are distributed along with the 4 Mb of the HLA locus, many variants located in non-coding portions of the HLA region may play a role in regulatory functions. Also, the neutral polymorphic sites in introns or in 3′ and 5′ UTR of HLA genes are beneficial for identifying regions with the lowest deviation from neutrality caused by the strong LD effect observed within the HLA.

Third, large sizeable NGS-based population sequencing projects help recognize rare genetic variants that frequently become prevalent and contribute substantially to the overall allele database. The discovery of rare alleles shared by populations located in geographically distant regions may shed light on common origins. In Finland, the narrower gene pool showing HLA haplotype frequencies is well characterized by the Finnish population compared to European HLA alleles. The imputation success for *HLA-DRB1* and *B* alleles was very low, indicating the importance of population-specific reference database(57). Although the existing GWAS data describing HLA genotypes are appropriate in disease associations, it is necessary to observe that these tag single nucleotide polymorphisms (SNPs) are merely substitutes for given HLA alleles. Fourth, compared to genome-wide SNP data, NGS reads reaching all types of variations over the whole locus, significantly improving haplotype phasing and, consequently, enhancing LD estimation to assess populations' ancestry, as this hallmark feature decrease expectedly with time through recombination.

The impact of NGS on genetic analysis in autoimmune disease studied is vast and can be compiled into two main aspects. First, NGS can target extended, in-depth sequencing of the HLA, particularly class II alleles *DRB1, DRB3, DRB4, DRB5, DQA1, and DQB1*. Second, HLA haplotypes analysis and molecular dissection of disease associations or identification of HLA-linked causative variants (58). The HLA community's current efforts are directed toward the standardization of data reporting in HLA studies to allow comparisons among genetic predisposition studies (59). In the near prospect, the time may come when whole-genome sequencing becomes a standard measure for every individual through applying HLA genotyping methods and nomenclature for accurate interpretation of disease association studies.

Type 1 Diabetes (T1D)

Type 1 diabetes (T1D) is a chronic autoimmune disease characterized by increased blood glucose levels (hyperglycemia) due to the insulin deficiency that occurs as the consequence of pancreatic islet β-cells destruction (60). Since the early classification T1D issued in 1955, it is still one of the most common endocrine and metabolic conditions affecting children. The majority of patients have a loss of β-cells concomitant with the formation of associated autoantibodies; herein is considered as autoimmune T1D. Only in a minority of patients, are autoantibodies not detected, and the cause of β-cell destruction is idiopathic.

HLA gene associations with T1D

T1D is a polygenic disease that is determined by both genetic and environmental factors. Genetic risk factors are required but not adequate for disease development. Candidate-gene studies in the early 1970s reported the first HLA association for T1D in HL-A antigen called previously "specificity W15" (18), assigned by new nomenclature to *HLA-DRB1*04:01* risk allele. Afterward, Nerup et al. proved a subsequent association with W15 and what was then called "*HL-A8*" (17), which was named later as the *HLA-B*08:01* HLA class I gene located on the conserved T1D risk "*A1-B8-DR3*" haplotype (61). Subsequent hereditary family-based studies replicated and confirmed these alleles risk associations (62). Studies in the 1980s reported the high-risk heterozygous genotype (commonly referred to as DR3/4)with a haplotype including *DRB1*03* allele on one chromosome and *DRB1*04* allele on the opposite by linkage analysis in families with T1D (63). Many studies have since confirmed that *DRB1*04* as the major candidate-gene in T1D in different ethnic populations (64-69).

HLA is considered the major susceptibility locus in T1D with a high likelihood ratio (OR of at least 6 in most studies), and accounts for approximately 40% of its heritability (70). The molecular structure of the HLA-encoded risk is principally amino acid variants in the MHC class II antigen-binding grooves that determine the repertory of bound antigen peptides and regulate T cell responses determining self-tolerance and immune system activation. In T1D, the individual and additive effects of specific amino acid variants in *HLA-DR* and *DQ* have been subjected to extensive studies (71, 72). Extensive international T1D genetics collaborative studies like the

Type 1 Diabetes Genetics Consortium (T1DGC), worked to define genetic determinants in T1D and provide resources to identify all of TID genes susceptibility(73). Through these efforts, the linkage to *HLA-DR3* and *DR4* established the vast majority of the genetic component of this severe life-long disease (74, 75). Later on, the association with the HLA-DQ haplotypes that include *HLA-DQ2* and *HLA-DQ8* for *HLA-DR3* and *HLA-DR4* defined the risk associated with these two HLA-DR risk haplotypes. Extreme genetic risk studies showed that the relative risk gradient is higher for *HLA-DQ2* and *HLA-DQ8* homozygotes than for heterozygotes and maximal for *HLA-DQ2/HLA-DQ8* heterozygosity (76).

DRB1-DQA1-DQB1 haplotype T1D risk associations

Most definite HLA-associated T1D risk is thus confirmed to the *HLA-DR* and *DQ* encoding loci (77, 78). Because genes in the HLA region exhibit extensive LD, the presence of one allele allows the estimation of other alleles on the same haplotype. Genes encoding DR and DQ express this feature, so that, for a presented population, a single allele is regularly observed in only one or a few haplotype sequences given the known distance between them and expected recombination frequencies. This concept is widely studied in disease association studies. As an excellent example, the *DRB1* gene is present in all individuals. Allelic variants of *DRB1* are in LD with either none or one of class II *DRB3*, *DRB4*, and *DRB5* genes and that make it target of genotyping association (79). Additionally, there are several related genes that can be estimated by the LD concept, like the pseudogenes (*DRB2*, *DRB6*, *DRB7*, *DRB8*, and *DRB9*).

To explain DRB1 combination with other alleles, the *DRB1*03:01* is frequently found linked to the *DQA1*05:01* and *DQB1*02:01* alleles almost exclusively, to form the high-risk haplotype *DRB1*03:01-DQA1*05:01-DQB1*02:01*, commonly called "DR3". Similarity, the *DRB1*04:01* is frequently found linked with *DQA1*03:01* alleles, but it can differ in their linkage with either *DQB1*03:01* or *DQB1*03:02* to form the high-risk haplotype *DRB1*04:01-DQA*03:01-DQB1*03:01/ DQB1*03:02*, referred as "DR4" haplotypes. However, the T1D risk for these two haplotypes is very different and that not only due to DQ-encoding genes driving risk for T1D or that the DRB1 locus is of limited relevance. T1D haplotype risk contribution comparison between *DRB1*04:01-DQA1*03:01-DQB1*03:02* and *DRB1*04:03-DQA1*03:01-DQB1*03:02* haplotypes could point to DRB1 locus as the chief determinant of T1D susceptibility, and the DQB1 locus is not. However, it is the total allele risk effects that most likely explain the combination of DR and DQ presence effect as HLA-DR-DQ accounts for ~40% to 50% of the overall T1D genetic risk (23, 80). Therefore, evaluating a single locus for T1D susceptibility can be misleading, as analysing *DRB1-DQA1-DQB1* haplotypes is considerably more appropriate.

GWAS, through recent landmark T1D genetic studies (77, 78, 81), not only have proved that single nucleotide polymorphisms (SNPs) as the most influential association in the *DR-DQ* sequence, but also reported less association of additional non-*HLA* loci (66, 82-84). Numerous studies attempted to dissect the relative risk confirmed by class II *DR* and *DQ* closely located loci. They extended it to the *HLA-DP* locus as well (85). The findings of *DRB1*03:01* allele association with the *DRB3*02:02* allele to express an independent risk for T1D compared with the *DRB3*01:01* allele, in which the high-risk *DRB1-DRB3* haplotypes containing the genetic marker of *DRB3*02:02* allele but also enhanced the predisposition for *DRB1*03:01* haplotypes particularly in individuals homozygous for *DRB1*03:01,* proved the benefits of extended sequencing to *DRB* alleles (79). Specifically, on the *HLA-DQ8* haplotype, *HLA-DRB1*04:01* and *HLA-DRB1*04:05* are associated with greater susceptibility to T1D than is *HLA-DRB1*04:04*, whereas *HLA-DRB1*04:03* is shown to be protective. Several recent studies have attempted to distinguish the relative risk between the adjacent *HLA-DR* and *DQ* class II loci (80) and that extended to *HLA-DP* locus (85). Early T1D investigations showed that *HLA-DQ* might have more significant roles in disease susceptibility than *HLA-DR* genes (86).

It was previously reported that *HLA-DR-DQ* genotypes might define various T1D autoimmunity response showing strong differential associations with insulin autoantibodies including insulin autoantibody (IAA), GAD 65 (GAD antibody), IA-2 antigen (IA-2A), or the three variants (amino acids R, W, or Q on position 325) of ZnT8A antibodies (ZnT8RA, ZnT8WA, ZnT8QA, respectively, that can help develop a diagnostic tool of T1D (87, 88). In these studies, the limited analyses studied just *HLA-DQ* alleles detected by allele-specific probes (89, 90). Another major limitation of the referred study was the lack of typing *DRB3*, *DRB4*, and *DRB5* alleles located between the *DRA* and *DRB1* loci (91, 92). These haplotypes are often associated with insulin autoantibodies, and further studies can apply NGS of all *DRB* alleles in successively diagnosed T1D patients with all islet autoantibodies analyzed (93, 94) and in geographically matched control subjects(95). In another study, the extended haplotype *HLA-DRB1*03:01-DQ2 (HLA-DQA1*05:01-DQB1*02:01)* was associated with GAD65 autoantibody (96). The study of patients with recently diagnosed T1D (1–34 years of age) shows that the age-dependent onset of T1D, which is associated with distinct *HLA-DR-DQ* genotypes, is considerably related to the appearance of β-cell-targeting autoantibodies. These genetic risk determinants are common in western populations and have low penetration (97, 98), which might explain why many people do not develop islet-targeted autoimmunity or T1D despite having these risk factors.

Previous T1D risk studies showed that risk depends additionally on the genotypic context. The total risk of the "DR3/DR4" heterozygous genotype (where either of the protective *DRB1*04:03* or *DQB1*03:01* alleles are not present in DR4 haplotype) is higher than the total combined risk for the individual DR3 and DR4

homozygous haplotypes (the *DR3/DR4* genotype reported having an OR of 16.6) as shown in a meta-analysis study (99). The proposed theoretical explanation for the combined risk is due to DQ heterodimers trans position (i.e., on the opposite chromosome) and to the DQ molecules encoding in cis position (i.e., on the same chromosome), on the cell surface. Each chromosome carries the *DQA1* gene encodes an α polypeptide chain and a β polypeptide chain from the *DQB1* gene. On the same chromosome, both the encoded α and β chains can create heterodimers, and commonly the α and β chains encoded on different chromosomes can form heterodimers also, therefore giving the possible expression of a total four different DQ molecules on the cell surface. Schematically, the trans configuration of *DQA1* and *DQB1* pairs were found in on a typical high-risk DR3/DR4 T1D-predisposing genotype. In comparison, the combination of *DQA1*05:01* and *DQB1*03:02* has not been noticed encoded in a cis arrangement, but studies showed that this configuration might confer a very high T1D risk (80).

DP gene contributions to T1D risk

The *DPA1* and *DPB1* genes encode the DP molecule might contribute to T1D risk. As mentioned earlier, because DPA1 gene low polymorphism, studies have shown that *DPB1*03:01* (*DPA1*01:03-DPB1*03:01*) is associated with T1D susceptibility and *DPB1*04:02* (*DPA1*01:03-DPB1*04:02*) and *DPA1*01:03-DPB1*01:01* with protection (100). Studies reported also that *HLA-DPB1*04:02* protects against T1D autoimmunity in the highest risk *DR3-DQB1*02:01/DR4-DQB1*03:02*, while *DPB1*03:01* and *DPB1*02:02* are predisposing to T1D risk (101, 102). In T1D genetic association studies, LD analyses of the DR and DQ-encoding genes on the chromosome as genetic variants of the HLA-A, B, and AIF1 loci show independent associations (103).

HLA class I gene associations with T1D risk

The autoimmunity inflammatory process that leads to immune destruction of the insulin-producing beta-cells in the pancreas is an inflammation produced by cytotoxic ($CD8^+$) T cell killing. MHC class I molecules help shape the T cell repertoire and present selected antigens to $CD8^+$ T cells to initiate the cytotoxic T cell killing process. The role of specific combinations of HLA class I is likely to influence beta-cell destruction. It is shown in *HLA-A*24* correlates with low residual beta-cell function in T1D patients (104). Many studies have localized HLA class I susceptibility with T1D, also combining the estimated LD with the *DR* and *DQ*-encoding genes (105-107). Results showed *HLA-B*39:06* to have the strongest predisposing effect on T1D risk (OR=10.31). In comparison, *HLA-B*57:01* showed a protective effect (OR=0.19)(107).

Additional studies using LD association in HLA class I with DR-DQ locus (106-108), showed specifically the association of *B*39* and *A*24* alleles (68). As *B*39* has been shown to contribute risk and early onset at T1D diagnosis (106), the subtypes *B*39:06*, *A*24:02*, and *A*29:02*, incorporated into a precise genetic risk score (T1D GRS2) to distinguish diabetes subtypes and to prognosticate T1D in newborn screenings (109). The presence of the *B*39:06* allele enhances the risk of T1D when present on distinct HLA-DR-DQ haplotypes (*DRB1*08:01-DQB1*04:02* and *DRB1*01:01-DQB1*05:01*) (110). By contrast, *B*18* was associated with accelerated progression from autoimmunity to T1D, but only in subjects carrying DQ2, while *A*24* promoted rapid progression to T1D in the presence of DQ8 (111).

Many studies observed risk association of the HLA class I *B*39:06* risk allele, with one study recommending that it can improve T1D risk prediction, particularly patients carrying the moderately predisposing *DRB1*04:04-DQA1*03:01-DQB1*03:02* (DR4) haplotype and *HLA-DRB1∗08-DQB1∗04* haplotype (112). Initial studies showed *HLA-A*01:01* class I allele is part of the conserved T1D risk recognized as "A1-B8-DR3" haplotype, defined by using LD analysis of the HLA region association with T1D(61). However, *A*01:01* was significantly protective for T1D when LD with the DR3 haplotype arranged the predicted allele frequencies (108). In Addition to HLA class I alleles risk effects, they can predict the age of T1D onset (108, 113, 114).

Genetic associations of T1D autoantibodies

Recent studies investigated the appearance of initial β-cell targeting autoantibody after childbirth changed the view of genetic risk factors. As mentioned, it is well-known that the *HLA-DR4-DQ8* and *HLA-DR3-DQ2* haplotypes are the two principal risk contributors for T1D risk, these two haplotypes are also the critical risk factors for the development of T1D autoimmunity and β-cell targeting autoantibodies (115, 116). Furthermore, these HLA risk haplotypes might increase T1D development risk and can determine which type of autoantibody appears at the beginning (117). GAD65 autoantibody (GADA) was detected as the first β-cell-targeting autoantibody than insulin autoantibodies in individuals with the *HLA-DR3-DQ2* haplotype. In contrast, individuals with *HLA-DR4-DQ8* are more likely to develop insulin autoantibodies first, albeit they can develop GADA autoantibodies (118). The age at which autoantibody seroconversion is detected thus appears to be associated with these haplotypes. This finding suggests that individuals carrying any of these two risk haplotypes are at an increased risk of developing autoantibodies at a young age. Additional gene variants associated with T1D have also been investigated in association with the prevalence of

autoantibodies. T1D antibodies allele-risk associations show *HLA-DQB1*03:02* (DR4) allele is positively correlated with IA-2A, IAA, and ICA (96, 119, 120), whereas *DQB1*02* (DR3) allele is negatively correlated with IA-2A (96, 119). In the case of HLA genotypes and GADA, it is found to be different according to the diabetes stage and the age at onset of T1D autoimmunity. Actually, the association of beta-cell autoimmunity with *DQB1*02/*03:02* or *DQB1*03:02* was shown in unaffected healthy children (121-123), and in young adults with late autoimmune diabetes onset (119) or with T1D new onset (121).

T1D in different populations

T1D incidence (proportion of new cases to population per year) is increasing worldwide, and the estimated incidence is nearly 90,000 children each year (124). The T1D prevalence (proportion of T1D cases to the general population) is shown to be the highest in the European population at approximately 1 in 300 (35). The incidence is highest in Scandinavian countries (such as Finland), followed by other European countries (such as the United Kingdom), North America, and Australia, while Asian countries — such as China, Korea, and Japan, T1D is considered rare diseases. The vast majority of T1D genetic studies have studied European descent subjects. The hypothesis for population risk variation remains not fully elucidated but may be related to genetic susceptibility and environmental and lifestyle factors, including personal hygiene and infections. *HLA-DR-DQ* genotypes also vary between countries (125). Although approximately 27000 HLA alleles have been discovered (33, 34), more than 65%% of these alleles are designated as common and well-documented (CWD) (126). Some common alleles, like *DRB1*03:01*, appear in almost every studied population, whereas others are population have specific HLA distribution. The alleles frequency distribution frequency in any certain population differs from other populations and usually overlaps with each other. For instance, HLA high-risk genotypes for T1D are common in Scandinavia but are less common in Asia. Still, T1D incidence estimates in kids under 15 years old range from as high as 64 per 100,000 per year in Finland to as low as 0.1 or less per 100,000 children per year in China as shown in data from the International Diabetes Federation.

However, the distribution of *HLA-DR-DQ* haplotypes with low genetic risk in the original country may confer risk in children born to parents who immigrate to a high-risk country such as Sweden (93). Nearly 90% of children diagnosed with T1D in Scandinavia have one or both of *HLA-DR3-DQ2* and *DR4-DQ8* haplotypes (96). In Finnish T1D children, studies showed that *DQB1*03:02* allele on the *DR4* haplotype is associated with the highest risk for T1D, especially the combination with the *DQB1*02:01* allele on the *DR3* haplotype. The *DQB1*06:02* allele on the

DR2 haplotype is protective for T1D and found in about 36% of Finnish new-borns compared with about 3% of Finnish T1D children (127, 128). In Swedish studies, the prevalence of the high-risk genotypes *DQB1*02/*03:02* and **03:02* is negatively associated with age at diagnosis of T1D, whereas the *DQB1*06:02* genotype has been positively associated with age at diagnosis (82, 129).

The study population must include enough subjects to include multiple allele or haplotype frequencies to explain the susceptibility effect of an individual allele or haplotype on T1D risk. For example, unlike most *DRB1*04* alleles, *DRB1*04:03* is protective for T1D (130), but it is not easy to interpret and compare in studies of European populations without large-scale studies (80). Notable population difference is shown in *DRB1*04:03* high-frequency association (3.5%) in Asian lineage than for subjects of European descent (0.6%) in USA (131).

World populations are growing in the admixture, especially in the era of globalization. The genetic inheritance pattern is not apparent from the general disease phenotype, nor can it be obtained from clinical assessment. Therefore, the development of risk assessment modeling for all individuals, regardless of racial or ethnic background, requires the determination of the susceptibility attributed to specific alleles in extended haplotypes and genotypes readings. Genotyping DR3/4 individual with a limited, low-resolution genetic screening test for DR, will classify the individual as very high risk if positive. However, both of the haplotypes are protective of T1D. Genotyping for both DR and DQ at two-field resolution clearly shows that this individual is highly unlikely to get T1D. Approximately 40% of T1D patients carry the DR3/4 genotype; therefore, searching for high-risk DR3/4 positive individuals, even for peoples of European descent, will eliminate the remaining 60% of the projected non-DR3/4 genotype carrier patients. The need for the study of non-Caucasian populations to fully understand HLA susceptibility for T1D is emphasized. Extending the research by publishing extensive studies on non-Caucasian, underserved, and understudied populations, as seen in the increase of T1D in the temporal population (56), is the best approach for complete understanding of HLA risk.

Non-HLA genes in T1D

More than 60 genetic loci have been implicated in T1D risk (65, 132-134). These non-HLA genes can be situated both inside and outside the HLA region. They are associated with immune reactions, and their associations are of far lower in effect than those of HLA genes; however, they have reproducible T1D autoimmunity effects. These genetic factors are essential to the immune system, and only a limited number is associated with the formation of β-cell-targeting autoantibodies (117). Many of these genes are associated with other autoimmune disorders. Association

analyses for these non-HLA genes must consider the overall nature of the HLA region, where LD with DR-DQ haplotypes associated with T1D can be mistaken for verified disease association.

Genetic prediction of T1D

Several comprehensive genetic screenings for T1D risk have been conducted, including DAISY (135), TEDDY (90), TrialNet (136) use genetic analysis to select subjects for following up autoantibody testing and intervention measures. T1D autoimmunity precedes the onset of clinical disease by months to years. Proven classification and use of disease-specific prevention or intervention strategies before the occurrence of overt disease presentation depends on the ability to identify prospective patients. T1D autoimmunity starts months to years before the onset of clinical disease. Identifying future patients is not straightforward, and prevention is the ultimate goal to alter the course of the disease (137). The T1D prevention strategies require screening on a large scale for the high-risk population, and risk-individuals get appropriate test intervention. These screening measures can significantly reduce the expense and duration of clinical studies. Therefore, HLA genetic testing presents a convenient screening for T1D.

T1D prediction is becoming more attainable for risk prediction in Europeans, by utilizing the clinical predictive value estimated by risk score of the highest-risk HLA alleles with chosen single nucleotide polymorphisms (SNPs) (138). Development of risk score to disease susceptibility is a feasible strategy (139). A recent risk assessment model used 67 (SNPs) to incorporate HLA alleles, their interactions, and recently discovered non-HLA loci in T1D genetic risk score (termed the “T1D GRS2”) for proper classification of diabetes subtypes and T1D prediction in new-born screening studies (109). Risk Predicting by analyzing different HLA alleles and haplotypes is also necessary to construct valid predictive models that include different populations.

Celiac Disease (CD)

CD is a systemic immune-mediated condition triggered by gluten ingestion in genetically susceptible individuals and characterized by enteropathy, ranging from intraepithelial lymphocytosis (IEL) to total villous atrophy. The clinical presentation comprises a broad spectrum of features from gastrointestinal symptoms of diarrhea and weight loss to extra-intestinal manifestations, including iron deficiency anemia, bone loss, and neurological symptoms to no symptoms. Described for the first time in the first century before Christ, the understanding of CD has profoundly developed aided by the advances in epidemiological, clinical, and genetic research that have led to a better knowledge of the pathogenesis of the disease (140). The identification of CD associated antibodies, mainly those directed against tissue transglutaminase (tTG) (also named type 2 transglutaminase, TG2) (anti-tTG or anti-TG2 antibodies) and anti-endomysium antibodies (141), the characteristic histological features of intestinal mucosa from duodenal biopsies, the strong association with *HLA-DQ2* and/or *DQ8* in genetic analysis, and improvement after the introduction of a gluten-free diet (GFD) (142) constitute the hallmarks in CD diagnosis (143).

The etiology of CD was not completely understood until the Dutch pediatrician Dr.Dicke found the association between intake of cereals and severe malabsorption syndrome in children (144). Later studies revealed that it is the protein gluten, a protein-rich in prolines and glutamines, present in wheat, barley, and rye that drives the inflammation in the gut. Studies of CD familial nature using small intestine biopsies in the 1960s established the hereditary characteristics (145). The familial tendency in CD directed researchers to the study of genetic markers. Initial studies of HLA hinted a relationship with *HL-B 8* (19), and later more specific associations showing a very much higher prevalence of A1 and B8 in CD than in control populations (20). However, it is clear that although a vast majority of CD has these antigens, the genetic was not universal, and also, a substantial minority of non-celiacs has them. CD was thus not ascribable fully to these markers (146). Studies also suggested that non-HLA genes might share responsibility, and some indicated that a distinct haplotype might have a negative association with CD (147).

The impact of deamidation (addition of amino-functional group in post-translational modification) has been revealed through studies of T cell recognition in CD, where the disease-associated *HLA-DQ2.5* and *DQ8* molecules present gluten-derived peptides (gliadins) that have been deamidated by TG2 (148). Gliadins are bound to

HLA-DQ2.5 with high kinetic stability that leads to their sustained presentation, which is a requirement for the generation of the pathogenic T cell response. By contrast, *HLA-DQ2.2* confers a low risk of CD, and its geometry is such that gliadins cannot be stably accommodated in the peptide-binding cleft(149, 150). The CD associated *HLA-DQ8* lacks an aspartic acid residue at position β57, which creates a positively charged P9 pocket with an affinity toward negatively charged peptides (151). In CD, the deficiency of aspartic acid may adequately accommodate deamidated gliadins and may promote the recruitment of cross-reactive TCRs that carry a negative signature charge in CDR3β, which can respond to both the modified and unmodified gluten peptides (151). The deficiency of aspartic acid may adequately accommodate deamidated gliadins and may promote the recruitment of cross-reactive TCRs that carry a negative signature charge in CDR3β, which can respond to both the modified and unmodified gluten peptides (151). In chronic conditions, the range of epitopes driving pathological responses may change by epitope focusing, which is when T cell clones that directed against immunodominant epitopes outcompete other clones as reported in gluten response directed toward multiple gliadin and glutenin peptides (152). Thus, antigenic diversity in the context of polymorphic HLA molecules can promote the evolution of the T cell repertoire, from the early events that occur in the thymus during the establishment of central tolerance through to the events that occur throughout a lifetime in the periphery.

Nevertheless, the ingestion of gluten does not necessarily cause CD. Only around 1% of individuals in Western populations develop CD despite a widespread consumption of cereals. Furthermore, patients with CD present very differently with a broad spectrum of clinical manifestations at different ages of onset and variations in the severity of the mucosal lesions. The individual genetic background, most likely interacting with other environmental triggers, is seemingly contributing to this variability.

Genetic risk in CD

CD carries a strong genetic component, which has been demonstrated by epidemiological studies of siblings, showing a concordance of around 70–85% in monozygotic twins compared to approximately 20% in dizygotic twins (7, 11). Through familial aggregation studies, the risk of a patient's sibling to develop CD (relative sibling risk) is estimated at 20–60% CD (153, 154). The CD is considered a complex genetic disorder with genetic and environmental factors that likely contribute to disease to have high heritability and strong HLA association (155). It is considered a polygenic disease with an intricate non-Mendelian pattern of inheritance, involving HLA and non-HLA genes, which collectively contribute to the genetic risk of developing the disease. The evidence of a strong genetic

association reflects the central role of $CD4^+$ T cells and the contribution of particular HLA molecules associated with CD in binding with specific gluten peptides that activate the T cells response (156). The identified genetic variability is estimated to be responsible for approximately 54% of the heritability in CD as reported (157, 158) of which associated *HLA-DR-DQ* risk genotypes account for 53% of the genetic risk using recent prevalence estimates (12).

Further investigation of the *DRB3, DRB4, DRB5,* and *DRB1* haplotypes concerning DQ remains mostly incomplete, and research in this field is expanding (159). Many previous studies have assessed the risk associated with different HLA genotypes using statistical methods based on case-control studies (160, 161). The genetic risk still constitutes a substantial part of the inheritance, and the need for further genetic association studies to reveal additional heritability factors are required.

The population prevalence of CD is about 1/91 and heritability is estimated to be 87%, the influence of shared environmental factors to 12%, and the contribution of the unshared environmental component of variance at the 1% level (7). In a register-based twin study of CD cases, the heritability of diagnosed CD was estimated at 75% (55% to 96%), and the non-HLA heritability contributed to 68% (40%-96%)(155, 162). Surprisingly, the HLA loci only account for an additional 6% of the heritability of the diagnosed CD. A possible reason for the low estimate could be that the HLA alleles were estimated from population frequencies and not genotyped(163). Further exploring the unexplored heritability of CD and including extended HLA genotyping will conceivably reveal additional HLA alleles outside the *DR-DQ* region in the population carrying the genetic peril.

HLA class II genes in CD

The strongest and best-characterized genetic susceptibilities in the CD are HLA class II genes, known as *HLA-DQ2* and *DQ8.* Overall, *HLA-DQ2* and *DQ8* are present in almost 40% of the Swedish population, whereas only approximately 1% of the population is diagnosed with CD (164), meaning that other factors besides these genes are implicated in disease progression. In the 1980s, the alleles-encoding *HLA-DQ2* were identified as the main factors responsible for the genetic risk conferred by the HLA genes (156, 165), which had been previously attributed to the *HLA-B*8* and *HLA-DR3* alleles (19, 166). *HLA-DP* alleles were reported to be associated with CD risk in 1989 (167) and *HLA-DPB1*04:01* has been suggested to decrease the risk to develop anti-TG2 antibodies among *DR3-DQ2* positive children (168). The following series explains the studies HLA genetic associations with CD.

HLA-DQ2.5

About 90% of CD patients carry alleles encoding HLA-DQ2(169), explicitly, HLA-DQ2.5 encoded by the *DQB1*02* and *DQA1*05* alleles, which can be inherited in cis (in the presence of *HLA-DRB1*03*) or trans configuration. Less than 1% of CD patients lack these HLA-DQ2.5 haplotypes (169). The resulting *cis* and *trans*-HLA-DQ2.5 molecules differ in residues, which are not required in peptide identification (one residue in the leader peptide in the α-chain and one residue in the membrane-proximal domain in the β-chain) and are described as conferring similar risk. These genetic configurations are similarly associated with CD (170). The frequency of HLA-DR3–DQ2 homozygosity conferred the highest risk of CD autoimmunity and prevalence in countries including the United States, Finland, Germany, and Sweden and was associated with the earliest CD onset as TEDDY study showed the CD risk in children with DR3–DQ2 homozygosity was increased by more than 2.5 times compared with the children group with a single DR3–DQ2 haplotype and more than five times that in the lowest-risk groups that was studied (DR4–DQ8 homozygotes and DR4–DQ8/DR8–DQ4 genotype)(171).

In addition to the increased risk of developing CD in patients encoding two permissive heterodimers, e.g., being homozygous for DQ2.5, are at risk of developing a more severe CD phenotype (172, 173) with earlier disease onset, diarrhea, anemia at presentation, more profound villous atrophy by histology (174), and a slower rate of villous healing on a GFD (175), and a higher rate of refractory (non-responsive) CD and enteropathy-related T cell lymphoma (176). This "HLA gene dose" effect is supposed to be related to an increased presentation of gluten-derived peptides by DQ2.5-homozygous individuals and used in stratifying risk for CD in a large at-risk population (177).

HLA-DQ8

Besides DQ2.5, there is an *HLA-DQB1*02* gene dosage effect encoded by the *DQB1*03:02* and *DQA1*03* alleles. The denotation of "8" is not apparent in either of these alleles because several *DQ3* variants are recognized differently by serology. Therefore, the *DQ3* group was "split" into *DQ7*, *DQ8*, and *DQ9* proteins, encoded by *DQB1*03:01*, *DQB1*03:02*, and *DQB1*03:03* alleles, respectively. Importantly, only the *DQ8* protein, encoded by the *DQB1*03:02* alleles, is approved for CD and is present in approximately 20% of patients with CD (171). Earlier studies showed that when both progenitors inherit this allele, the risk of developing CD is higher than when the child inherits this allele from only one of the two progenitors (166). This pattern of inheritance is conditioned to the presence of at least one copy of *HLA-DQA1*05.* Most of the remaining patients not carrying *HLA-DQ2.5* carry *DQA1*03* and *DQB1*03:02* , encoding the *HLA-DQ8* molecule (178). A similar "gene dosage" effect for HLA-DQ8 was also proposed (169). In

nearly all CD patients who carry neither *HLA-DQ2.5* nor *HLA-DQ8*, one of the two alleles encoding *HLA-DQ2.5* is present: most commonly *DQB1*02 (HLA-DQ2.2)* and less commonly *DQA1*05 (HLA-DQ7.5)*(169).

HLA-DQ2.2

HLA-DQ2.2 heterodimer share similarities with *DQ2.5* heterodimer, except that *DQA1*02* allele, encodes the DQα chain, instead of *DQA1*05* allele in *DQ2.5*. Although the resulting "2.2" heterodimer is significantly less able to presenting gluten-derived peptides compared with the 2.5 heterodimers (179, 180), most celiac patients without *DQ2.5* or *DQ8* (approximately 5%) encode the *DQ2.2* heterodimer (181). However, patients encoding either *DQ2.5* or *DQ8*, can carry *DQ2.2* heterodimer. The hereditary hazard conferred by these described HLA variants ranges from the most significant effect of *HLA-DQ2.5* to no effect attributed to *HLA-DQ7.5*. However, the role of *HLA-DQ7.5* in CD should be considered since this variant is present in nearly all CD patients lacking the known HLA associated risk variants. Risk factors in *HLA-DQA1* and *HLA-DQB1* account for 22% of CD heritability (182). Recent HLA region fine mapping using high-density imputation have identified five new CD risk independent variants in this region: *HLA-DPβ1* (position 9), *HLA-B* (the classical *HLA-B*08* and *HLA-B*39:06* alleles) and two SNPs, rs1611710, which shows a *cis*-eQTL effect on *HLA-F* expression, and rs2301226, which shows a *cis*-eQTL effect on B3GALT4 and *HLA-DPB1* expression (182). These five factors explain an additional 2.5–3% of the disease heritability. Nevertheless, the *HLA-DPB1*04:01* alleles may be in linkage disequilibrium with risk alleles in HLA-DQ genes, and further studies are required to reinforce the independence between those association signals.

HLA class I genes in CD

The HLA class I complexes function of peptide binding and presentation to form the T-cell supply in the immune system is fundamental for antigen-specific T-cell mediated cytotoxicity, explaining the genetic association of this immunological reaction. HLA class II genes strong genetic association and the identification of several *DQ2.5/DQ8* restricted gluten epitopes, explicitly recognized by CD patients, highlight the critical role of adaptive immunity mediated by $CD4^+$ T lymphocytes in CD pathogenesis (3, 4). However, the massive infiltration of $CD8^+$ T lymphocytes in the epithelium and lamina in the immune response is one of the main features of all states of CD pathogenesis, as seen in silent, active, or even refractory CD form. As discussed earlier, the primary genetic link of HLA with CD is with HLA class II *DQA1*05/DQB1*02* genes, the earliest genetic studies performed in the 1970s of HLA class I association with CD revealed associations with the *A*01* and *B*08*

alleles (21). GWAS fine mapping of the HLA region explained an additional 18% of CD heritability, independent of the DQ region, and identified *B*08:01* and *B*39:06* alleles in strong LD with the DR3-DQ2.5 haplotype (182).

Furthermore, GWAS of the HLA region identified additional risk regions independent of the *HLA DQA1*05* and *DQB1*02* genes, among them, the *HLA-B*08:01* allele, in strong LD with *DR3-DQ2.5* genes, contributes to genetic susceptibility to CD (183). It is reported that *A*01:01* and *B*08:01* influence the immune response by confining the adaptive $CD8^+$ T cell responses to gluten in subjects with CD (184).

Non-HLA genetic risk factors

It has been estimated that the 6 identified HLA and 57 non-HLA genetic variants explain around 31% of CD heritability. It is worth mentioning that non-HLA variants have been estimated to account for 6.5% of the CD heritability, mainly explained by the classically known MHC variants. Furthermore, up to now, multiple common variants with low effect (except the MHC variants) seem to be responsible for the most significant part of this known heritability, as it corresponds with the polygenic disease. The contribution of non-HLA genes to CD risk susceptibility is much less substantial (OR < 1.5) compared with the HLA-associated haplotypes risk effect (OR >5), and the collective significance in the total risk effect of these non-HLA variants is relatively modest, estimated to account for ~15% of the genetic risk (185). Overall, all the genetic variants identified to date, including HLA, explain only ~50% of the genetic variance in CD, and additional hereditary factors may potentially exist that await identification. Recent association studies showed that more than 70 candidate risk genes in over 40 non-HLA loci had been linked to CD heritability (158, 182, 186-189). These non-HLA loci encode proteins implicated in a series of immune process including stimulating T and B cell, cell migration, chemokine receptor activity, cytokine binding, thymic differentiation of $CD4^+$ and $CD8^+$ T cells, and innate immunity. Only one gene is gut-specific (RGS1), underlining the systemic character of immune dysregulation in CD (158). No current evidence for specific alleles encoding gastrointestinal proteases or tTG explaining frequent CD co-occurrence and considerable overlap between genetic risk factors with other autoimmune diseases such as rheumatoid arthritis, multiple sclerosis, and T1D (190, 191). Despite that, the overlap between CD genetic risk loci and inflammatory bowel disease such as Crohn's disease has been documented despite the weak clinical correlation (192, 193). Interestingly, 90% of the identified risk loci ,critical for CD susceptibility, are located in non-coding regions such as promoter regions, enhancers, or non-coding RNA genes, suggesting the importance of gene expression regulation rather than changes at the protein-coding level (194).

CD in different populations

The frequency of the CD-predisposing HLA haplotypes varies worldwide, but notably, CD prevalence varies in populations with a similar HLA background. Such diversity in CD prevalence may be explained by environmental factors rather than genetics. Potential environmental factors include the consumption of gluten-containing cereals (195), infection in the early years of life (196), lower economic status, and a substandard hygienic environment. There are numerous studies in the literature focused on the incidence of *HLA-DQ* in patients diagnosed with CD. In most of these studies, *HLA-DQ* was evaluated, and in some studies, only *HLA-DQA* or *DQB* alleles were evaluated. *HLA-DQ 2.5* has a high frequency in north and west Europe and portions of Africa, while DQ8 has a broader distribution, particularly in Central and South America.

European Genetics Cluster results on CD patients recruited from Finland, France, Italy, Norway, Sweden, and the UK, showed a distribution between 89.4-96.7% for *HLA-DQ2* and/or *HLA-DQ8*87-93.7% for *HLA-DQ2* and 5-8% for *HLA-DQ8*(169). In Spain, the determined prevalence was 95.6% for *HLA-DQ2* and/or *HLA-DQ8,* 93.4% for *HLA-DQ2* risk and 2.4% for *HLA-DQ8*, and (197). An Indian study showed a 100% prevalence of *HLA-DQB1*02:01*, whereas the *DQ2* occurred in 97.1% of CD patients (198). A study in Turkey found HLA-DQ2 in 67% and *HLA-DQ2* and/or *HLA-DQ8* in 76% (199). A Swedish birth cohort study showed that HLA genotyped more than 70% children born in the Skåne region, South of Sweden showed that *HLA-DQB1*02* and *DQB1*03:02* alleles occurred in 56.8% of the general population of which CD was exclusively detected in children having any of these two alleles (200). The comparison of all these studies reveals that the studies evaluating *HLA-DQ* haplotype and *HLA-DQB1* allele are almost comparable. All these genetic analyses confirm variability of CD incidence across various geographical regions in presence *HLA-DQ8*.

According to the data originating from western countries, the percentage of negative *HLA-DQ2* and *HLA-DQ8* CD patients can reach 10%, but there are also other studies reporting contradicting higher percentages. A study conducted in Chile reported that 62.9% of patients were *HLA-DQ2* and *DQ8* negative (201). A cohort study from the south of Italy found that 4.2% of CD patients were *DQ2/DQ8* negative and *DQ7* was the most frequent haplotypes in all CD patients and significantly less frequent in *DQ2/DQ8* positive CD patients (24%) than in *DQ2/DQ8* negative patients (38%)(202). These results show that the incidence of CD varies independently to HLA-*DQ2* and *DQ8* negative among different populations, and *DQ7* play as an additive or independent CD risk haplotype. Furthermore, the negative predictive value is frequently attributed to the absence of *HLA-DQ2* and *DQ8* in subjects at risk of CD.

Overlap of CD and T1D

Patients with T1D and CD patients are at a higher risk of developing other autoimmune diseases, partially explained by shared genetics (203). It has been reported that between 10-30% of T1D (204) and 35% of CD patients (205) develop a second autoimmune disease, which is partially due to active screening of patients with other organ-specific autoantibodies (206). Still, the risk is considered to be higher than in the general population. A Danish epidemiological study showed that the prevalence of autoimmune diseases was 16.4% among CD patients compared with 5.3% in the general population in 2016 (207). In particular, the presence of either T1D or CD appears to enhance the risk of developing the other disease in the same individual. Both diseases share common risk factors in genes, environmental, and immune dysregulation mechanisms. Approximately 5% of patients with CD have T1D diagnosis, and ~6% of T1D patients have a CD diagnosis (208, 209). Generally, the hypothesized causes for T1D occurrence with CD include shared HLA genetic risk (190, 210, 211) or shared environmental conditions (212, 213). Other studies suggest that double disease co-occurrence is difficult than it can be described by shared genetic risk loci only (214). Another plausible explanation for this difference in prevalence could be regional differences in extended HLA class II risk genotypes for the two autoimmune diseases.

T1D w/CD epidemiology

Walker-Smith first described the coexistence of T1D-CD in 1969 (215). While the CD prevalence is 0.3% to 1% in the general population of all ages (216), its weighted pooled prevalence was 5.1% in patients with T1D (217). The CD occurrence in T1D patients is 5–7 times more common than the overall population. 4% -9% of the T1D patients have been diagnosed with CD, compared with 1% CD incidence of CD in the general population (218). CD occurs in 3-16% of patients with previously diagnosed T1D (208, 219). Conversely, individuals with prior CD are at a three-fold increased risk for T1D before the age of 20 (220). The risk of developing both diseases is thus significantly higher than that of the general population, which is suggested to be partially explained by shared genetics (213, 221). CD in T1D patients is commonly asymptomatic or slightly symptomatic, and the diagnosis is reached through the routine screening. The hypotheses behind the

CD prevalence has been portrayed as a model of an "iceberg," where the asymptomatic cases represent the bulk of the iceberg that is not noticeable and symptomatic cases represent only the visible tip of the iceberg. Thus, the hidden CD prevalence is hypothesized to be the bottom layers of the CD “iceberg”. In most subjects, the diagnosis of T1D precedes the development of CD, while it is less likely cases with double diseases are diagnosed simultaneously. In many cases, the CD is approximately diagnosed within two years after diabetes presentation (222). Although approximately 11% to 25% are diagnosed with CD first (217), most epidemiological studies have investigated T1D patient's risk for developing CD. Recent epidemiological studies from different countries that are mainly wheat consuming showed higher T1D w/CD prevalence rate including Saudi Arabia (11.3%), Denmark (10.4%), Sweden (9.67%), Canada (7.7%), Italy (6.65%), and Iran (6.2%) (222-227). In comparison, lower coincidence rates are seen in other countries from Tunisia (5.3%), Austria (5%), Australia (5.7%), United Kingdom (4.42%), and Egypt (4%) (228-232). All these epidemiological estimates have been described for cleared biopsy-proven CD cases, confirmed by serological screening with autoantibodies. Consequently, these figures represent the classically diagnosed CD incidences and lack the estimation of latent or silent CD forms.

Genetics in patients with T1D w/CD

Studies showed that shared genetic predisposition to both T1D and CD is attributed to being homozygous for *HLA-DQ2* or heterozygous for *DQ2.5/DQ8*, of which the latter genotype seem to predispose patients to develop both diseases (233, 234). Although both diseases are also associated with HLA class I gene variants as discussed previously (67, 235), there is a paucity of genetic epidemiology studies on patients with T1D w/CD.

Possibly this shared genetic overlap might be the primary factor leading to the concomitant occurrence of T1D and CD. Reasons for concurrent CD and T1D include shared HLA genetic risk (190, 211, 236) and shared environmental exposures.(212, 213). However, data suggest that disease co-occurrence is higher than can be described by shared genetic risk loci (214). Several clinical studies described shared risk factors and mechanisms of T1D and CD, including TEDDY study that found T1D usually precedes CD autoimmunity. It also found that co-occurrence is greater than the frequency explained by demographic and genetic factors (214). A main finding of the T1DGC study was that common genetic variants contribute to T1D and autoantibodies associated with CD (237).

Suggested hypothesis for the increased susceptibility to CD and T1D coexistence is the presumed presence of DQ2.5 and DQ8 heterodimers encoded by alleles in trans, in addition to the DQ molecules encoded by alleles in cis, on the cell surface of

immune cells demonstrated in two alleles in LD, *DQA1*05*, and *DQB1*02*, encoding the DQ2.5 molecule, and *DQA1*03* and *DQB1*03*, encoding the DQ8 molecule (238). Consequently, the associations of these two *DR-DQ* genotypes (*DR3-DQ2.5/DR3-DQ2.5* and *DR3-DQ2/DR4-DQ8*) with CD and T1D indicate that the mechanism of autoimmune susceptibility may partly be overlapping.

In GWAS, only three non-HLA loci were identified as related to T1D w/CD: RGS1 on chromosome 1q31, IL18RAP on chromosome 2q12, and TAGAP on chromosome 6q25 (190), respectively. However, the association of these single nucleotide polymorphisms is weak compared to the association seen with HLA (190). Analysis of reported shared loci in T1D and CD show considerable variants overlap associated with these two conditions demonstrated in of 8 CD loci, 6 loci showed association with T1D as well, and over of the 17 loci associated in T1D, 8 showed an association with CD. Many of non-HLA loci are also associated with other autoimmune diseases (239).

Association of HLA genes

Nearly all patients with T1D w/CD carry either the *DRB1*03-DQA1*05:01-DQB1*02:01* (*DR3-DQ2.5*) or *DRB1*04-DQA1*03-DQB1*03:02* (*DR4-DQ8*) haplotypes (240, 241). As previously discussed, the *DR4-DQ8/DR3-DQ2.5* genotype is associated with the most significant risk of T1D (71) and *DR3-DQ2.5/DR3-DQ2.5* genotype the highest risk for CD (242, 243). Although DR3-DQ2.5/DR4-DQ8 and DR3-DQ2.5/DR3-DQ2.5 positive individuals are at higher risk to develop T1D w/CD than individuals carrying other HLA genotypes (244), a Norwegian study showed that patients with T1D w/CD have HLA profile more similar to T1D patients than CD patients (211). However, additional risk factors on the extent of developing T1D w/CD independent of known HLA class II including extended HLA haplotypes and non-HLA genes are yet to be determined (244, 245).

Although the NGS technology has facilitated the definition of full-length HLA gene sequences, allowing an in-depth characterization of population HLA heterogeneity (246), extended HLA class I allelic and haplotype diversity in T1D and CD have not been well studied. There are different genetic associations between patients with T1D w/CD, T1D only, or CD only. The result of genetic risk is based, principally, on particular shared alleles and genotypes in the HLA class II region, with some support for HLA class I that may be linked through extending the genotyping analysis. The HLA and non-HLA loci found in further studies can be used as stratification factors in the building of risk models to predict double autoimmunity. Previous studies were based on lower resolution (1st and 2nd field) HLA typing of selected HLA loci in individuals who share both diseases, providing incomplete insight into the HLA diversity (235).

Aims of the thesis

The aim of the thesis was to test the hypothesis if subtypes of *HLA-DR, DQ,* and *DP* are associated with T1D, CD as well as T1D w/CD using NGS. To achieve these aims, four studies were conducted (Paper I-IV), each with the following specific aims:

I. To test the hypothesis that *HLA-DRB3, DRB4*, and *DRB5* alleles modify the risk conferred by *DRB1* for islet autoantibodies and T1D genetic risk (Paper I).

II. To test the hypothesis that *HLA-DRB1, DRB3, DRB4,* and *DRB5* affect the risk of CD in relation to the *DQA1* and *DQB1* haplotypes (Paper II).

III. To extend the findings of our two previous investigations and perform high-resolution genotyping using high-throughput NGS technique in children with T1D and CD in the search for shared extended HLA class II loci in children that develop T1D w/CD (Paper III)

IV. To investigate if HLA class I alleles differ between children with T1D and CD and if a specific HLA class I alleles contribute to disease risk in T1D w/CD children (Paper IV).

Study populations

Paper I

A case-control study design included nine hundred seventy patients (n=970) given a diagnosis of diabetes between 9 months and 18 years of age were sequentially enrolled from a nationwide Swedish BDD study (87, 93, 247). American Diabetes Association and World health organization criteria were used for the diagnosis and classification of diabetes. Patients included had at the time of clinical diagnosis one or several of T1D autoantibodies against insulin: IA-2A, and ZnT8RA, ZnT8WA, or ZnT8QA. Four hundred forty-eight control subjects (n=448) matched for age (1–18 years), sex, and place of residence were analyzed at the same time (95).

Paper II

This case-control study comprised of 278 patients (174 females and 104 males) recruited as part of the GENEX study (248). Study participants were investigated with an upper endoscopy with serial intestinal biopsies taken from the bulb and duodenum at median age 9.8 years (1.4-18.3 years) between 2010 and 2012 at the Department of Pediatrics, Skåne university hospital situated in Malmö, Sweden. All patients serum samples were assessed at time for intestinal biopsy for both IgA and IgG autoantibodies against tissue transglutaminase (tTGA) using radioligand binding assays previously described (249).Among the 143 patients selected as cases, 118 had untreated celiac disease, 4 treated celiac disease and 21 were persistently tTGA positive and classified as having potential celiac disease (**Table 3**). For the 135 patients selected as controls, celiac disease was ruled out by the findings of intestinal biopsy and/or serology. Study groups were selected to be of Scandinavian ethnicity and non-Scandinavian ethnicity based on parents' place of birth.

Table 3.
Proportion (%) of children with CD (cases) compared to non-celiac disease controls. Grading of intestinal biopsies were applied according to the Marsh-Oberhuber classification and a Marsh score >1 was compliant with biopsy-proven CD (250).

characteristics	Cases (n=143)	Controls (n=134)	*P*-value
Gender	71.6	58.2	<0.001
Female	32.9	39.2	
Male			
Age (year)			<0.0001
0-	6.7	38.1	
5-	14.7	34.3	
10-	39.6	15.7	
15-20	42	9.1	
Marsh score			<0.0001
M0	8.2	72.4	
M1	9.8	25.2	
M2	2.2	0.00	
M3A	21.7	7.00	
M3B	34.3	0.00	
M3C	26.6	0.00	
IgA-tTGA			<0.0001
Negative	8.2	91	
Positive	92.2	8.5	
IgG-tTGA	9.7	93.3	<0.0001
Negative	90.8	6.4	
Positive			

Papers III & IV

Both studies III and IV used the same study subjects. Included children that were prospectively followed in a birth cohort that screened for T1D and CD between 2004 and 2010 at the Unit of Diabetes and Celiac disease, Department of Clinical Sciences, Lund University, Malmö, Sweden, as described (251). A total of 219 children were diagnosed with CD (137 females, 82 males, n=219) at median age 4.5 (range 1.1-11.0 years) according to ESPGHAN criteria (252), 68 children were diagnosed with T1D (39 females, 29 males, n=68) at median age 5.5 (range 0.9-11.3) years according to the American Diabetes Association criteria (253), and seven children (5 females, 2 males, n=7) with double diagnosisT1D w/CD . Representing the general population (GP), 448 healthy Swedish children (254) and 188 healthy children randomly selected from the LifeGene prospective cohort study

were included as controls (n=636)(255) in study III were matched for high-resolution NGS of HLA class II.

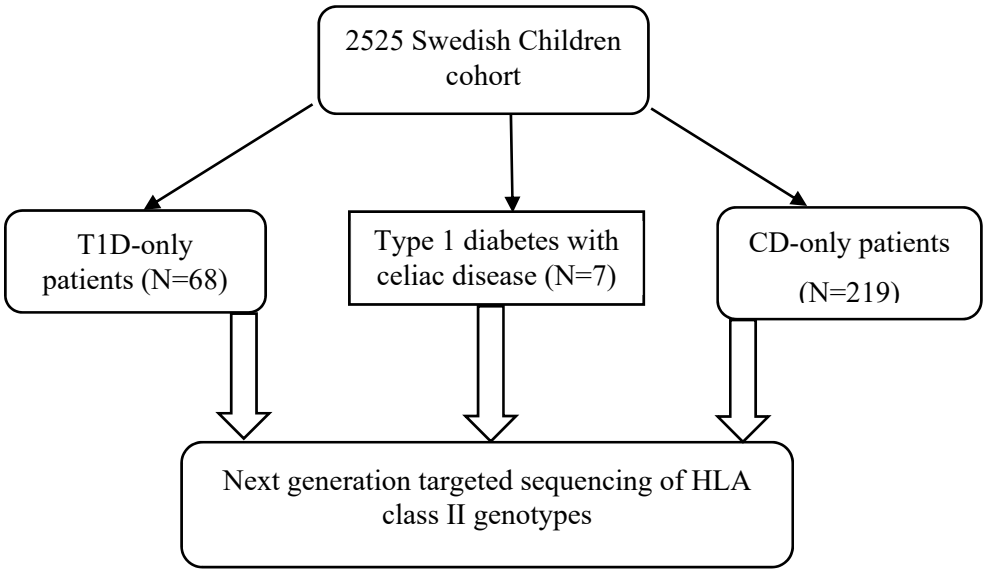


Figure 7. Flow chart of study groups

Methods

Laboratory Methods

DNA Extraction

CPT™ (Cell preparation Tubes) was used for blood collection from patients and controls at assigned clinical localities. After collection, CPT was sent to the laboratory for DNA extraction. After centrifugation, blood was phased into layers, plasma on top, a cloud of mononuclear cells above the gel-component in the tube, and the dense bottom layer of red blood cells. This granulocyte containing layer at the bottom of the tubes was used for DNA isolation using The Plasmid Maxiprep Kit (QIAGEN) according to the manufacturer's instructions (Qiagen, Hilden, Germany) from frozen whole-blood samples of patients and control subjects.

Islet Autoantibodies

Paper I included quantifications of insulin autoantibodies including GADA, IA-2A, IAA, and the three variants of ZnT8A (ZnT8RA, ZnT8WA, or ZnT8QA) that were determined in quantitative radiobinding assays by using in-house standards to determine biochemical levels as previously described in detail in previous studies (247, 256).

HLA High-resolution NGS analysis

HLA high-resolution sequencing of both class I and II was performed with the ScisGo HLA v4.0 typing kit of the genotyping system according to the manufacturer's instructions (Scisco Genetics Inc., Seattle WA, USA) using MiSeq v2 PE500 (Illumina, San Diego, CA)(53, 54, 257). Robust assays for each target loci of all class I & II loci were used providing a depth of genotyping extending. The principle behind NGS technology is DNA polymerase catalyzes the incorporation of fluorescently labeled deoxyribonucleotide triphosphates (dNTPs) into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, at the point of incorporation, the nucleotides are identified by fluorophore excitation. NGS extends the process across millions of fragments in a massively

parallel fashion. It delivers high accuracy, a high yield of short error-free reads, and a high percentage of base calls above Q30 than methods with long reads.

Illumina NGS workflows include four serial steps:

1. **Target generation**: PCR amplification of specific HLA loci to generate a large amount of target DNA. Longer amplicons copies "PCR product" are produced, and DNA concentration is calculated.
2. **Library Preparation**: Three sequential steps prepare the sequencing library; random fragmentation of the DNA or cDNA sample, followed by 5′and 3′adapter ligation and finally, Barcoding or "tagmentation" that combines the fragmentation and ligation reactions into a single step that significantly increases the efficiency of the library preparation process. DNA fragmentation is a random process, which guarantees that shorter overlapping sequencing reads encompass the complete gene. Enzymatic cleavage is directly followed by DNA repair and A-tailing in a common reaction. Adapter-ligated fragments aim to add primers to the end of the DNA fragments to enable sequencing. The fragmentation process forms DNA fragments with A-overhang; adaptors come with T-overhang to enable ligation. Illumina NGS later uses the sequences of adaptors. Following adaptor ligations is DNA cleaning and size selection. This step removes components that could eventually interfere with sequencing. Selection with SPRI (Solid Phase Reversible Immobilization) beads aim to sort larger fragments, which is preferable in the later process of phasing. Barcoding enables samples from different individuals to be pooled and run in single sequencing analysis. Indexing PCR is performed as each fragment is elongated to contain individual barcodes and flow cell attachment sites. The resulting library-pooled sample includes DNA from several individuals and loci.
3. **Cluster Generation**: For cluster generation, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the DNA templates are ready for sequencing, each including copies of the same DNA fragment.
4. **Sequencing**: Sequencing-by-synthesis is sequencing technology used primarily by Illumina, in which a DNA polymerase synthesizes a strand of DNA complementary to a template by incorporating a fluorescently labeled deoxynucleoside triphosphate that is imaged to identify the base and then cleaved before the process is repeated to determine the order and identity of each base in the DNA strand. It employs a proprietary reversible terminator–based method that detects single bases as they are incorporated into DNA template strands. As all four reversible terminator–bound dNTPs are present

during each sequencing cycle, natural competition minimizes incorporation bias and significantly reduces raw error rates. The result is highly accurate base-by-base sequencing that virtually eliminates sequence context-specific errors, even within repetitive sequence regions and homopolymers.

Data Analysis

During data analysis and alignment, the newly named sequence reads are aligned to a reference genome. Following alignment, many variations of HLA analysis are potential, such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for DNA methods, phylogenetic or metagenomic analysis. HLA typing results for the new samples are analyzed using known sequences by software analysis.

In short-read sequencing by Illumina technology, DNA fragments are ligated to adapters. The adapters contain unique molecular identifiers as well as sequences complementary to the oligonucleotides attached to the surface of a flow cell. Adapter-tagged DNA is loaded onto a flow cell, and the adapters from the modified DNA hybridize to the oligonucleotides that coat the surface of the flow cell. Once the DNA fragments have attached, cluster generation begins, where thousands of copies of each fragment are generated through a process known as bridge amplification. In this process, one strand folds over, and the adapter on the end of the molecule hybridizes to another oligonucleotide in the flow cell. A polymerase incorporates nucleotides to build double-stranded bridges of the DNA molecules, which are subsequently denatured to leave single-stranded DNA fragments tethered to the flow cell. This process is repeated over and over, generating several million dense clusters of double-stranded DNA. After bridge amplification, the reverse DNA strands are cleaved and washed away, leaving only the forward strands. Then, sequencing by synthesis begins, in which fluorescently labeled deoxynucleoside triphosphates are incorporated into the newly synthesized DNA strand at each cycle. After incorporation, a laser excites the fluorophore on the strand, emitting a characteristic fluorescence signal that corresponds to the base.

Statistical Methods

Paper I

The project was performed to determine allelic association analysis of the multiallelic genes of *HLA-DRB1, -DRB3, -DRB4*, and -*DRB5* in the diabetes group (n=970) and control subjects (n=448). Under Hardy-Weinberg equilibrium, the allelic analysis computes allelic frequencies among patients only, control subjects only, and the combination of both by calculating sets of alleles of varying frequency, effect size and direction that disrupt the same gene. Given excessive HLA polymorphism, a score test used test allele-specific associations with T1D (258, 259). The score test estimates the allele-specific arrangement following the null hypothesis with no allelic associations and is referred to as the H-score because it was particularly developed for testing haplotype associations. Under the null hypothesis, H-score has an asymptotic normal distribution, which is utilized to calculate the P-value. The statistical approach was to consider all potential haplotype combinations for ambiguous HLA data. For all estimates, we used the R function haplo.cc found on (http://cran.r project.org/web/packages/haplo.stats/index.html). The function haplo.cc computes the haplotype-based association used for the haplotype analysis described next and assesses the allelic association by introducing a monomorphic locus as the second locus. The algorithm estimates the likelihood probability by counting all compatible haplotypes by the frequency of haplotypes pairs. The significance of individual haplotype effects was estimated using Wald statistics. To estimate allele-specific magnitudes of the T1D association, we chose the reference allele with comparable allelic frequencies between patients and control subjects and had a relatively high allelic frequency. In comparison with this reference allele, we calculated the OR for every allele. The OR of the reference allele is given the 1 value, OR <1 implies a protective allele, and an OR >1 is a risk allele.

In the *HLA-DR* locus, alleles of *HLA-DRB1* are in LD with alleles of either *HLA-DRB3 or -DRB4 or -DRB5*. Hence, by treating these subunits and their allelic variations as different alleles, analysis of HLA-DRB1 and -*DRB3, -DRB4, or -DRB5* by haplo.cc produces estimates of frequencies for all haplotypes provided that expectation numbers of corresponding haplotypes are five or more copies in both patients and control subjects. Similarly, the function haplo.cc produces H-scores and P values for all included haplotypes. By following the same principle of

choosing the reference haplotype, we computed haplotype-specific ORs. A two-dimensional table was created to facilitate the understanding of estimated haplotype-specific ORs by showing H-scores for specific pairs of alleles at *HLA-DRB1 and -DRB3, -DRB4, or -DRB5*.

Paper II

A frequency matching strategy, matching cases, and controls by gender was applied in Paper II. Allelic association analysis of multiallelic genes of *HLA-DRB1, DRB3, DRB4, DRB5, DQA1, and DQB1*, was performed. Alleles of *DRB3, DRB4, and DRB5* at the DR locus do not jointly reside on the same chromosome. These 3 genes are assigned as a "single gene" DRB345 for the analytic facilitations. For each gene, allelic frequencies were computed among cases, controls, and combined cases and controls, under the Hardy Weinberg equilibrium. To test if the allele is significantly associated with CD, the same allele-specific in Paper I score test was applied(258). The H-score test evaluates the allele-specific score under the null hypothesis with no allelic associations and it was developed specifically for testing haplotype-association. H score has an asymptotic normal distribution, which is then used to compute P-value under the null hypothesis.

To retain meaningful interpretation on allele-specific odds ratios, it is estimated that under the null hypothesis for a reference allele, its allelic frequency in cases should be comparable to that in controls, and thus their ratio equals 1. In comparison with this presumed reference allele, odds ratios were calculated marginally for every allele as the ratio of allelic frequency in cases over that in the controls, that is, the ratio of exposure rates (RR). RR of less than 1 implies a protective allele, and RR of greater than 1 is a risk allele. Within the DR locus, alleles of *DRB*1 are in linkage disequilibrium with alleles of DRB345.

Similarly, *DQA1* and *DQB1* are in high LD with each other in the DQ locus. Further, DR and DQ loci are physically adjacent to each other, and many of their alleles are highly associated with each other. Since phases of these individual genes are incomplete, statistically driven haplotype analysis on DRB1-DRB345, DQA1-DQB1, or DR-DQ was performed.

Without assigning haplotypes directly, the statistical approach is to enumerate all possible haplotype configurations with empirically computed prior probabilities and produces estimates of haplotype frequencies for all possible haplotypes, provided that expectation numbers of corresponding haplotypes are 10 or more copies in the combined cases and controls. Just as those statistics produced for allelic association analysis, haplotypic frequencies were computed among cases, controls, and pooled case-control. Then, the function "haplo.cc" produces H-scores and P-values for all included haplotypes. Following the same principle of choosing the null reference

haplotype, in other words, comparing haplotypic frequencies between cases and controls, the haplotype-specific RR was computed. During the evaluation of allelic associations within individual genes, the bootstrap technique was used for random sampling with replacement to form a bootstrap sample to estimate the allele-specific rate of ratios (allelic frequency) among cases over controls, which is equivalent to odds ratio gave the assumption that 2 allelic frequencies are the same between cases and controls. Using 1000 bootstrap samples, we estimated associated standard errors and thus associated confidence intervals for individually estimated RRs.

Paper III and Paper IV

Allelic frequency distribution analysis of HLA-class II in paper III and HLA-class I genes in paper IV was determined by the relative predispositional effects (RPE) analysis(260). Crude odds ratios (ORs) and their associated 95% confidence intervals (95% CI) were calculated, and $\chi 2$ tests and Fisher's exact tests (if any cell contained fewer than three observations) were used to test whether the frequencies of a given allele/haplotype differed between cases and non-cases. The RPE method was used to identify the disease risk alleles, haplotypes, or genotypes with the strongest predisposing or protective effects at each iteration. The selected alleles were then removed from the dataset, and the analysis was repeated until no risk or protective alleles were identified. Comparisons of DR, DQ, and DP allele frequencies were performed for both exons 2 and 3 of chromosome 6p21 by performing pairwise comparisons between all Paper III study groups and listed in order of increasing P-value, followed by the extended haplotype and genotype frequencies. Comparisons of HLA-A, -B, and -C allele frequencies were performed for exons 1–7 of chromosome 6p21 by performing pairwise comparisons between all paper IV study groups and listed in order of increasing P-value, followed by the extended haplotype and genotype frequencies. P-values ≤ 0.05 were considered to be statistically significant, and alleles with a low frequency ($\leq 1\%$) were not shown in the analysis. The p-values presented are nominal and not adjusted for multiple comparisons. Analyses were performed in R (r-project.org) version 3.6.1 and R package epiDisplay version 3.5.0.1.

Results & Discussion

Paper I**:** *"Next-Generation Sequencing Reveal that HLA-DRB3, -DRB4, and -DRB5 May Be Associated with Islet Autoantibodies and Risk for Childhood Type 1 Diabetes"*

Main findings in Paper I showed first that by analyzing the 25 *HLA-DRB1* genotyped alleles, only four (*DRB1*03:01:01, DRB1*04:01:01, DRB1*04:04:01,* and *DRB1*04:05:01*) were found positively associated with T1D risk. The H-score was used to rank the relative degree of risk by decreasing order as follows: *DRB1*04:01:01 >*03:01:01 >*04:05:01 >*04:04:01*.

Second, NGS revealed nine genotyped *DRB345* alleles; eight of them are alleles of *DRB3, DRB4, and DRB5* plus one including chromosomes with only non-amplified loci. Only two alleles; *DRB4*01:03:01* and *DRB3*01:01:02* were positively associated with T1D risk; the remaining five alleles were negatively associated with the disease. *DRB4*01:03:01* has a high association ranking (H-score 12), which ranked higher than *DRB3*01:01:02* (H-score 4). Another important finding, *DRB4* alleles were divided in their risk effect in which *DRB4*01:03:01* was positively associated and *DRB4*01:01:01* was negatively associated with T1D as evidenced in haplotype association with *DRB1*07:01:01*. Likewise, *DRB3*01:01:02* was positively associated and *DRB3*02:02:01* negatively associated with T1D. Because either one of these two alleles may be shown on a haplotype bearing *DRB1*03:01:01*, it cannot be excluded that the risk of this allele for T1D is influenced by the DRB3 alleles to either increase or decrease the risk.

A third finding when analyzing the extended *DRB1-DRB3-DRB4-DRB5* haplotypes (28 haplotypes were identified), was that the two *DRB1*03:01:01*-containing haplotypes remained positively associated with T1D whether either *DRB3*01:01:02* or *DRB3*02:02:01* was shown (although the positively associated *DRB3*01:01:02* showed a P-value for risk that was three times higher than for the *DRB3*02:02:01*-containing haplotype). More importantly, the borderline associations of the *DRB3*, *DRB4*, and *DRB5* alleles were extensively affected by *DRB1* alleles. The interpretations of the different extended *DRB1-DRB3-DRB4-DRB5* haplotypes strongly infer that the risk for T1D cannot be attributed to a single *DRB1* allele but that *DRB3, DRB4*, and *DRB5* in LD have to be considered when dissecting the role of *HLA-DR* in T1D.

The application of NGS method in the Paper I enabled extended *DRB1-DRB3-DRB4-DRB5* haplotypes to be calculated without information on the linear descent relationship. Sequencing of coding regions of exon 1–4 of both *DRB1* and *DRB3*, *DRB4*, and *DRB5* allowed the detection of all functional genes, whereas pseudogenes were not magnified. Despite the presence of pseudogenes in the *DRB3-DRB4-DRB5* region, it was possible to compute full sequenced haplotypes without ambiguities, including in subjects with pseudogenes such as *DRB1*01:01:01*, *DRB1*01:02:01*, *DRB1*08:01:01*, and *DRB1*10:01:01*, respectively. The utilized method is a study improvement because it deducted high-resolution typing of alleles, which has been understudied. Due to reduced costs of PCR, novel instrumentations and approaches in bioinformatics make NGS HLA typing affordable and accurate.

The study population was appropriate because it represents coherent patients with lately diagnosed T1D in Sweden (87, 93). Moreover, patients in the Paper I were all born to parents born in the same geographic locations, as was the fact for their grandparents (93). According to Swedish pediatric diabetes guidelines, all patients with T1D <18 years old are seen by pediatricians at one of the pediatric diabetes clinics in Sweden. The implied vulnerability of this study is that it could not analyze a control group of equal numbers as the patients. However, the current control subjects were selected to represent the geographical location of the patients (261).

By reviewing the literature, only one published study found analyzed the NGS of *DRB* genes (79). It included 143 control subjects and 337 patients of a much larger cohort and reported that both *DRB3*01:01* and *DRB3*02:02* alleles showed an increased risk for T1D. In particular, *DRB1*03:01* and *DRB3*02:02* contribute to T1D risk. By comparison, Paper I analysis of patients with newly diagnosed T1D and control subjects shows results compatible with the mentioned outcome. However, paper I results differ because *DRB3*02:02* was negatively associated with T1D, and when considered with *DRB1*03:01:01*, *DRB3*02:02:01* reduced the risk for T1D compared with the *DRB1*03:01:01-DRB3*01:01:02* haplotype.

T1D risk haplotypes containing *DRB1*03:01:01* may either carry the *DRB3*02:02:01* or *DRB3*02:02:01* alleles. In the TEDDY study, *DRB1*03:01:01/DRB1*03:01:01* homozygous children were at an increased risk for developing GADA as their first islet autoantibody (115). Additional studies are required to determine whether *DRB3*02:02:01* affects the risk of GADA as to the first islet autoantibody (262). The negative association of *DRB1*01:03* with T1D was independent of *DRB3*, *DRB4*, and *DRB5* because it resides on a haplotype unqualified to express any of these DR subtypes. The analysis is that the *DRB1* protein heterodimer confers protection, presumably by inducing immunological tolerance. Future studies of amino acids of *HLA-DRB1* and *HLA-DRB3*, *4*, and *5* can include potentially causal residues responsible for the risk of T1D (55).

The significance of the results in Paper I is that *DRB3*, *DRB4*, and *DRB5* may affect not only the risk for T1D but also the risk of having specific islet autoantibodies. The data strongly suggest that *DRB3* on the *DRB1*03:01:01* haplotype affects the risk of having GADA or ZnT8RA (positive association) or IA-2A (negative association). Patients with *DRB1*07:01:01-DRB4*01:01:01* are at an increased risk for GADA detection at a clinical diagnosis. Therefore, determining to what extent the *DRB3*, *DRB4*, and *DRB5* β-chains can form heterodimers with the *DRA* α-chains is significant. It cannot be excluded that peptide presentation on *DRB3*, *DRB4*, or *DRB5* heterodimers may induce immune responses related to autoimmunity.

Paper II: *"Different DRB1*03:01-DQB1*02:01 haplotypes confer different risk for celiac disease"*

Paper II is the first study that implements NGS in HLA haplotype analysis in CD. The main findings showed that the highest risk haplotype for CD, *DRB1 *03:01:01-DQA1 *05:01:01-DQB1 *02:01:01*(DR3-DQ2.5) is modified at the *DRB3* locus. The most frequent of these haplotype among the Scandinavian population is *DRB3 *01:01:02* in linkage with *DQA1 *05:01-DQB1 *02:01* (hereafter called "8.1AH" for the extended ancestral haplotype *HLA-A1-B8-DRB3*01:01-DRB1*03:01-DQB1*02:01*) and the less frequent version of this haplotype carries another *DRB3*02:02:01* in linkage with *DQA1*05:01-DQB1*02:01* (hereafter called 18.2AH for the extended ancestral haplotype *HLA-A30-B18-DRB3*02:02-DRB1*03:01-DQB1*02:01*) (**Fig 8**). Even though both of these haplotypes are associated with CD, no patient carried the 18.2AH haplotype only.

Second, among CD patients who carried the 18.2AH haplotype on 1 chromosome, all carried different CD risk haplotype (i.e. either the *DRB3*01-DQA1*05:01-DQB1*02:01 (DQ2.5)*, *DQA1*02:01-DQB1*02:01 (DQ2.2)* or *DQA1*05:05-DQB1*03:01 (DQ7.5)* haplotype) on the opposite chromosome. The lack of patients carrying a single risk haplotype 18.2AH emphasizes that other HLA-linked loci are required in the disease process. If only DQ alleles were involved, the two haplotypes A8.1AH and 18.2AH (identical at the *DQA1* and *DQB1* loci) should carry the identical risk.

Third, the results of Paper II suggest that 18.2AH has a regressive conservative pattern of inheritance which is line with a previous Sardinian study (263). In discrepancy, the 8.1AH haplotype is inherited to affected individuals with an additive influence and often as a single risk haplotype where the opposite haplotype can be any other haplotype. The 18.2AH haplotype confers likely a synergistic effect combined with the 8.1AH haplotype in increasing the risk of CD in those individuals carrying homogenous variants, while the 18.2AH haplotype does not appear to confer risk on its own. These interpretations of the different extended *DRB1-DRB345-DQ* haplotypes suggest that the risk of CD cannot be attributed to the DQ locus only.

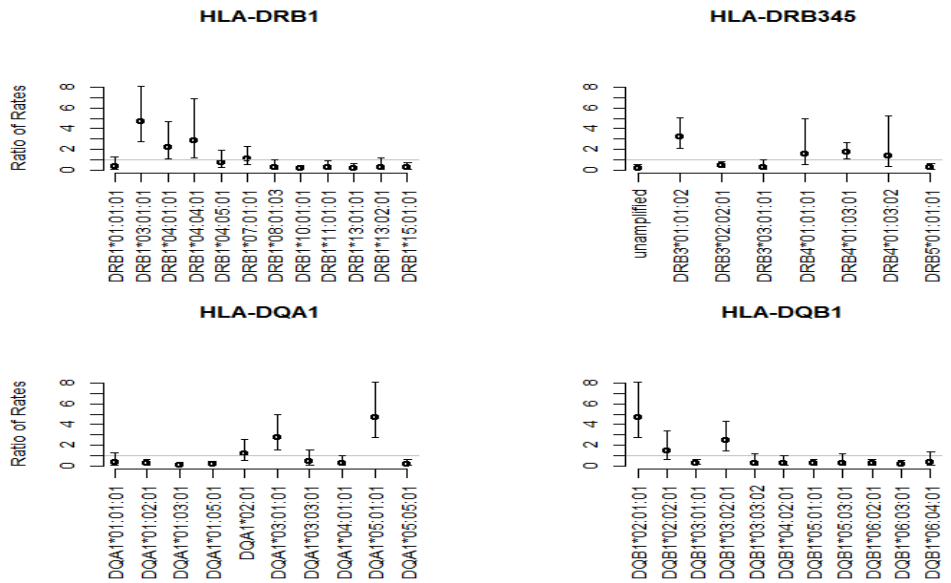


Figure 8. Association result from extended analysis of HLA class II alleles to the risk of CD.

The translation product of the *DRB3*01* and *DRB3*02* alleles generate class II heterodimer molecules with the essentially non-polymorphic alpha (DRA) chain, which is essential to the immune system by introducing peptides derived from extracellular proteins (264). However, there is inadequate data about the functional role of the heterodimers formed of *DRB3*01* and *DRB3*02*. The *DRB3*02* alleles are critical in hepatitis virus clearance, and *DRB3*02* on the *DRB1*03:01* haplotypes contributed to an increased risk for T1D when compared to the *DRB1*03:01/*03:01* homozygotes carrying only the *DRB3*01* haplotype on both chromosomes (79). *DRB3*02* may confer a synergistic effect increases the risk for T1D as well as CD in combination with another *DRB3*01* or *DRB3*02* alleles, but a potentially protective effect for CD, counteracting the *DQA1*05:01-DQB1*02:01* molecules except coupled with other HLA risk haplotypes.

The results showed that the *DQ2.5* haplotype risk can be divided into 2 distinct haplotypes; *DRB3*01-DQA1*05:01-DQB*02:01* (8.1AH) and *DRB3*02-DQA1*05:01-DQB*02:01* (18.2AH), with a seemingly different risk associated with each haplotype. Unlike 8.1AH, the 18.2AH haplotype is likely to confer only a minimal CD risk on its own, despite being identical at the *DQA1* and *DQB1* loci. The advantage of NGS for estimating *HLA-DR* and *DQ* polymorphism has presented a total of 15 novel haplotypes found to have significant haplotypic frequencies in a general Swedish population and allowed for deep dissection using

this "HLA-omic" methodology as seen also in Paper I (265). Although this study showed differences between *DRB3*01-DQA1 *05:01-DQB1*02:01* and *DRB3*02-DQA1*05:01-DQB1*02:01* on the disease risk, there are potential shortcomings with Paper II. Firstly, the sample size was somewhat modest and heterogeneous, which can affect the full spectrum of studied and unaffected variants in the population of concern. The number of informative haplotypes (*DRB3*02-DQA1*05:01-DQB1*02:01*) in our population was limited; consequently, the statistical power to distinguish the differential effect of *DRB3*02-DQA1*05:01-DQB1*02:01* and *DRB3*01-DQA1*05:01-DQB1*02:01* was modest. Secondly, the study included study participants from an exclusive clinical site in Sweden, which can contain selection probability bias. The findings of Paper II, therefore, need to be replicated in more general populations before these risk estimations of individuals for CD can be established.

Key genetic risk factors in CD are DQ2 (*DQA1*05:01-DQB1 *02:01*), and DQ8 (*DQA1*03:01-DQB1 *03:02*) has physicochemical properties and binding of gliadin-derived peptides deamidated by tissue transglutaminase 2 (tTG2). Both DQ2 and DQ8 contain positively charged pockets with a preference for binding negatively charged particles. Specifically, in DQ2, the lysine position at β71 has a preference for binding with negatively charged residues at P4, P6, and P7 positions. The autoimmune disease relevance of the *HLA-DQ8* polymorphic residue 57β was first identified for T1D (266). The DQ8 β57 polymorphism creates a basic environment with a preference for binding the negatively charged residue at P9 (267). DR3 and DQ2 are in strong LD; therefore, now it is considered that it is the DQ2 molecule that predisposes to T1D as well, as DQ2 influences the selection and binding of the autoantigenic peptide. This is well elucidated for CD, in which negatively charged gliadin peptides as such or modified by tTG bind to DQ2/DQ8 with high affinity. The lysine position at β71 in DQ2 binds to these residues at positions P4, P6, P7, and position β57 in DQ8 bind at P9 (268, 269). However, these mechanisms have not been fully resolved in T1D, as the triggering factor is not known in the last case, but it is anticipated that the "diabetogenic peptide" may be binding to DQ2 and DR3 accompanies it due to LD. Also, individuals who are homozygous for DQ2.5 or DQ8 have a five-fold higher risk of developing T1D than those who are heterozygous (270). The loss of aspartic acid at position 57 of diabetogenic HLA-DQβ chains supports Class II association with; this single amino acid change determines how TCRs recognize peptides in HLA-DQ8, and I-Ag7 using a mechanism termed the P9 switch. Recently, studies provided a mechanistic molecular explanation that links the specific HLA class II polymorphism of T1D with the recognition of islet autoantigens and disease onset in mice (271).

Besides, decreased costs of PCR, novel instrumentations, and approaches in bioinformatics, made NGS HLA typing affordable and accurate. Although Paper II showed differences between *DRB3 *01-DQA1 *05:01-DQB1 *02:01* and *DRB3 *02-DQA1 *05:01-DQB1 *02:01* on the CD risk, there are noticeable potential

shortcomings with Paper II. Firstly, the sample size was somewhat modest and heterogeneous, which can affect the full spectrum of studied and unaffected variants in the population of concern. It can misrepresent the characterization of HLA diversity in the studied population. The number of informative haplotypes (*DRB3 *02-DQA1 *05:01-DQB1 *02:01*) in the study population was limited; consequently, the statistical power to distinguish the differential effect of *DRB3 *02-DQA1 *05:01-DQB1*02:*01 and *DRB3 *01-DQA1 *05:01-DQB1 *02:01* was low. Secondly, the study included study participants from a particular clinical site in Sweden, which can contain selection probability bias. The sequencing of the HLA coding region solely will be inadequate for a complete understanding of the CD. Further analyses are required to determine the transcription of the key genes involved in the HLA functional pathway, along with physically interacting targets and/or further investigation of regulatory regions such as those containing transcription factor-binding sites.

Paper III: *"High-resolution genotyping suggests that children with type 1 diabetes and celiac disease share three HLA class II loci in DRB3, DRB4, and DRB5 genes"*

The main finding in Paper III demonstrated that three alleles, *DRB4*01:03:01, DRB3*01:01:02*, and *DRB3*02:02:01*, were found to be associated with T1D w/CD. Of these genes, *DRB4*01:03:01* was associated with T1D only, *DRB3*01:01:02* was associated with CD only, but inversely associated with T1D only. The *DRB3*02:02:01* allele was associated with a low predisposition in all three groups. The novelty of the study lies in the fact that several alleles in *DRB3*, *DRB4*, and *DRB5* genes in T1D w/CD children seem to have an extended HLA profile more similar to that in children with T1D only than that in children with CD only.

Second, *DRB4*01:03:01* containing haplotypes conferred an association with T1D w/CD as well as T1D in comparison to *DRB3*01:01:02* containing haplotypes that were highly associated with CD. Around 50% of T1D and 70% of T1D w/CD haplotypes carry the former allele compared with 60% of CD who carry the latter allele. The analyses of the extended *DRB1-DRB3-DRB4-DRB5* haplotypes likely suggests that the risk for developing both diseases likely resembles T1D risk.

Third, differences between T1D w/CD and T1D children were found of whom T1D w/CD children were more likely to be homozygous for *DR4-DQ8/DR4-DQ8* compared to T1D children.

In line with a previous study (80) the *DR3-DQ2* haplotype occurred in over a third of T1D children, and as previously demonstrated, *DR3-DQ2/DR4-DQ8* was the most frequent genotype (272). Moreover, *HLA-DQ2.5* homozygosity was more common among CD children compared with T1D w/CD and T1D children, confirming the HLA dosage effect of *DR3-DQ2* on the risk of CD (273). These

results are all in line with a previous Norwegian study (211), and a study conducted on Dutch patients (274) which showed that the T1D risk heterozygous genotype (*DQ2.5/DQ8*) provided a comparable frequency with T1D w/CD. In contrast to Bakker et al. (275) study that reported HLA-DQ2.5 homozygosity is expected in 30% of the T1D w/CD group, indicating that a double dose of DQ2.5 confers the highest modifier for T1D patients to develop CD as shown previously (276).

The strengths of the present study were the use of high-resolution NGS for extended HLA genotyping, which enabled examining the disease susceptibility between T1D w/CD and extended *HLA-DRB3, DRB4,* and *DRB5* alleles. The RPE statistical analysis method used to estimate the association between HLA alleles (or haplotypes) and each of the outcomes (T1D only, CD only, T1D w/CD) accounts for the fact that a high frequency of a given allele "induces" a lower frequency of all other alleles, as their total must remain constant (260). A limitation of the study was the small sample size for the comparison of genotype effects both within and between the three disease groups. Secondly, the study included study participants from a single site comparing children at high-risk genotypes constituting a relatively homogeneous population with little HLA diversity. It cannot be ruled out that shared HLA loci may be different in other populations.

Paper IV: *"High-resolution genotyping of HLA class I loci in children with type 1 diabetes and celiac disease"*

In Paper IV, extended polymorphism of HLA class I genes identified genetic diversities and similarities in children with T1D, CD, and in a subgroup of children having T1D w/CD of which eight alleles in the *HLA-A* region and six alleles in both *HLA-B* and *HLA-C* regions, respectively. Among those, only *A*29:02:01* and *C*05:01:01* showed a similar positive association between T1D w/CD, T1D and CD, indicating that shared genetic HLA class I loci for both diseases.

Second, NGS identified *A*68:01:02* as additional allele positively associated between T1D w/CD and T1D patients in addition to *A*29:02:01*; however, that was not shown in CD association frequency as it is only associated with *A*29:02:01*.

Third, NGS results in T1D associated *HLA-A* and *HLA-B* alleles showed a distribution that was closer to T1D w/ CD than with CD. *HLA-B* alleles were mostly represented by two alleles, *B*08:01:01 and B*15:01:01*, similarly comparable between T1D and T1D w/ CD. Only one genotype, *A*03:01:01-B*08:01:01-C*03:04:01/A*24:02:01-B*15:01:01-C*07:01:01* was shared between T1D and T1D w/CD, albeit this genotype was only found in two children. All this points to the direction that HLA class I loci in T1D w/CD patients are more similar to that of T1D as compared to CD patients and suggest significant risk contribution of T1D risk alleles.

The positive association of *B*18* and *A*24* with T1D is in line with previous reports (111, 277). In contrast to other studies, we found *B*39:06* not to be associated with

T1D (105, 278). Since *B*39* is relatively rare and found in 0.5–1.2% of Europeans, the different results between our and those previous studies might relate to regional differences in the study populations (279). Two previous studies have investigated the role of HLA class I and class II *DRB1* and *DQB1* on the risk of developing both T1D, CD or both diseases (T1D w/CD) and found that individuals with T1D w/CD are genetically more similar to T1D than to CD patients (210, 211). Two previous studies have investigated the role of HLA class I and class II *DRB1* and *DQB1* on the risk of developing both T1D, CD or both diseases (T1D w/CD) and found that individuals with T1D w/CD are genetically more similar to T1D than to CD patients (210, 211). In contrast to these previous studies that directly analyzed HLA class II in patients and control subjects, we utilized the NGTS for extended genotyping of *HLA-A, B*, and *C* without information on the descent. The sequencing of the coding region of exon 1-7 of HLA class I allowed the high-resolution detection of all associated genes differences in T1D and CD groups and subgroup T1D w/CD cohort.

Our objective was to enhance our understanding of CD compared to T1D by asking if HLA class I genes contribute to T1D w/CD. In the future, this information might help in building precise genetic risk models to identify individuals with either T1D or CD who are at high risk of developing both. In comparison to our analysis, it is still the HLA class II loci that presents the most significant association with disease coexistence (280). However, our association study shows that the HLA class I alleles that are related to T1D w/CD are not the same as those related to the frequency distribution of either T1D or CD separately.

A limitation of the present study is the inclusion of only a few patients with T1D w/CD. Despite this, it was possible through the RPE analysis to identify HLA class I alleles effect size that may have contributed to the development of both diseases. Another limitation is that it was not possible to include healthy children from the general population as a reference group for comparison. Although the underlying populations from which T1D and CD sets were collected from the same clinic, we addressed whether the allelic variation in the HLA region could be different between the two sample sets, which in turn could have resulted in predictable future disease risk.

Overall, our conclusions indicate significant differences in the genetic structure in the HLA region, especially class I markers, between CD and T1D.

Conclusions

Using high-resolution NGTS for extended HLA class I and class II genotyping in patients with T1D and CD, the conclusions from this thesis demonstrated that:

I. *HLA-DRB4*01:03:01* and *DRB3*01:01:02* have a modifying effect to *DRB1* on the risk of T1D (Paper I)

II. *HLA-DRB3*01:01:02* and *DRB3*02:02:01* alleles in linkage with *DQA1*05:01-DQB1*02:01* showed a differentiated effect on the risk of CD, which seem more predominant among patients of Scandinavian ethnicity (Paper II).

III. *HLA-DRB4*01:03:01, DRB3*01:01:02* and *DRB3*02:02:01* were identified as three risk alleles shared between T1D and CD patients of which DRB4*01:03:01 conferred the strongest risk allele for developing T1D w/CD (Paper III).

IV. The distribution of extended *HLA-A* and *HLA-B* alleles in T1D w/CD showed patterns closer to T1D but are different between children with T1D and CD. Only *A*29:02:01* and *C*05:01:01* showed a similar positive association between T1D w/CD, T1D and CD. (Paper IV)

Summary and Future perspectives

The progress in characterizing HLA gene diversity, HLA associations with human diseases, and HLA–peptide–TCR interactions and their mechanistic immunologic implications have addressed the fundamental basis of this research area: to harness the improved understanding of HLA risk contribution for T1D and CD clinical benefit. With the ultimate aim of achieving precision or even personalized medicine, stratifying patients based on their HLA genotypes and administering antigen-specific, patient-tailored disease prevention or immunotherapy presents a uniquely attainable target.

Over the prior years, long-read, single-molecule DNA sequencing technologies have emerged as advancements in genomics. These platforms have demonstrated their strength to solve some of the most challenging human genome regions through generating reads tens to thousands of kilobases in length with precision approaching and high-resolution. Next-generation sequencing permits the full-length extended genotyping of diploid genomes, which revolutionized genomics by revealing the full spectrum of HLA genetic variation, resolving some of the missing heritability, and identifying novel mechanisms of autoimmune HLA-linked diseases.

T1D is a mortifying disease that is a chronic condition requiring intensive treatment and, in the worst circumstances, may lead to an early death. The overall T1D burden must be decreased through early disease detection and early prevention by using immunogenetic characterization that draws a crucial factor in the pathway to T1D prevention. A prediction model by a combined risk score can enhance the prediction of T1D among susceptible patients. These types of interventions can significantly improve the cost and care of such patients.

Despite the tremendous advances accomplished on the understanding of CD in the last decades, many issues remain unresolved, like why only a small fraction of HLA-DQ2/DQ8 individuals develop CD or why this disease can appear at any time in life, from early childhood to late adulthood. Moreover, the percentage of diagnosed individuals who lack blood circulating antibodies against TG2 is reaching non-negligible values. These fields of study may also be underlying clinical disease manifestations and have a lot to say in the development and progression of human diseases.

Precision medicine has found a sturdy hold in the diagnosis and treatment of diabetes. The ability to combine the diagnosis of diabetes into routine clinical care is one example where genetic diagnostics are inevitable and meet many of the properties of the prototypical test. Despite an excellent diagnostic standard, there are no known ways for preventing diabetes, although careful monitoring in pre-symptomatic variant carriers may lead to early detection of diabetes and rapid treatment. Future precision diabetes medicine approaches are expected to introduce diagnostic genetic testing for predicting and determining diabetes subtypes to select the best interventional and therapeutic approaches.

Development in transforming advances in biology and technology will be administered by the classification, precise estimation, and scalable deployment of diagnostic tools for diagnosis and therapy. Precision methods must be available to the full variety of human populations and societal settings, such that precision medicine does not widen health differences but delivers the most meaningful benefits for all individuals and society.

Decades of research show that genetics represents an integral part of the etiology and pathogenesis of T1D and CD. An extended basis of knowledge has grown on the genetic factors that influence the development of both diseases. However, the genetics of T1D and CD are complex and polygenic. Besides HLA genetic factors directly associated with controlling immune response and beta-cell function, there is mounting proof that other missing hereditary may be involved.

A meaningful application of genetics is to develop prediction in approaches that can be designed and achieved to counter disease in individuals at risk will be directed towards establishing:

I. Extensive records of genotype-phenotype relationships in different populations and environments;
II. Systematic assays of the variant- and gene-level function.
III. Enhanced scalable strategies for using this basic genetic knowledge into completely developed molecular genetics models of T1D and CD pathogenesis; and
IV. Application of genetic insights to stimulate new preventative and therapeutic alternatives.

These advantages obtained by magnificent sequencing technology and collaborative research group outcomes will improve the understanding of the genetic basis of human disease. Data will drive translational innovations and widen access to large-scale genetic resources.

Acknowledgments

The work of this thesis comes not only from one student research work. It is the result of contribution and support from the kind and significant people around me.

I want to express my heartfelt gratefulness to my supervisor Dr. **Daniel Agardh**, who has the character and the essence of a real passionate scientist pediatrician. He continually and convincingly conveyed a spirit of adventure concerning research and scholarship and excitement around academics. Without his guidance and determined help, this dissertation would not have been possible. One of the best qualities he has, is excellent work intension, always available, dedicated to supporting and guiding my doctorate studies. I am hugely blessed for your generous guidance and support by you and from your family. Thank you.

My immense gratitude and thanks are for my co-supervisor, Professor **Åke Lernmark**, the celebrated teacher of many science generations and well-respected figures in the academic research. An early meeting with him introduced me to be a research student and be part of his fantastic research group. Then a success journey.

Special thanks to my co-supervisor, Dr. **Helena Elding Larsson**, for her precious contribution to my research effort and supervision during her clinical practice and care of my manuscript writing.

Colleagues of my studies deserve the best of my recognition for their productive discussions and guidance. Alexander Lind has a massive contribution to my expertise in writing, overseeing, discussing, and incredible work ethics. Gratitude to my doctorate colleagues: Falastin Salami, Agnes Svärd. You were the best supporters and fans during all the research years. All of you have the best praise.

This work's contributions are extended to my lab colleagues: Anita, Charlotte, Ida, Linda, Rasmus, and Zeliha. Thank you for all companionship and support for my lab work. Special thanks to Thomas for his excellent and swift support for my studies. A particular greetings to my previous colleagues in the lab: Omar, Ariana, Josefin, Samia.

I would thank all the collaborators in my research work, including:

Marlena Maziarz, thank you for your statistical assistance and for teaching me the technical advice.

Lue Ping Zhao for his important statistical analysis and genetic interest

Daniel E Geraghty and the NGS team at Scisco Genetics labs and Fred Hutchinson institute in Seattle, USA

My thanks extend to co-authors and every contributor behind the scene, nurses, biomedicine analysts, administrators, and journal reviewers.

I would like to thank my committee members, the main opponent Professor Riccardo Troncone and the examination committee for their work and giving effort in assessing my work.

And not least, to all type 1 diabetes and celiac patients and families, who donated and contributed to this work

My thanks are not confined to who just contributed to my study works.
I have massive gratefulness to my family; My Father and Mother, you are my beautiful heart. Thank you for raising me and teaching me to attain this degree. My brothers and sister. Thank you for your graceful feelings and support.
Sincere gratitude to all my friends around me, from my country and around the world. Dr. **Mohammad Abdul Razaq** has a special appreciation for accompanying me during our doctorate studies. We made it!

Finally, I dedicate this work for the sake of Allah; I hope that he accepts me in his mercy.

References

1. Schlosser M, Mueller P, Törn C, Bonifacio E, Bingley P. Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies. Diabetologia. 2010;53(12):2611-20.
2. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. The Lancet. 2014;383(9911):69-82.
3. Association AD. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Diabetes care. 2020;43(Suppl 1):S14.
4. Abadie V, Sollid LM, Barreiro LB, Jabri B. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. Annual review of immunology. 2011;29:493-525.
5. Lindfors K, Ciacci C, Kurppa K, Lundin KEA, Makharia GK, Mearin ML, et al. Coeliac disease. Nature Reviews Disease Primers. 2019;5(1):3.
6. Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. Concordance for islet autoimmunity among monozygotic twins. New England Journal of Medicine. 2008;359(26):2849-50.
7. Nisticò L, Fagnani C, Coto I, Percopo S, Cotichini R, Limongelli MG, et al. Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut. 2006;55(6):803-8.
8. Steck AK, Barriga KJ, Emery LM, Fiallo-Scharer RV, Gottlieb PA, Rewers MJ. Secondary attack rate of type 1 diabetes in Colorado families. Diabetes care. 2005;28(2):296-300.
9. Singh P, Arora S, Lal S, Strand TA, Makharia GK. Risk of celiac disease in the first- and second-degree relatives of patients with celiac disease: a systematic review and meta-analysis. American Journal of Gastroenterology. 2015;110(11):1539-48.
10. Wessels MMS, de Rooij N, Roovers L, Verhage J, de Vries W, Mearin ML. Towards an individual screening strategy for first-degree relatives of celiac patients. Eur J Pediatr. 2018;177(11):1585-92.
11. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, et al. The first large population based twin study of coeliac disease. Gut. 2002;50(5):624-8.
12. Sollid LM, Lie BA. Celiac disease genetics: current concepts and practical applications. Clinical Gastroenterology and Hepatology. 2005;3(9):843-51.
13. Pociot F, McDermott M. Genetics of type 1 diabetes mellitus. Genes & Immunity. 2002;3(5):235-49.
14. Amiel J. Study of leucocyte phenotypes in Hodgkins' disease. Histocompatibility testing 1967. 1967:79-81.

15. Mcdevitt H, Bodmer W. HL-A, immune-response genes, and disease. The Lancet. 1974;303(7869):1269-75.
16. Schlosstein L, Terasaki PI, Bluestone R, Pearson CM. High association of an HL-A antigen, W27, with ankylosing spondylitis. New England Journal of Medicine. 1973;288(14):704-6.
17. Nerup J. Platz R Andersen OO et al.(1974) HL-A antigens and diabetes mellitus. Lancet II.864-6.
18. Singal DP, Blajchman MA. Histocompatibility (HL-A) Antigens, Lymphocytotoxic Antibodies and Tissue Antibodies in Patients with Diabetes Mellitus. Diabetes. 1973;22(6):429-32.
19. Falchuk ZM, Rogentine GN, Strober W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. The Journal of Clinical Investigation. 1972;51(6):1602-5.
20. Stokes P, Holmes G, Asquith P, Mackintosh P, Cooke W. Histocompatibility antigens associated with adult coeliac disease. The Lancet. 1972;300(7769):162-4.
21. Falchuk ZM, Strober W. HL-A antigens and adult coeliac disease. Lancet. 1972;2(7790):1310.
22. Höhler T, Gerken G, Notghi A, Lubjuhn R, Taheri H, Protzer U, et al. HLA-DRB1* 1301 and* 1302 protect against chronic hepatitis B. Journal of hepatology. 1997;26(3):503-7.
23. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. American journal of human genetics. 1996;59(5):1134.
24. Bjorkman PJ, Saper M, Samraoui B, Bennett WS, Strominger JL, Wiley D. Structure of the human class I histocompatibility antigen, HLA-A2. Nature. 1987;329(6139):506-12.
25. Saper M, Bjorkman P, Wiley D. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. Journal of molecular biology. 1991;219(2):277-319.
26. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. Journal of human genetics. 2009;54(1):15-39.
27. Slatkin M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics. 2008;9(6):477-85.
28. Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, et al. Complete MHC haplotype sequencing for common disease gene mapping. Genome Res. 2004;14(6):1176-87.
29. Ceppellini R, Curtoni E, Mattiuz P, Miggiano V, Scudeller G, Serra A. Genetics of leukocyte antigens: a family study of segregation and linkage. Histocompatibility testing. 1967;1967:149.
30. Vina MAF, Hollenbach JA, Lyke KE, Sztein MB, Maiers M, Klitz W, et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and

haplotypes in closed and open populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367(1590):820-9.

31. Mungall AJ, Palmer S, Sims S, Edwards C, Ashurst J, Wilming L, et al. The DNA sequence and analysis of human chromosome 6. Nature. 2003;425(6960):805-11.
32. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annual review of genomics and human genetics. 2013;14:301-23.
33. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SG. IPD-IMGT/HLA Database. Nucleic acids research. 2020;48(D1):D948-D55.
34. Gonzalez-Galarza Faviel F, McCabe A, Santos Eduardo J Md, Jones J, Takeshita L, Ortega-Rivera Nestor D, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Research. 2019;48(D1):D783-D8.
35. Noble JA. Immunogenetics of type 1 diabetes: a comprehensive review. Journal of autoimmunity. 2015;64:101-12.
36. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature. 1999;401(6756):921-3.
37. Grumet F, Fish L, Moossazadeh J, Ness D, Duceman B. An HLA-B locus probe clarifies endonuclease polymorphism of major histocompatibility complex class I genes. Molecular Biology & Medicine. 1983;1(5):501-9.
38. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature. 1986;324(6093):163-6.
39. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue antigens. 1992;39(5):225-35.
40. Santamaria P, Boyce-Jacino MT, Lindstrom AL, Barbosa JJ, Faras AJ, Rich SS. HLA class II "typing": direct sequencing of DRB, DQB, and DQA genes. Human immunology. 1992;33(2):69-81.
41. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. Nature. 2017;550(7676):345-53.
42. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nature Reviews Genetics. 2020.
43. Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg E, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue antigens. 2009;74(5):393-403.
44. Bentley G, Erlich H, Higuchi R. High resolution, high throughput hla genotyping by clonal sequencing. Google Patents; 2010.
45. Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Human immunology. 2010;71(10):1033-42.

46. Duke J, Lind C, Mackiewicz K, Ferriola D, Papazoglou A, Gasiewski A, et al. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. Hla. 2016;87(3):141-52.
47. Profaizer T, Lázár-Molnár E, Close D, Delgado J, Kumánovics A. HLA genotyping in the clinical laboratory: comparison of next-generation sequencing methods. Hla. 2016;88(1-2):14-24.
48. Monos D, Maiers MJ. Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): Consequences, opportunities and challenges. Human immunology. 2015;76(12):883-6.
49. Shiina T, Suzuki S, Ozaki Y, Taira H, Kikkawa E, Shigenari A, et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. Tissue antigens. 2012;80(4):305-16.
50. Erlich H. HLA DNA typing: past, present, and future. Tissue antigens. 2012;80(1):1-11.
51. Erlich HA. HLA typing using next generation sequencing: An overview. Hum Immunol. 2015;76(12):887-90.
52. Holcomb CL, Höglund B, Anderson MW, Blake LA, Böhme I, Egholm M, et al. A multi-site study using high-resolution HLA genotyping by next generation sequencing. Tissue Antigens. 2011;77(3):206-17.
53. Smith AG, Pyo CW, Nelson W, Gow E, Wang R, Shen S, et al. Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors. Hum Immunol. 2014;75(10):1040-6.
54. Nelson WC, Pyo CW, Vogan D, Wang R, Pyon YS, Hennessey C, et al. An integrated genotyping approach for HLA and other complex genetic systems. Hum Immunol. 2015;76(12):928-38.
55. Zhao LP, Papadopoulos GK, Kwok WW, Moustakas AK, Bondinas GP, Larsson HE, et al. Motifs of three HLA-DQ amino acid residues (α44, β57, β135) capture full association with the risk of type 1 diabetes in DQ2 and DQ8 children. Diabetes. 2020.
56. Hermann R, Knip M, Veijola R, Simell O, Laine AP, Akerblom HK, et al. Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes--indication of an increased environmental pressure? Diabetologia. 2003;46(3):420-5.
57. Vlachopoulou E, Lahtela E, Wennerström A, Havulinna A, Salo P, Perola M, et al. Evaluation of HLA-DRB1 imputation using a Finnish dataset. Tissue Antigens. 2014;83(5):350-5.
58. Carapito R, Radosavljevic M, Bahram S. Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. Hum Immunol. 2016.
59. Milius RP, Mack SJ, Hollenbach JA, Pollack J, Heuer ML, Gragert L, et al. Genotype List String: a grammar for describing HLA and KIR genotyping results in a text string. Tissue Antigens. 2013;82(2):106-12.
60. Katsarou A, Gudbjörnsdottir S, Rawshani A, Dabelea D, Bonifacio E, Anderson BJ, et al. Type 1 diabetes mellitus. Nature Reviews Disease Primers. 2017;3(1):17016.

61. Aly TA, Eller E, Ide A, Gowan K, Babu SR, Erlich HA, et al. Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype. Diabetes. 2006;55(5):1265-9.
62. Cudworth A, Woodrow J. HL-A antigens and diabetes mellitus. The Lancet. 1974;304(7889):1153.
63. Barbosa J, Chern MM, Anderson VE, Noreen H, Johnson S, Reinsmoen N, et al. Linkage analysis between the major histocompatibility system and insulin-dependent diabetes in families with patients in two consecutive generations. The Journal of clinical investigation. 1980;65(3):592-601.
64. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661.
65. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nature genetics. 2007;39(7):857-64.
66. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature genetics. 2009;41(6):703-7.
67. Nejentsev S, Howson JMM, Walker NM, Szeszko J, Field SF, Stevens HE, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. 2007;450(7171):887-U19.
68. Howson JM, Walker N, Clayton D, Todd J, Consortium DG. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes, Obesity and Metabolism. 2009;11:31-45.
69. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7(9):e1002293.
70. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, et al. A genome-wide search for human type 1 diabetes susceptibility genes. Nature. 1994;371(6493):130-6.
71. Noble JA, Erlich HA. Genetics of type 1 diabetes. Cold Spring Harbor perspectives in medicine. 2012;2(1):a007732.
72. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen W-M, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nature genetics. 2015;47(8):898.
73. Concannon P, Chen W-M, Julier C, Morahan G, Akolkar B, Erlich HA, et al. Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the Type 1 Diabetes Genetics Consortium. Diabetes. 2009;58(4):1018-22.
74. Polychronakos C, Li Q. Understanding type 1 diabetes through genetics: advances and prospects. Nature Reviews Genetics. 2011;12(11):781-92.
75. Todd JA. Etiology of type 1 diabetes. Immunity. 2010;32(4):457-67.

76. Aly TA, Ide A, Jahromi MM, Barker JM, Fernando MS, Babu SR, et al. Extreme genetic risk for type 1A diabetes. Proceedings of the National Academy of Sciences. 2006;103(38):14074-9.
77. Rich SS, Concannon P, Erlich H, Julier C, Morahan G, Nerup J, et al. The type 1 diabetes genetics consortium. Annals of the New York Academy of Sciences. 2006;1079(1):1-8.
78. Cooper J, Howson J, Smyth D, Walker N, Stevens H, Yang J, et al. Confirmation of novel type 1 diabetes risk loci in families. Diabetologia. 2012;55(4):996-1000.
79. Erlich HA, Valdes AM, McDevitt SL, Simen BB, Blake LA, McGowan KR, et al. Next generation sequencing reveals the association of DRB3*02:02 with type 1 diabetes. Diabetes. 2013;62(7):2618-22.
80. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2008.
81. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PloS one. 2013;8(6):e64683.
82. Graham J, Kockum I, Sanjeevi CB, Landin-Olsson M, Nyström L, Sundkvist G, et al. Negative association between type 1 diabetes and HLA DQB1* 0602-DQA1* 0102 is attenuated with age at onset. European Journal of Immunogenetics. 1999;26(2):117-27.
83. Kockum I, Sanjeevi CB, Eastman S, Landin-Olsson M, Dahlquist G, Lernmark A. Complex interaction between HLA DR and DQ in conferring risk for childhood type 1 diabetes. Eur J Immunogenet. 1999;26(5):361-72.
84. Concannon P, Erlich HA, Julier C, Morahan G, Nerup J, Pociot F, et al. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. Diabetes. 2005;54(10):2995-3001.
85. Valdes AM, Thomson G, Erlich HA, Noble JA. Association between type 1 diabetes age of onset and HLA among sibling pairs. Diabetes. 1999;48(8):1658-61.
86. She J-X. Susceptibility to type I diabetes: HLA-DQ and DR revisited. Immunology today. 1996;17(7):323-9.
87. Carlsson A, Kockum I, Lindblad B, Engleson L, Nilsson A, Forsander G, et al. Low risk HLA-DQ and increased body mass index in newly diagnosed type 1 diabetes children in the Better Diabetes Diagnosis study in Sweden. International journal of obesity. 2012;36(5):718-24.
88. Andersson C, Larsson K, Vaziri-Sani F, Lynch K, Carlsson A, Cedervall E, et al. The three ZNT8 autoantibody variants together improve the diagnostic sensitivity of childhood and adolescent type 1 diabetes. Autoimmunity. 2011;44(5):394-405.
89. Kiviniemi M, Hermann R, Nurmi J, Ziegler AG, Knip M, Simell O, et al. A high-throughput population screening system for the estimation of genetic risk for type 1 diabetes: an application for the TEDDY (the Environmental Determinants of Diabetes in the Young) study. Diabetes technology & therapeutics. 2007;9(5):460-72.

90. Hagopian WA, Erlich H, Lernmark Å, Rewers M, Ziegler AG, Simell O, et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatric diabetes. 2011;12(8):733-43.
91. Texier C, Pouvelle-Moratille S, Busson M, Charron D, Ménez A, Maillere B. Complementarity and redundancy of the binding specificity of HLA-DRB1,-DRB3,-DRB4 and-DRB5 molecules. European journal of immunology. 2001;31(6):1837-46.
92. Holdsworth R, Hurley C, Marsh S, Lau M, Noreen H, Kempenich J, et al. The HLA dictionary 2008: a summary of HLA-A,-B,-C,-DRB1/3/4/5, and-DQB1 alleles and their association with serologically defined HLA-A,-B,-C,-DR, and-DQ antigens. Tissue antigens. 2009;73(2):95-170.
93. Delli AJ, Lindblad B, Carlsson A, Forsander G, Ivarsson SA, Ludvigsson J, et al. Type 1 diabetes patients born to immigrants to Sweden increase their native diabetes risk and differ from Swedish patients in HLA types and islet autoantibodies. Pediatric diabetes. 2010;11(8):513-20.
94. Andersson C, Kolmodin M, Ivarsson SA, Carlsson A, Forsander G, Lindblad B, et al. Islet cell antibodies (ICA) identify autoimmunity in children with new onset diabetes mellitus negative for other islet cell antibodies. Pediatric diabetes. 2014;15(5):336-44.
95. Gyllenberg A, Asad S, Piehl F, Swanberg M, Padyukov L, Van Yserloo B, et al. Age-dependent variation of genotypes in MHC II transactivator gene (CIITA) in controls and association to type 1 diabetes. Genes and immunity. 2012;13(8):632-40.
96. Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, Lowe RM, et al. Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. Diabetes. 2002;51(5):1346-55.
97. Dahlquist G, Blom L, Holmgren G, Hägglöf B, Larsson Y, Sterky G, et al. The epidemiology of diabetes in Swedish children 0–14 years—a six-year prospective study. Diabetologia. 1985;28(11):802-8.
98. Parkkola A, Härkönen T, Ryhänen SJ, Ilonen J, Knip M, Register FPD. Extended family history of type 1 diabetes and phenotype and genotype of newly diagnosed children. Diabetes care. 2013;36(2):348-54.
99. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. Curr Diab Rep. 2011;11(6):533-42.
100. Varney MD, Valdes AM, Carlson JA, Noble JA, Tait BD, Bonella P, et al. HLA DPA1, DPB1 alleles and haplotypes contribute to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2010;59(8):2055-62.
101. Baschal EE, Aly TA, Babu SR, Fernando MS, Yu L, Miao D, et al. HLA-DPB1* 0402 protects against type 1A diabetes autoimmunity in the highest risk DR3-DQB1* 0201/DR4-DQB1* 0302 DAISY population. Diabetes. 2007;56(9):2405-9.
102. Cruz TD, Valdes AM, Santiago A, de Llado TF, Raffel LJ, Zeidler A, et al. DPB1 alleles are associated with type 1 diabetes susceptibility in multiple ethnic groups. diabetes. 2004;53(8):2158-63.

103. Eike MC, Olsson M, Undlien D, Dahl-Jørgensen K, Joner G, Rønningen K, et al. Genetic variants of the HLA-A, HLA-B and AIF1 loci show independent associations with type 1 diabetes in Norwegian families. Genes & Immunity. 2009;10(2):141-50.
104. Nakanishi K, Kobayashi T, Murase T, Nakatsuji T, Inoko H, Tsuji K, et al. Association of HLA-A24 with complete β-cell destruction in IDDM. Diabetes. 1993;42(7):1086-93.
105. Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. 2007;450(7171):887-92.
106. Valdes AM, Erlich HA, Noble JA. Human leukocyte antigen class IB and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. Human immunology. 2005;66(3):301-13.
107. Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, Fear AL, et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes. 2010;59(11):2972-9.
108. Noble JA, Valdes AM, Bugawan TL, Apple RJ, Thomson G, Erlich HA. The HLA class IA locus affects susceptibility to type 1 diabetes. Human immunology. 2002;63(8):657-64.
109. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. Diabetes care. 2019;42(2):200-7.
110. Baschal E, Baker P, Eyring K, Siebert J, Jasinski J, Eisenbarth G. The HLA-B* 3906 allele imparts a high risk of diabetes only on specific HLA-DR/DQ haplotypes. Diabetologia. 2011;54(7):1702-9.
111. Mbunwe E, Van der Auwera BJ, Weets I, Van Crombrugge P, Crenier L, Coeckelberghs M, et al. In antibody-positive first-degree relatives of patients with type 1 diabetes, HLA-A*24 and HLA-B*18, but not HLA-B*39, are predictors of impending diabetes with distinct HLA-DQ interactions. Diabetologia. 2013;56(9):1964-70.
112. Mikk M-L, Kiviniemi M, Laine A-P, Härkönen T, Veijola R, Simell O, et al. The HLA-B∗ 39 allele increases type 1 diabetes risk conferred by HLA-DRB1∗ 04: 04-DQB1∗ 03: 02 and HLA-DRB1∗ 08-DQB1∗ 04 class II haplotypes. Human Immunology. 2014;75(1):65-70.
113. Roark CL, Anderson KM, Simon LJ, Schuyler RP, Aubrey MT, Freed BM. Multiple HLA epitopes contribute to type 1 diabetes susceptibility. Diabetes. 2014;63(1):323-31.
114. Valdes AM, Erlich HA, Carlson J, Varney M, Moonsamy PV, Noble JA. Use of class I and class II HLA loci for predicting age at onset of type 1 diabetes in multiple populations. Diabetologia. 2012;55(9):2394-401.
115. Krischer JP, Lynch KF, Schatz DA, Ilonen J, Lernmark Å, Hagopian WA, et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. Diabetologia. 2015;58(5):980-7.

116. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA. 2013;309(23):2473-9.
117. Törn C, Hadley D, Lee H-S, Hagopian W, Lernmark Å, Simell O, et al. Role of type 1 diabetes–associated SNPs on risk of autoantibody positivity in the TEDDY Study. Diabetes. 2015;64(5):1818-29.
118. Wester A, Skärstrand H, Lind A, Ramelius A, Carlsson A, Cedervall E, et al. An increased diagnostic sensitivity of truncated GAD65 autoantibodies in type 1 diabetes may be related to HLA-DQ8. Diabetes. 2017;66(3):735-40.
119. Howson JM, Rosinger S, Smyth DJ, Boehm BO, Todd JA, group A-Es. Genetic analysis of adult-onset autoimmune diabetes. Diabetes. 2011;60(10):2645-53.
120. Hagopian WA, Sanjeevi CB, Kockum I, Landin-Olsson M, Karlsen AE, Sundkvist G, et al. Glutamate decarboxylase-, insulin-, and islet cell-antibodies and HLA typing to detect diabetes in a general population-based study of Swedish children. The journal of clinical investigation. 1995;95(4):1505-11.
121. Vandewalle CL, Falorni A, Lernmark Å, Goubert P, Dorchy H, Coucke W, et al. Associations of GAD65-and IA-2-autoantibodies with genetic risk markers in new-onset IDDM patients and their siblings. The Belgian Diabetes Registry. Diabetes care. 1997;20(10):1547-52.
122. Kulmala P, Rahko J, Savola K, Vähäsalo P, Sjöroos M, Reunanen A, et al. Beta-cell autoimmunity, genetic susceptibility, and progression to type 1 diabetes in unaffected schoolchildren. Diabetes care. 2001;24(1):171-3.
123. Gullstrand C, Wahlberg J, Ilonen J, Vaarala O, Ludvigsson J. Progression to type 1 diabetes and autoantibody positivity in relation to HLA-risk genotypes in children participating in the ABIS study. Pediatr Diabetes. 2008;9(3 Pt 1):182-90.
124. Diaz-Valencia PA, Bougnères P, Valleron A-J. Global epidemiology of type 1 diabetes in young adults and adults: a systematic review. BMC public health. 2015;15(1):255.
125. Askar M, Daghstani J, Thomas D, Leahy N, Dunn P, Claas F, et al. 16th IHIW: Global distribution of extended HLA haplotypes. International journal of immunogenetics. 2013;40(1):31-8.
126. Hurley CK, Kempenich J, Wadsworth K, Sauter J, Hofmann JA, Schefzyk D, et al. Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0. 0. Hla. 2020;95(6):516-31.
127. Ilonen J, Sjöroos M, Knip M, Veijola R, Simell O, Åkerblom HK, et al. Estimation of genetic risk for type 1 diabetes. American journal of medical genetics. 2002;115(1):30-6.
128. Hermann R, Bartsocas C, Soltész G, Vazeou A, Paschou P, Bozas E, et al. Genetic screening for individuals at high risk for type 1 diabetes in the general population using HLA Class II alleles as disease markers. A comparison between three European populations with variable rates of disease incidence. Diabetes/metabolism research and reviews. 2004;20(4):322-9.

129. Kockum I, Lernmark Å, Dahlquist G, Falorni A, Hagopian W, Landin-Olsson M, et al. Genetic and immunological findings in patients with newly diagnosed insulin-dependent diabetes mellitus. Hormone and metabolic research. 1996;28(07):344-7.
130. Zhang X-m, Wang H-y, Luo Y-y, Ji L-n. HLA-DQ, DR allele polymorphism of type 1 diabetes in the Chinese population: a meta-analysis. LWW; 2009.
131. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Human immunology. 2013;74(10):1313-20.
132. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature genetics. 2009;41(6):703.
133. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nature genetics. 2015;47(4):381-6.
134. Bakay M, Pandey R, Grant SFA, Hakonarson H. The Genetic Contribution to Type 1 Diabetes. Curr Diabetes Rep. 2019;19(11):116.
135. Rewers M, Bugawan T, Norris J, Blair A, Beaty B, Hoffman M, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). Diabetologia. 1996;39(7):807.
136. Mahon JL, Sosenko JM, Rafkin-Mervis L, Krause-Steinrauf H, Lachin JM, Thompson C, et al. The TrialNet Natural History Study of the Development of Type 1 Diabetes: objectives, design, and initial results. Pediatric diabetes. 2009;10(2):97-104.
137. Thomas HR, Gitelman SE. Altering the course of type 1 diabetes: an update on prevention and new onset clinical trials. Pediatric diabetes. 2013;14(5):311.
138. Valdes AM, Varney MD, Erlich HA, Noble JA. Receiver operating characteristic analysis of HLA, CTLA4, and insulin genotypes for type 1 diabetes. Diabetes care. 2013;36(9):2504-7.
139. Ferrat LA, Vehik K, Sharp SA, Lernmark Å, Rewers MJ, She J-X, et al. A combined risk score enhances prediction of type 1 diabetes among susceptible children. Nature Medicine. 2020:1-9.
140. Ludvigsson JF, Green PH. Clinical management of coeliac disease. J Intern Med. 2011;269(6):560-71.
141. Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Riecken EO, et al. Identification of tissue transglutaminase as the autoantigen of celiac disease. Nat Med. 1997;3(7):797-801.
142. Rubio-Tapia A, Hill ID, Kelly CP, Calderwood AH, Murray JA. ACG clinical guidelines: diagnosis and management of celiac disease. Am J Gastroenterol. 2013;108(5):656-76; quiz 77.
143. Husby S, Koletzko S, Korponay-Szabó I, Mearin M, Phillips A, Shamir R, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. Journal of pediatric gastroenterology and nutrition. 2012;54(1):136-60.

144. Dicke W-K, Weijers H, KAMER JvD. Coeliac Disease The Presence in Wheat of a Factor Having a Deleterious Effect in Cases of Coeliac Disease. Acta Paediatrica. 1953;42(1):34-42.
145. Macdonald WC, Dobbins WO, 3rd, Rubin CE. STUDIES OF THE FAMILIAL NATURE OF CELIAC SPRUE USING BIOPSY OF THE SMALL INTESTINE. N Engl J Med. 1965;272:448-56.
146. Brett P, Yiannakou J, Morris MA, Vaughan R, Curtis D, Ciclitira P. Common HLA alleles, rather than rare mutants, confer susceptibility to coeliac disease. Annals of human genetics. 1999;63(3):217-25.
147. Partanen J. Major histocompatibility complex and coeliac disease. Coeliac Disease M Maki, P Collin, J Visakorpi (eds) Tampere, Coeliac Disease Study Group. 1997:253-64.
148. Molberg Ø, Mcadam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. Nature medicine. 1998;4(6):713-7.
149. Bodd M, Kim CY, Lundin KE, Sollid LM. T-cell response to gluten in patients with HLA-DQ2. 2 reveals requirement of peptide-MHC stability in celiac disease. Gastroenterology. 2012;142(3):552-61.
150. Fallang L-E, Bergseng E, Hotta K, Berg-Larsen A, Kim C-Y, Sollid LM. Differences in the risk of celiac disease associated with HLA-DQ2. 5 or HLA-DQ2. 2 are related to sustained gluten antigen presentation. Nature immunology. 2009;10(10):1096-101.
151. Hovhannisyan Z, Weiss A, Martin A, Wiesner M, Tollefsen S, Yoshida K, et al. The role of HLA-DQ8 β57 polymorphism in the anti-gluten T-cell response in coeliac disease. Nature. 2008;456(7221):534-8.
152. Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, Benckhuijsen W, et al. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology. 2002;122(7):1729-37.
153. Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. Annals of human genetics. 1997;61(4):307-17.
154. Bevan S, Popat S, Braegger C, Busch A, O'Donoghue D, Falth-Magnusson K, et al. Contribution of the MHC region to the familial risk of coeliac disease. Journal of medical genetics. 1999;36(9):687-90.
155. Kuja-Halkola R, Lebwohl B, Halfvarson J, Wijmenga C, Magnusson PK, Ludvigsson JF. Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins. Gut. 2016.
156. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. The Journal of experimental medicine. 1989;169(1):345-50.
157. Romanos J, Rosen A, Kumar V, Trynka G, Franke L, Szperl A, et al. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. Gut. 2014;63(3):415-22.

158. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nature genetics. 2008;40(4):395-402.
159. Djilali-Saiah I, Caillat-Zucman S, Schmitz J, Chaves-Vieira ML, Bach JF. Polymorphism of antigen processing (TAP, LMP) and HLA class II genes in celiac disease. Hum Immunol. 1994;40(1):8-16.
160. Hadley D, Hagopian W, Liu E, She JX, Simell O, Akolkar B, et al. HLA-DPB1*04:01 Protects Genetically Susceptible Children from Celiac Disease Autoimmunity in the TEDDY Study. Am J Gastroenterol. 2015;110(6):915-20.
161. Paunić V, Gragert L, Madbouly A, Freeman J, Maiers M. Measuring ambiguity in HLA typing methods. PLoS One. 2012;7(8):e43585.
162. Ludvigsson JF, Brandt L, Montgomery SM, Granath F, Ekbom A. Validation study of villous atrophy and small intestinal inflammation in Swedish biopsy registers. BMC Gastroenterol. 2009;9:19.
163. Harris RA, Sugimoto K, Kaplan DE, Ikeda F, Kamoun M, Chang KM. Human leukocyte antigen class II associations with hepatitis C virus clearance and virus-specific CD4 T cell response among Caucasians and African Americans. Hepatology. 2008;48(1):70-9.
164. Megiorni F, Mora B, Bonamico M, Barbato M, Nenna R, Maiella G, et al. HLA-DQ and risk gradient for celiac disease. Hum Immunol. 2009;70(1):55-9.
165. Tosi R, Vismara D, Tanigaki N, Ferrara GB, Cicimarra F, Buffolano W, et al. Evidence that celiac disease is primarily associated with a DC locus allelic specificity. Clinical immunology and immunopathology. 1983;28(3):395-404.
166. Keuning J, Pena A, Van Hooff J, Van Leeuwen A, Van Rood J. HLA-DW3 associated with coeliac disease. The Lancet. 1976;307(7958):506-8.
167. Bugawan TL, Angelini G, Larrick J, Auricchio S, Ferrara GB, Erlich HA. A combination of a particular HLA-DPβ allele and an HLA-DQ heterodimer confers susceptibility to coeliac disease. Nature. 1989;339(6224):470-3.
168. Hadley D, Hagopian W, Liu E, She J-X, Simell O, Akolkar B, et al. HLA-DPB1* 04: 01 protects genetically susceptible children from celiac disease autoimmunity in the TEDDY study. The American journal of gastroenterology. 2015;110(6):915.
169. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, 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. Human immunology. 2003;64(4):469-77.
170. Mubarak A, Spierings E, Wolters V, van Hoogstraten I, Kneepkens CM, Houwen R. Human leukocyte antigen DQ2.2 and celiac disease. Journal of pediatric gastroenterology and nutrition. 2013;56(4):428-30.
171. Liu E, Lee HS, Aronsson CA, Hagopian WA, Koletzko S, Rewers MJ, et al. Risk of pediatric celiac disease according to HLA haplotype and country. N Engl J Med. 2014;371(1):42-9.
172. Karinen H, Kärkkäinen P, Pihlajamäki J, Janatuinen E, Heikkinen M, Julkunen R, et al. Gene dose effect of the DQB1*0201 allele contributes to severity of coeliac disease. Scandinavian journal of gastroenterology. 2006;41(2):191-9.

173. Biagi F, Bianchi PI, Vattiato C, Marchese A, Trotta L, Badulli C, et al. Influence of HLA-DQ2 and DQ8 on severity in celiac Disease. J Clin Gastroenterol. 2012;46(1):46-50.
174. Jores RD, Frau F, Cucca F, Grazia Clemente M, Orrù S, Rais M, et al. HLA-DQB1*0201 homozygosis predisposes to severe intestinal damage in celiac disease. Scandinavian journal of gastroenterology. 2007;42(1):48-53.
175. Karinen H, Kärkkäinen P, Pihlajamäki J, Janatuinen E, Heikkinen M, Julkunen R, et al. Gene dose effect of the DQB1* 0201 allele contributes to severity of coeliac disease. Scandinavian journal of gastroenterology. 2006;41(2):191-9.
176. Al–Toma A, Goerres MS, Meijer JW, Peña AS, Crusius JBA, Mulder CJ. Human leukocyte antigen–DQ2 homozygosity and the development of refractory celiac disease and enteropathy-associated T-cell lymphoma. Clinical Gastroenterology and Hepatology. 2006;4(3):315-9.
177. Pietzak MM, Schofield TC, McGinniss MJ, Nakamura RM. Stratifying risk for celiac disease in a large at-risk United States population by using HLA alleles. Clinical Gastroenterology and Hepatology. 2009;7(9):966-71.
178. Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E. HLA-DR and-DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. Human immunology. 1992;35(3):188-92.
179. Fallang LE, Bergseng E, Hotta K, Berg-Larsen A, Kim CY, Sollid LM. Differences in the risk of celiac disease associated with HLA-DQ2.5 or HLA-DQ2.2 are related to sustained gluten antigen presentation. Nat Immunol. 2009;10(10):1096-101.
180. Bodd M, Kim CY, Lundin KE, Sollid LM. T-cell response to gluten in patients with HLA-DQ2.2 reveals requirement of peptide-MHC stability in celiac disease. Gastroenterology. 2012;142(3):552-61.
181. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, 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):469-77.
182. Gutierrez-Achury J, Zhernakova A, Pulit SL, Trynka G, Hunt KA, Romanos J, et al. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nature genetics. 2015;47(6):577-8.
183. Medrano LM, Dema B, López-Larios A, Maluenda C, Bodas A, López-Palacios N, et al. HLA and celiac disease susceptibility: new genetic factors bring open questions about the HLA influence and gene-dosage effects. PloS one. 2012;7(10):e48403.
184. Picascia S, Sidney J, Camarca A, Mazzarella G, Giardullo N, Greco L, et al. Gliadin-Specific CD8(+) T Cell Responses Restricted by HLA Class I A*0101 and B*0801 Molecules in Celiac Disease Patients. J Immunol. 2017;198(5):1838-45.
185. Tye-Din JA, Galipeau HJ, Agardh D. Celiac Disease: A Review of Current Concepts in Pathogenesis, Prevention, and Novel Therapies. Frontiers in Pediatrics. 2018;6(350).
186. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. Nature genetics. 2010;42(4):295-302.

187. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nature genetics. 2007;39(7):827-9.
188. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nature genetics. 2011;43(12):1193-201.
189. Sharma A, Liu X, Hadley D, Hagopian W, Liu E, Chen W-M, et al. Identification of Non-HLA Genes Associated with Celiac Disease and Country-Specific Differences in a Large, International Pediatric Cohort. PLoS One. 2016;11(3):e0152476-Article No.: e.
190. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008;359(26):2767-77.
191. Zhernakova A, Van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature Reviews Genetics. 2009;10(1):43-55.
192. Lundin KE, Wijmenga C. Coeliac disease and autoimmune disease—genetic overlap and screening. Nature reviews Gastroenterology & hepatology. 2015;12(9):507.
193. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet. 2011;7(1):e1001283.
194. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, Luo X, Bhagat G, Green PH, et al. A long noncoding RNA associated with susceptibility to celiac disease. Science. 2016;352(6281):91-5.
195. Aronsson CA, Lee H-S, af Segerstad EMH, Uusitalo U, Yang J, Koletzko S, et al. Association of gluten intake during the first 5 years of life with incidence of celiac disease autoimmunity and celiac disease among children at increased risk. Jama. 2019;322(6):514-23.
196. Lindfors K, Lin J, Lee H-S, Hyöty H, Nykter M, Kurppa K, et al. Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study. Gut. 2020;69(8):1416-22.
197. Vidales MC, Zubillaga P, Zubillaga I, Alfonso-Sanchez MA. Allele and haplotype frequencies for HLA class II (DQA1 and DQB1) loci in patients with celiac disease from Spain. Human immunology. 2004;65(4):352-8.
198. Kaur G, Sarkar N, Bhatnagar S, Kumar S, Rapthap C, Bhan M, et al. Pediatric celiac disease in India is associated with multiple DR3-DQ2 haplotypes. Human immunology. 2002;63(8):677-82.
199. Basturk A, Artan R, Yilmaz A. The incidence of HLA-DQ2/DQ8 in Turkish children with celiac disease and a comparison of the geographical distribution of HLA-DQ. Prz Gastroenterol. 2017;12(4):256-61.

200. Björck S, Brundin C, Lörinc E, Lynch KF, Agardh D. Screening Detects a High Proportion of Celiac Disease in Young HLA-genotyped Children. Journal of pediatric gastroenterology and nutrition. 2010;50(1):49-53.
201. Perez-Bravo F, Araya M, Mondragon A, Ríos G, Alarcon T, Roessler J, et al. Genetic differences in HLA-DQA1∗ and DQB1∗ allelic distributions between celiac and control children in Santiago, Chile. Human immunology. 1999;60(3):262-7.
202. Tinto N, Cola A, Piscopo C, Capuano M, Galatola M, Greco L, et al. High frequency of haplotype HLA-DQ7 in celiac disease patients from South Italy: retrospective evaluation of 5,535 subjects at risk of celiac disease. PLoS One. 2015;10(9):e0138324.
203. Parkkola A, Härkönen T, Ryhänen SJ, Uibo R, Ilonen J, Knip M, et al. Transglutaminase antibodies and celiac disease in children with type 1 diabetes and in their family members. Pediatric diabetes. 2018;19(2):305-13.
204. Husebye ES, Anderson MS. Autoimmune polyendocrine syndromes: clues to type 1 diabetes pathogenesis. Immunity. 2010;32(4):479-87.
205. Bibbò S, Pes GM, Usai-Satta P, Salis R, Soro S, Colosso BMQ, et al. Chronic autoimmune disorders are increased in coeliac disease: A case–control study. Medicine. 2017;96(47).
206. Triolo TM, Armstrong TK, McFann K, Yu L, Rewers MJ, Klingensmith GJ, et al. Additional autoimmune disease found in 33% of patients at type 1 diabetes onset. Diabetes care. 2011;34(5):1211-3.
207. Canova C, Pitter G, Ludvigsson JF, Romor P, Zanier L, Zanotti R, et al. Celiac disease and risk of autoimmune disorders: a population-based matched birth cohort study. The journal of pediatrics. 2016;174:146-52. e1.
208. Elfström P, Sundström J, Ludvigsson JF. Systematic review with meta-analysis: associations between coeliac disease and type 1 diabetes. Alimentary pharmacology & therapeutics. 2014;40(10):1123-32.
209. Barker JM. Type 1 diabetes-associated autoimmunity: natural history, genetic associations, and screening. The Journal of Clinical Endocrinology & Metabolism. 2006;91(4):1210-7.
210. Bratanic N, Smigoc Schweiger D, Mendez A, Bratina N, Battelino T, Vidan-Jeras B. An influence of HLA-A, B, DR, DQ, and MICA on the occurrence of Celiac disease in patients with type 1 diabetes. Tissue Antigens. 2010;76(3):208-15.
211. Viken MK, Flam ST, Skrivarhaug T, Amundsen SS, Sollid LM, Drivvoll AK, et al. HLA class II alleles in Norwegian patients with coexisting type 1 diabetes and celiac disease. Hla. 2017;89(5):278-84.
212. Viljamaa M, Kaukinen K, Huhtala H, Kyrönpalo S, Rasmussen M, Collin P. Coeliac disease, autoimmune diseases and gluten exposure. Scandinavian journal of gastroenterology. 2005;40(4):437-43.
213. Frisk G, Hansson T, Dahlbom I, Tuvemo T. A unifying hypothesis on the development of type 1 diabetes and celiac disease: Gluten consumption may be a shared causative factor. Med Hypotheses. 2008;70(6):1207-9.
214. Hagopian W, Lee HS, Liu E, Rewers M, She JX, Ziegler AG, et al. Co-occurrence of Type 1 Diabetes and Celiac Disease Autoimmunity. Pediatrics. 2017;140(5).

215. Walker-Smith JA, Grigor W. Coeliac disease in a diabetic child. Lancet. 1969;1(7603):1021.
216. Bai JC, Fried M, Corazza GR, Schuppan D, Farthing M, Catassi C, et al. World Gastroenterology Organisation global guidelines on celiac disease. Journal of clinical gastroenterology. 2013;47(2):121-6.
217. Cerutti F, Bruno G, Chiarelli F, Lorini R, Meschi F, Sacchetti C. Younger age at onset and sex predict celiac disease in children and adolescents with type 1 diabetes: an Italian multicenter study. Diabetes care. 2004;27(6):1294-8.
218. Camarca ME, Mozzillo E, Nugnes R, Zito E, Falco M, Fattorusso V, et al. Celiac disease in type 1 diabetes mellitus. Ital J Pediatr. 2012;38:10.
219. Craig ME, Prinz N, Boyle CT, Campbell FM, Jones TW, Hofer SE, et al. Prevalence of Celiac Disease in 52,721 Youth With Type 1 Diabetes: International Comparison Across Three Continents. Diabetes care. 2017;40(8):1034-40.
220. Ludvigsson JF, Ludvigsson J, Ekbom A, Montgomery SM. Celiac disease and risk of subsequent type 1 diabetes: a general population cohort study of children and adolescents. Diabetes Care. 2006;29(11):2483-8.
221. Wijmenga C, Gutierrez-Achury J. Celiac disease genetics: past, present and future challenges. Journal of pediatric gastroenterology and nutrition. 2014;59 Suppl 1:S4-7.
222. Larsson K, Carlsson A, Cederwall E, Jönsson B, Neiderud J, Jonsson B, et al. Annual screening detects celiac disease in children with type 1 diabetes. Pediatr Diabetes. 2008;9(4 Pt 2):354-9.
223. Al-Ashwal AA, Shabib SM, Sakati NA, Attia NA. Prevalence and characteristics of celiac disease in type 1 diabetes mellitus in Saudi Arabia. Saudi medical journal. 2003;24(10):1113-5.
224. Hansen D, Brock-Jacobsen B, Lund E, Bjørn C, Hansen LP, Nielsen C, et al. Clinical benefit of a gluten-free diet in type 1 diabetic children with screening-detected celiac disease: a population-based screening study with 2 years' follow-up. Diabetes care. 2006;29(11):2452-6.
225. Gillett P, Gillett H, Israel D, Metzger D, Stewart L, Chanoine J, et al. High prevalence of celiac disease in patients with type 1 diabetes detected by antibodies to endomysium and tissue transglutaminase. Canadian Journal of Gastroenterology. 2001;15.
226. Salardi S, Volta U, Zucchini S, Fiorini E, Maltoni G, Vaira B, et al. Prevalence of celiac disease in children with type 1 diabetes mellitus increased in the mid-1990s: an 18-year longitudinal study based on anti-endomysial antibodies. Journal of pediatric gastroenterology and nutrition. 2008;46(5):612-4.
227. Fallahi G-H, Ahmadian JH, Rabbani A, Yousefnezhad A, Rezaei N. Screening for celiac disease in diabetic children from Iran. Indian pediatrics. 2010;47(3):268-70.
228. Crone J, Rami B, Huber W, Granditsch G, Schober E. Prevalence of celiac disease and follow-up of EMA in children and adolescents with type 1 diabetes mellitus. Journal of pediatric gastroenterology and nutrition. 2003;37(1):67-71.

229. Goh C, Banerjee K. Prevalence of coeliac disease in children and adolescents with type 1 diabetes mellitus in a clinic based population. Postgraduate medical journal. 2007;83(976):132-6.
230. Salah N, Abdel Hamid F, Abdel Ghaffar S, El Sayem M. Prevalence and type of anaemia in young Egyptian patients with type 1 diabetes mellitus. EMHJ-Eastern Mediterranean Health Journal, 11 (5-6), 959-967, 2005. 2005.
231. Mankaï A, Hamouda HB, Amri F, Ghedira-Besbes L, Harbi A, Sfar MT, et al. Screening by anti-endomysium antibodies for celiac disease in Tunisian children with type 1 diabetes mellitus. Gastroenterologie clinique et biologique. 2007;31(5):462-6.
232. Smith CM, Clarke CF, Porteous LE, Elsori H, Cameron DJ. Prevalence of coeliac disease and longitudinal follow-up of antigliadin antibody status in children and adolescents with type 1 diabetes mellitus. Pediatric diabetes. 2000;1(4):199-203.
233. Elias J, Hoorweg-Nijman JJG, Balemans WA. Clinical relevance and cost-effectiveness of HLA genotyping in children with Type 1 diabetes mellitus in screening for coeliac disease in the Netherlands. Diabetic Medicine. 2015;32(6):834-8.
234. Smigoc Schweiger D, Mendez A, Kunilo Jamnik S, Bratanic N, Bratina N, Battelino T, et al. High-risk genotypes HLA-DR3-DQ2/DR3-DQ2 and DR3-DQ2/DR4-DQ8 in co-occurrence of type 1 diabetes and celiac disease. Autoimmunity. 2016;49(4):240-7.
235. Gutierrez-Achury J, Romanos J, Bakker SF, Kumar V, de Haas EC, Trynka G, et al. Contrasting the Genetic Background of Type 1 Diabetes and Celiac Disease Autoimmunity. Diabetes care. 2015;38:S37-S44.
236. Bratanic N, Smigoc Schweiger D, Mendez A, Bratina N, Battelino T, Vidan-Jeras B. An influence of HLA-A, B, DR, DQ, and MICA on the occurrence of Celiac disease in patients with type 1 diabetes. Tissue antigens. 2010;76(3):208-15.
237. Brorsson CA, Pociot F, Consortium TDG. Shared genetic basis for type 1 diabetes, islet autoantibodies, and autoantibodies associated with other immune-mediated diseases in families with type 1 diabetes. Diabetes care. 2015;38(Supplement 2):S8-S13.
238. Farina F, Picascia S, Pisapia L, Barba P, Vitale S, Franzese A, et al. HLA-DQA1 and HLA-DQB1 Alleles, Conferring Susceptibility to Celiac Disease and Type 1 Diabetes, are More Expressed Than Non-Predisposing Alleles and are Coordinately Regulated. Cells. 2019;8(7):751.
239. Kaur N, Bhadada SK, Minz RW, Dayal D, Kochhar R. Interplay between Type 1 Diabetes Mellitus and Celiac Disease: Implications in Treatment. Dig Dis. 2018;36(6):399-408.
240. Tjon JM, van Bergen J, Koning F. Celiac disease: how complicated can it get? Immunogenetics. 2010;62(10):641-51.
241. Qiao SW, Iversen R, Ráki M, Sollid LM. The adaptive immune response in celiac disease. Semin Immunopathol. 2012;34(4):523-40.
242. Bourgey M, Calcagno G, Tinto N, Gennarelli D, Margaritte-Jeannin P, Greco L, et al. HLA related genetic risk for coeliac disease. Gut. 2007;56(8):1054-9.

243. Margaritte-Jeannin P, Babron M, Bourgey M, Louka A, Clot F, Percopo S, et al. HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. Tissue antigens. 2004;63(6):562-7.
244. Gutierrez-Achury J, Romanos J, Bakker SF, Kumar V, de Haas EC, Trynka G, et al. Contrasting the genetic background of type 1 diabetes and celiac disease autoimmunity. Diabetes care. 2015;38(Supplement 2):S37-S44.
245. Li YR, Zhao SD, Li J, Bradfield JP, Mohebnasab M, Steel L, et al. Genetic sharing and heritability of paediatric age of onset autoimmune diseases. Nature communications. 2015;6(1):1-10.
246. Erlich RL, Jia X, Anderson S, Banks E, Gao X, Carrington M, et al. Next-generation sequencing for HLA typing of class I loci. BMC Genomics. 2011;12(1):42.
247. Delli AJ, Vaziri-Sani F, Lindblad B, Elding-Larsson H, Carlsson A, Forsander G, et al. Zinc transporter 8 autoantibodies and their association with SLC30A8 and HLA-DQ genes differ between immigrant and Swedish patients with newly diagnosed type 1 diabetes in the Better Diabetes Diagnosis study. Diabetes. 2012;61(10):2556-64.
248. Monten C, Bjelkenkrantz K, Gudjonsdottir AH, Browaldh L, Arnell H, Naluai AT, et al. Validity of histology for the diagnosis of paediatric coeliac disease: a Swedish multicentre study. Scandinavian journal of gastroenterology. 2016;51(4):427-33.
249. Grubin C, Daniels T, Toivola B, Landin-Olsson M, Hagopian W, Li L, et al. A novel radioligand binding assay to determine diagnostic accuracy of isoform-specific glutamic acid decarboxylase antibodies in childhood IDDM. Diabetologia. 1994;37(4):344-50.
250. Marsh MN. 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):330-54.
251. Group TS. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. Pediatr Diabetes. 2007;8(5):286-98.
252. Husby S, Koletzko S, Korponay-Szabo IR, Mearin ML, Phillips A, Shamir R, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. Journal of pediatric gastroenterology and nutrition. 2012;54(1):136-60.
253. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018. Diabetes care. 2018;41(Supplement 1):S13-S27.
254. Persson M, Becker C, Larsson HE, Lernmark Å, Forsander G, Ivarsson S, et al. The Better Diabetes Diagnosis (BDD) study–A review of a nationwide prospective cohort study in Sweden. diabetes research and clinical practice. 2018;140:236-44.
255. Almqvist C, Adami H-O, Franks PW, Groop L, Ingelsson E, Kere J, et al. LifeGene—a large prospective population-based study of global relevance. European journal of epidemiology. 2011;26(1):67-77.
256. Vaziri-Sani F, Delli AJ, Elding-Larsson H, Lindblad B, Carlsson A, Forsander G, et al. A novel triple mix radiobinding assay for the three ZnT8 (ZnT8-RWQ) autoantibody variants in children with newly diagnosed diabetes. Journal of immunological methods. 2011;371(1-2):25-37.

257. Lind A, Akel O, Wallenius M, Ramelius A, Maziarz M, Zhao LP, et al. HLA high-resolution typing by next-generation sequencing in Pandemrix-induced narcolepsy. PloS one. 2019;14(10):e0222882.
258. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. The American Journal of Human Genetics. 2002;70(2):425-34.
259. Lake SL, Lyon H, Tantisira K, Silverman E, Weiss S, Laird N, et al. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Human heredity. 2003;55(1):56-65.
260. Payami H, Joe S, Farid N, Stenszky V, Chan S, Yeo P, et al. Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. American journal of human genetics. 1989;45(4):541.
261. Association AD. Diagnosis and classification of diabetes mellitus. Diabetes care. 2014;37(Supplement 1):S81-S90.
262. Bonifacio E, Beyerlein A, Hippich M, Winkler C, Vehik K, Weedon MN, et al. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: A prospective study in children. PLOS Medicine. 2018;15(4):e1002548.
263. Congia M, Frau F, Lampis R, Frau R, Mele R, Cucca F, et al. A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201. Tissue Antigens. 1992;39(2):78-83.
264. Ting JP, Trowsdale J. Genetic control of MHC class II expression. Cell. 2002;109 Suppl:S21-33.
265. Zhao LP, Alshiekh S, Zhao M, Carlsson A, Larsson HE, Forsander G, et al. Next-Generation Sequencing Reveals That HLA-DRB3, -DRB4, and -DRB5 May Be Associated With Islet Autoantibodies and Risk for Childhood Type 1 Diabetes. Diabetes. 2016;65(3):710-8.
266. Todd JA, Bell JI, McDevitt HO. HLA-DQ β gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature. 1987;329(6140):599-604.
267. Kim C-Y, Quarsten H, Bergseng E, Khosla C, Sollid LM. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proceedings of the National Academy of Sciences. 2004;101(12):4175-9.
268. Busch R, De Riva A, Hadjinicolaou AV, Jiang W, Hou T, Mellins ED. On the perils of poor editing: regulation of peptide loading by HLA-DQ and H2-A molecules associated with celiac disease and type 1 diabetes. Expert reviews in molecular medicine. 2012;14:e15.
269. Antvorskov JC, Josefsen K, Engkilde K, Funda DP, Buschard K. Dietary gluten and the development of type 1 diabetes. Diabetologia. 2014;57(9):1770-80.
270. Di Sabatino A, Vanoli A, Giuffrida P, Luinetti O, Solcia E, Corazza GR. The function of tissue transglutaminase in celiac disease. Autoimmunity reviews. 2012;11(10):746-53.
271. Gioia L, Holt M, Costanzo A, Sharma S, Abe B, Kain L, et al. Position β57 of I-A^{g7} controls early anti-insulin responses in NOD mice, linking an

MHC susceptibility allele to type 1 diabetes onset. Science Immunology. 2019;4(38):eaaw6329.

272. Koeleman BP, Lie BA, Undlien DE, Dudbridge F, Thorsby E, De Vries RR, et al. Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. Genes & Immunity. 2004;5(5):381-8.
273. Murray JA, Moore SB, Van Dyke CT, Lahr BD, Dierkhising RA, Zinsmeister AR, et al. HLA DQ gene dosage and risk and severity of celiac disease. Clin Gastroenterol Hepatol. 2007;5(12):1406-12.
274. Gutierrez-Achury J, Romanos J, Bakker SF, Kumar V, de Haas EC, Trynka G, et al. Contrasting the Genetic Background of Type 1 Diabetes and Celiac Disease Autoimmunity. Diabetes Care. 2015;38 Suppl 2:S37-44.
275. Bakker SF, Tushuizen ME, Stokvis-Brantsma WH, Aanstoot HJ, Winterdijk P, van Setten PA, et al. Frequent delay of coeliac disease diagnosis in symptomatic patients with type 1 diabetes mellitus: Clinical and genetic characteristics. European Journal of Internal Medicine. 2013;24(5):456-60.
276. Sumnik Z, Cinek O, Bratanic N, Kordonouri O, Kulich M, Roszai B, et al. Risk of Celiac Disease in Children With Type 1 Diabetes Is Modified by Positivity for HLA-DQB1* 02-DQA1* 05 andTNF− 308A. Diabetes Care. 2006;29(4):858-63.
277. Tait B, Colman P, Morahan G, Marchinovska L, Dore E, Gellert S, et al. HLA genes associated with autoimmunity and progression to disease in type 1 diabetes. Tissue antigens. 2003;61(2):146-53.
278. Lipponen K, Gombos Z, Kiviniemi M, Siljander H, Lempainen J, Hermann R, et al. Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. Diabetes. 2010;59(12):3253-6.
279. Haimila K, Peräsaari J, Linjama T, Koskela S, Saarinenl T, Lauronen J, et al. HLA antigen, allele and haplotype frequencies and their use in virtual panel reactive antigen calculations in the Finnish population. Tissue Antigens. 2013;81(1):35-43.
280. Verdu EF, Danska JS. Common ground: shared risk factors for type 1 diabetes and celiac disease. Nature Immunology. 2018;19(7):685-95.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

Immunogenetics of Type 1 diabetes and Celiac disease

Shehab Alshiekh graduated as a medical doctor from King Abdulaziz University in Jeddah, Saudi Arabia in 2008 and then worked as a Physician at King Abdulaziz university hospital until he moved to work as a physician at Skåne university hospital and doctorate student at Lund University, Sweden. The main interest of his research is to study the genetics of diabetes and celiac disease.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences, Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2020:101
ISBN 978-91-7619-963-3
ISSN 1652-8220