

From THE DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

GENETIC ANALYSIS OF CANDIDATE SUSCEPTIBILITY GENES FOR TYPE 1 DIABETES

Samina Asad



**Karolinska
Institutet**

Stockholm 2012

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by Larserics Digital Print AB, Sweden

© Samina Asad, 2012

ISBN 978-91-7457-842-3

ABSTRACT

Type 1 diabetes (T1D) is a complex disease where the pancreatic β -cells are destroyed in an autoimmune attack. For the patients, this leads to lifelong daily insulin treatment and increased risk for various kinds of complications. It is thought that both environmental as well as genetic factors act in concert to cause T1D. The Human Leukocyte Antigen (HLA) region located on chromosome 6 accounts for about 50% of the genetic risk to develop T1D. Several other genes are also known to contribute to disease risk.

Paper I. Previous publications indicate that the *programmed cell death 1 (PDCDI)* gene (chr.2) is associated to various autoimmune diseases. *PDCDI* is involved in maintaining self tolerance. The aim of our study was to test the involvement of the *PDCDI* gene in T1D susceptibility. However, when two separate Swedish cohorts were analyzed no association or linkage was found between T1D and the *PDCDI* gene. Nor did we observe any association in a meta-analysis with a previous study reporting association between *PDCDI* and T1D.

Paper II. We have in a previous study observed suggestive linkage to the chromosome 5p13-q13 region in Scandinavian T1D families. This region showed stronger evidence of linkage, when only the Swedish families were investigated. Genotyping of more than 70 markers in the Swedish families revealed two associated candidate genes: *5-hydroxytryptamine (serotonin) receptor 1A (HTR1A)* and *ringfinger protein 180 (RNF180)*. Association of both genes has been confirmed by us in Danish families. The two genes are in strong linkage disequilibrium with each other. However conditional analysis data suggest that *HTR1A* may be most strongly associated. Further, we report that *HTR1A* is expressed in human β -cells and α -cells.

Paper III. The *class II transactivator (CIITA)* gene (chr.16) is crucial for MHC II gene regulation and has been reported to associate with susceptibility to a number of complex diseases. By genotyping SNPs in Swedish T1D cohorts and the combined control material from previous studies of *CIITA* we have observed significant difference in the genotype distribution for three markers in *CIITA* with respect to age, in the collected control material. This phenomenon was confirmed in an independent control material. After adjusting for age we detect association to T1D for two markers in our T1D material. Further, we observed interaction between markers in *CIITA* and the protective *HLA DR15* haplotype. These findings suggest that a polymorphism in the *CIITA* gene area may be associated with type 1 diabetes susceptibility. Importantly, results also suggest that control groups should be properly matched for the cases.

Paper IV. In complex diseases genes seldom act alone in disease susceptibility. Instead it is thought that genes may interact with each other. The aim of our investigation was to study the interaction of the most significantly associated genes in T1D (*HLA-DRB1*, *HLA-DQB1*, *INS* and *PTPN22*). This was done by comparing four different models for studying interaction; multiplicative and additive interaction models, Multifactor dimensionality reduction (MDR) model and Bayesian Networks (BN) model. Results indicate several interaction terms mainly in the additive model. Further, we show that the additive interaction model has the strongest prediction accuracy rate indicating that this is the model of preference.

In summary, in order to better understand the cause of T1D the aim of this thesis was to identify single genes as well as gene-gene interactions which may influence the risk of T1D development.

LIST OF PUBLICATIONS

- I. **Samina. Asad**, P. Nikamo, C. Törn, M. Landin-Olsson, Å. Lernmark, M. Alarcón-Riquelme, I. Kockum, Diabetes Incidence in Sweden Study Group. **No evidence of association of the PDCD1 gene with Type 1 diabetes.**
Diabetic Medicine, 2007, 24, 1473-1477

- II. **Samina. Asad***, P. Nikamo*, A. Gyllenberg, H. Bennet, O. Hansson, N. Wierup, Diabetes Incidence in Sweden Study Group, A. Carlsson, G. Forsander, S-A. Ivarsson, H. Larsson, Å. Lernmark, B. Lindblad, J. Ludvigsson, C. Marcus, K. S Rønningen, J. Nerup, F. Pociot, H. Luthman, M. Fex, I. Kockum.
HTRIA a novel type 1 diabetes susceptibility gene on chromosome 5p13-q13.
Plos ONE, 2012, vol. 7, Issue 5, e35439

- III. A. Gyllenberg, **Samina. Asad**, F. Piehl, M. Swanberg, L. Padyukov, B. Van Yserloo, E. A Rutledge, B. McNeney, J. Graham, M. Orho-Melander, E. Lindholm, C. Graff, C. Forsell, K. Åkesson, M. Landin-Olsson, A. Carlsson, G. Forsander, S-A. Ivarsson, H. Larsson, B. Lindblad, J. Ludvigsson, C. Marcus, Å. Lernmark, L. Alfredsson, K. Åkesson, T. Olsson, I. Kockum for the Swedish Childhood Diabetes Study Group, the Diabetes Incidence in Sweden Study Group, and the Better Diabetes Diagnosis Study group.
Age Dependent Variation of Genotypes in Major Histocompatibility Complex Class II Transactivator in Controls and Association to Type 1 Diabetes.
In Press, Genes and Immunity

- IV. **Samina. Asad***, H. Westerlind*, P. Nikamo, L. Goop, A. Carlsson, C. Marcus, J. Ludvigsson, Å. Lernmark, S-A. Ivarsson, H. Larsson, G. Forsander, B. Lindblad, H. Källberg, T. Koski, I. Kockum for the Swedish Childhood Diabetes Study Group and the Diabetes Incidence in Sweden Study Group.
Investigation of interaction between DRB1*04-DQA1*03:01-DQB1*03:02, DRB1*03-DQA1*05:01-DQB1*02:01, DRB1*15-DQA1*01:02-DQB1*06:02, Insulin and PTPN22 using four different interaction models.
Manuscript

* These authors contributed equally to the work

CONTENTS

1	BACKGROUND.....	1
1.1	GENETICS.....	1
1.1.1	Genetic Diseases.....	1
1.1.2	DNA Variation.....	2
1.1.2.1	Single Nucleotide Polymorphisms.....	2
1.1.2.2	Structural Variation.....	3
1.1.2.3	Repeats.....	3
1.1.3	Genetic Approaches to Identify Diseases Susceptibility Genes..	4
1.1.3.1	Linkage Analysis.....	4
1.1.3.2	Linkage Disequilibrium and Association Studies.....	6
1.1.3.3	Interaction Studies.....	8
1.2	DIABETES.....	11
1.2.1	Type 1 Diabetes.....	13
1.2.1.1	Symptoms and Complications.....	14
1.2.1.2	Incidence and Epidemiology.....	14
1.2.1.3	Innate and Adaptive Immunity.....	15
1.2.1.4	Immunologic Tolerance and Autoimmunity.....	17
1.2.1.5	Disease Susceptibility Factors.....	18
1.2.1.5.1	Environmental Factors.....	19
1.2.1.5.2	Genetic Predisposition.....	21
1.2.1.6	Animal Models.....	26
2	STUDY AIMS.....	27
3	MATERIALS.....	28
3.1	Scandinavian Families.....	28
3.2	Swedish Patients and Controls.....	29
3.3	Individuals for immunohistochemistry and expression studies.....	31
4	METHODS.....	32
4.1	Fine Mapping of Chromosome 5 (paper II).....	32
4.2	SNP Genotyping.....	32
4.3	Sequencing of the <i>HTR1A</i> gene (paper II).....	34
4.4	HLA Typing (paper III and IV).....	34
4.5	Imputation (paper II).....	35
4.6	Q-PCR of <i>HTR1A</i> and <i>RNF180</i> mRNA from human islets of Langerhans (paper II).....	35
4.7	Tissue preparation and immunohistochemistry (paper II).....	35
4.8	Statistical Analysis.....	36
4.9	Computational Analysis (paper II).....	37
4.10	Adjustment for Age.....	37
4.11	Imputation (paper II).....	38
4.12	Interaction Studies.....	38
5	RESULTS AND CONCLUSION.....	39
5.1	Paper I.....	39
5.2	Paper II.....	42
5.3	Paper III.....	44
5.4	Paper IV.....	45
6	CONCLUDING REMARKS.....	49
7	FUTURE PERSPECTIVES.....	51
8	ACKNOWLEDGEMENTS.....	52
9	REFERENCES.....	55

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AP	Attributable proportion due to interaction
BBrat	Biobreeding rat
BDD	Better Diabetes Diagnosis Study
BN	Bayesian Networks
CI	Confidence interval
<i>CIITA</i>	MHC class II transactivator
<i>CLEC16A</i>	C-type lectin domain family 16
CNV	Copy number variation
CTL	Cytotoxic T-cells
<i>CTLA-4</i>	Cytotoxic T-Lymphocyte Antigen 4
DASH	Dynamic allele-specific hybridization
DIEGG	Danish IDDM Epidemiology and Genetics Group
DISS1/2	Diabetes Incidence Study in Sweden 1/2
DR	Diabetes Registry in Southern Sweden
DSGD	Danish Study Group of Diabetes in Childhood
DZ	Dizygotic twin
GAD 65	Glutamic acid decarboxylase
GWAS	Genome wide association studies
HLA	Human leukocyte antigen
<i>HTR1A</i>	5-hydroxytryptamine (serotonin) receptor 1A
htSNP	Tagging Single Nucleotide Polymorphism
IAA	Insulin autoantibodies
IA-2	Protein tyrosine phosphatase-like molecule
IBD	Identical by descent
ICA	Islet Cell Antibodies
<i>IFIH1</i>	Interferon induced with helicase C domain 1
<i>IL2RA</i>	Interleukin-2 receptor α chain
<i>INS</i>	Insulin gene
LADA	Latent autoimmune diabetes in adults
LD	Linkage disequilibrium
LOD	logarithm (base 10) of odds
LYP	Lymphoid-specific phosphatase
MDR	Multifactor dimensionality reduction
MI	Myocardial Infarction
MODY	Maturity onset diabetes of the young
MS	Multiple Sclerosis
MZ	Monozygotic twin
NOD-mouse	Non-obese diabetic-mouse
OR	Odds Ratio
<i>PDCD-1</i>	Immunoreceptor PD-1
PDT	Pedigree disequilibrium test
<i>PTPN22</i>	Protein tyrosine phosphatase non-receptor type 22
RA	Rheumatoid Arthritis

RERI	Relative excess risk due to interaction
<i>RNF180</i>	Ring finger protein 180
ROC	Relative Operating Characteristic
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SV	Structural variation
SV2	Swedish childhood Study
TDT	Transmission disequilibrium test
T-reg	Regulatory T-cell
T1DGC	Type 1 diabetes Genetic Consortium
T1D	Type 1 diabetes
T2D	Type 2 diabetes
VNTR	Variable number of tandem repeats
5-HT	5-hydroxytryptamine (serotonin)

♥ *To My Family* ♥

1 BACKGROUND

1.1 GENETICS

The word genetics comes from the Greek word “genitive” which in turn originates from the word *genesis* plainly meaning “origin”. Genetics refers to the genes involved in the heredity and variation in all living organisms.

In 1856 an Austrian monk named Gregor Mendel also called the “Father of genetics”, planted pea seeds in his monastery garden and discovered that certain traits of the pea plant seemed to follow specific laws of so called *dominant* and *recessive* inheritance patterns. These findings later became known as the three Laws of Mendel’s Inheritance; Law of segregation, dominance and independent assortment.

1.1.1 Genetic Diseases

By using Mendel’s studies as a base, scientists later discovered that the genetic information of the cells of all living organisms is packed in what we call the “genome”. In some cases, errors occur in the genome leading to various genetic diseases. Genetic diseases can be divided into four categories: chromosomal, monogenic, mitochondrial and complex/multifactorial.

Diseases caused by changes in the chromosomes involve mutations in large chromosome segments. Sometimes even whole chromosomes may be involved which is the case in Down syndrome (trisomy 21) where affected individuals carry an extra copy of chromosome 21.

Monogenic diseases follow the Mendelian inheritance pattern (autosomal dominant/recessive, X-linked dominant/recessive or Y-linked). These types of diseases are relatively rare and their mutations are fairly easily identified. An example of a recessive condition is Cystic Fibrosis.

Mitochondrial diseases are extremely rare and are only passed on by mothers to offspring. Maternally inherited diabetes with deafness is an example of rare mitochondrial disease.

Multifactorial diseases do not follow traditional Mendelian inheritance patterns. Instead it is believed that these diseases are caused by multiple gene-gene interactions as well

as interaction with environmental factors. Most common diseases such as type 2 diabetes, cardiovascular and autoimmune diseases are examples of multifactorial diseases.

1.1.2 DNA Variation

In 1962 James Watson and Francis Crick shared the Noble prize for the discovery of the structure of deoxyribonucleic acid (DNA). The DNA molecule is the basis of all heredity and it is the basic coding block for proteins and enzymes in living organisms. Watson and Crick determined that DNA is constructed of two chains, so called helixes. The two chains are held together by hydrogen bonds between pairs of bases; adenine (A) binding to thymine (T) and guanine (G) binding to cytosine (C). All four bases are also individually attached to a sugar-and phosphate molecule forming so called nucleotides. It is the sequences and combinations of bases on the DNA helix that build up and maintain an organism. Approximately 99% of all bases in the genome are in the same order in all individuals. However, the remaining 1% of all bases varies between individuals. These DNA sequence variations have been important for the process of human evolution and the creation of population heterogeneity and making individuals more fit to adjust in new environments. Variations in the DNA also provide useful help when trying to identify genes that cause multifactorial diseases [1].

1.1.2.1 Single Nucleotide Polymorphisms

The most common example of DNA sequence variations are Single Nucleotide Polymorphisms (SNPs). A SNP is an alteration in a single base (A, T, C or G) in the DNA sequence, usually varying between two nucleotides in a specific base pair position (Figure 1).

Thankfully, the vast majority of all the millions of SNPs that each individual carries are so called silent mutations and do not cause any harm or damage.

However, in some cases SNPs located within genes may alter the expression level or lead to expression of alternative variants of a protein which in turn leads to a specific phenotype.

One way of identifying disease susceptibility genes is to study the changes caused by SNPs and compare them in patients and controls.

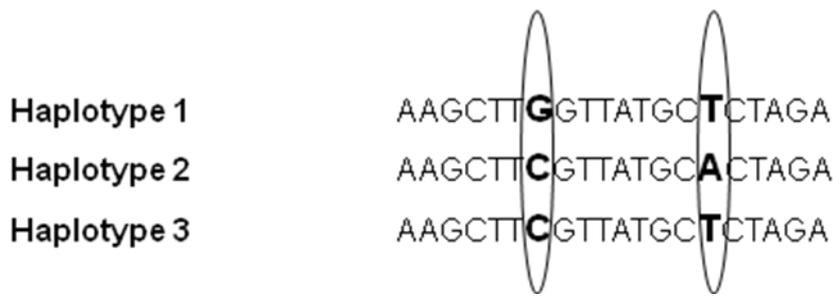


Figure 1. Illustration of an alternation in a DNA sequence caused by a Single Nucleotide Polymorphism (SNP) in three separate haplotypes

1.1.2.2 Structural Variation

The second most common type of DNA variation in humans is structural variations (SV's). Unlike SNPs SV's involve variation in more than only one base pair. Insertions, deletions, duplications, translocations and copy number variations are all examples of SV's. Insertions and deletions (often called INDELs) range from 1bp up to 10.000 bp's in length [2]. Translocations and duplications involve rearrangements of larger chromosomal segments. Copy number variations (CNV's) are often longer than 1kb and sometimes ranges up to 3Mb. Longer forms of deletions, insertions, duplications, inversions and translocations are all termed CNV's [3,4]. Like SNPs, all types of SV's may be involved in disease susceptibility by interference with gene expression and subsequently altering the translation of proteins [5]. For example there is an enhanced susceptibility of HIV infection if one has a lower copy number of the *CCL3L1* gene [6]. Furthermore, studies indicate that there is a higher risk of being affected by various types of cancer if one has homozygous deletions of the glutathione S-transferase genes (*GSTT1* and *GSTM1*) [7].

1.1.2.3 Repeats

Throughout the human genome, there are DNA sequences that are repeated in a row. The number of repeats varies between individuals and populations and are therefore an ideal tool for identifying disease genes [8]. The repeats can be classified into three different groups:

Microsatellites (also known as short tandem repeats) are short repeats of 1-6bp.

Minisatellites (also called variable number of tandem repeats, VNTR) are longer repeats ranging from 9-80bp.

Megasatellites can be up to several Kb long [9].

Examples of diseases caused by this type of polymorphisms include Huntingtons and Myotonic dystrophy.

1.1.3 Genetic Approaches to Identify Disease Susceptibility Genes

Since the mode of inheritance is unknown in complex diseases, finding disease causing genes and gene regions is a very difficult task. There are however two main ways to identify genetic regions or alleles that cause a specific phenotype; *Linkage analysis* and *Association studies*. Both methods can be used on candidate genes as well as in genome wide studies.

1.1.3.1 Linkage Analysis

Linkage analysis can only be used in family materials since it studies the inheritance pattern of certain markers together with a phenotype. In early days microsatellites were used for studying linkage but nowadays SNPs are used more often.

Chromosome pairs randomly exchange genetic material during the early stages of cell division, so called meiosis. This trade is called recombination.

The probability of two genes undergoing recombination is much higher in genes that are located far apart as compared to genes that are in close proximity to each other. Genes or markers located close to each other very rarely recombine and are therefore said to be linked (inherited together).

The extent of linkage is measured by the recombination fraction, denoted θ (theta).

Unlinked genes show 50% recombination and have a recombination fraction of 0.5.

When θ is 0 the studied genes are thought to be in complete linkage. Linkage is calculated using the LOD score (the score of the logarithm of odds) which is a statistical method to calculate the significance of obtained genotyping results given the observed phenotypes in a pedigree and given a mode of inheritance for the trait. The LOD score represents the ratio of two hypothesis; the null hypothesis where there is free recombination, H_0 (no linkage and $\theta=0.5$) and the H_1 hypothesis where linkage between loci is observed. The likelihood that the studied loci are linked rather than the likelihood of observing obtained data by chance is calculated as follows;

$LOD = \log_{10} \text{Likelihood of linkage } (H_1) / \text{Likelihood of free recombination } (H_0)$

There are two categories of linkage analysis (*parametric and non-parametric linkage analysis*) where the transmission of inherited DNA markers can be studied and compared from generation to generation for identifying a disease affecting gene region. Parametric (model based) linkage analysis has been very successful in the search for genes causing monogenic diseases. Parametric analysis however requires information of certain parameters such as mode of inheritance, allele frequencies and mutation rates. This creates problems while studying complex diseases where the mode of inheritance is unknown. Instead, non-parametric (model free) linkage analysis is preferred since it does not require specification of the inheritance mode [10]. In this method the studied families must have at least two affected individuals, often sib-pairs. In affected sib-pair analysis the families are genotyped to see how often a genetic marker is shared identical by descent (IBD) in the siblings. The expected IBD allele sharing in siblings if no linkage is present is 25% for sharing both alleles; 50% for sharing only one and 25% for not sharing any alleles. Increased LOD scores will be observed if family members share alleles more excessively than expected. The non parametric analysis is considered to have less power than the parametric analysis since only shared alleles among cases are studied and no genetic mode is assumed. Therefore, a large number of families with at least two affected relatives are required. In parametric analysis obtaining a LOD score of 3 is considered to be significant and basically indicates that the probability that the studied locus is linked is a 1000 times higher than that it is not linked assuming that there is only one linked polymorphism in the disease. Having observed a significant LOD score of 3 leads to the conclusion that the studied marker is located by a disease susceptibility marker/locus [11]. A LOD score of 2 is thought to be suggestively significant while a LOD score of less than 2 is non-significant. Depending on which statistical method a computer program uses, significant LOD scores should always be presented with corresponding p-values. In most statistical analysis a p-value of 0.05 is considered to correspond to significant results. This value indicates that if the study is repeated a 100 times the chance of obtaining similar results purely by chance is 5%. However, it is important to distinguish between point wise (nominal) significance levels (where only a single locus is studied)

and genome-wide significance levels (where a large number of markers are studied). In order to reach genome wide (GW) significance levels of 0.05, the nominal significance levels must therefore be set to much more stringent values. Thus for non-parametric analysis where many polymorphisms may be linked to the disease, a LOD score of 2.2 (nominal $p \leq 0.00074$) is regarded as suggestive linkage which means that the chance of obtaining similar results by chance is once in a genome-wide linkage analysis. A LOD score of ≥ 3.6 (nominal $p \leq 0.000022$) is evidence of genome-wide significant and indicates that in every 20 genome-wide linkage scans significant linkage will occur one time [12,13]. Even though linkage analysis has helped scientists to identify several disease susceptibility genes it has its draw backs. In monogenic diseases, linkage analysis manages to define small areas where only one or few genes are located. In complex diseases however, no robust methods which give sufficient statistical correlations with a specific locus have been developed. This leads to linked areas often contain hundreds and sometimes thousands of genes making it extremely tedious to find a susceptibility gene. A main reason for this is probably that we usually do not have large family materials with many affected individuals leading to few recombinations between phenotype and marker. Monogenic diseases on the other hand are more easily studied because of extended family pedigrees with large amount of recombination leading to fewer linked areas. An additional reason for these large linked regions is that many genes are involved sometimes mapping to the same region of the chromosome.

1.1.3.2 Linkage Disequilibrium (LD) and Association Studies

Identifying a SNP with strong association to a disease may not per se mean that the studied SNP is causative. Instead it may be in linkage disequilibrium (LD) with the true causative SNP. Therefore it is thought that the power of an association study increases with high LD. The definition of LD is as follows; non random association of two or more loci on the same chromosome. SNPs that are in LD with each other are therefore said to be inherited together on the same so called haplotypes. There are several factors that influence LD, such as random mating and migration, selection, rate of mutations, genetic drift and recombination fraction [14]. LD can be calculated in two ways; either using the X^2 -test [15] or by calculating the excess of alleles [16]. The effect of the LD may then be measured using either r^2 or D' [17] where r^2 is usually most preferably used.

Association studies can be divided into two types; *candidate gene studies* and *genome wide association studies (GWAS)*. In the candidate gene study approach the investigated gene is picked due to prior knowledge about the gene in the disease while in the latter method, no specific pre-existing knowledge of genes is required. Recently the latter approach has been widely used since it is a rapid way of scanning whole genomes in search of genetic variations which could lead to certain diseases. This method however requires large amounts of patients and controls.

Association studies involve testing whether a certain allele is more or less common in affected individuals compared to the healthy population. If association is found, it is thought that the studied allele is either directly involved in the susceptibility to disease or in LD with a susceptibility allele on the same or nearby locus.

Association studies can be both *population based* and *family based* studies. In the *family based* approach the transmission disequilibrium test (TDT) which detects association in the presence of linkage, is most commonly used.

Unlike in traditional linkage analysis where large families are required, the TDT can be calculated using only trio families (two parents and one offspring). No affected siblings are needed. It is however necessary that at least one parent is heterozygous for the allele that is studied. The TDT evaluates the transmission frequency of the disease/non disease associated alleles from parent to child [18]. A transmission frequency of more than 50% indicates that the studied allele is associated with disease. On the other hand, a transmission frequency of less than 50% is considered to indicate disease protection. A further development of the TDT has been generated; the pedigree disequilibrium test (PDT). This test is used when larger families are studied and when several affected individuals are involved [19].

Samples from affected and unaffected unrelated individuals are collected in the *population based* (case-control) approach. Association analysis can be done by various methods depending of question, but a common way to test for allele frequency differences is by using a χ^2 - test. In the population based studies it is crucial that the cases and controls are matched to each other in regard of ethnicity. Proper matching may minimize the risk of population stratification and false positive results. Many large GWAS studies use the Principal Component Analysis (PCA) to reduce population stratification effects.

1.1.3.3 Interaction Studies

Complex diseases arise from genetic as well as environmental factors yet not all genetically susceptible individuals respond to environmental factors in the same way. This difference can be explained by gene-environment interaction and gene-gene or so called epistasis. In epistasis it is considered that the effects of one gene is masked by one or several other genes. However, there is an disagreement regarding the meaning of the term epistasis. Most population geneticists claim that epistasis can only be applied on so called quantitative traits (due to certain statistical calculations) whereas others claim it refers to the same phenomena as genetic interaction [20]. In its original definition it refers to the masking of the effect of on gene by the genotype present at another gene. Lately more and more focus is being given to interaction studies. It is thought that these studies may shed more light into which factors are involved in different pathways leading to the development of a multifactorial disease.

Interaction studies can be performed using various statistical models. The most commonly used approaches are however the *Additive model* or the *Multiplicative model*.

In the Additive model, it is assumed that no interaction is present between the studied factors [21]. In that case the estimated risk for individuals exposed to two risk factors is the sum of the risk for the individual factors. For understanding a possible interaction, the so called “pie model” is used. In the pie model it is assumed that if the studied factors are not jointly required in a specific pathway leading to disease, the factors are not included in the same sufficient cause and hence are independent of each other (no interaction) (Figure 2) [22]. In other words, if the total effect of two factors deviates from additivity what is called causal interaction, is thought to be present [23]. Causal interaction is also referred to as biological interaction in the field of epidemiology.

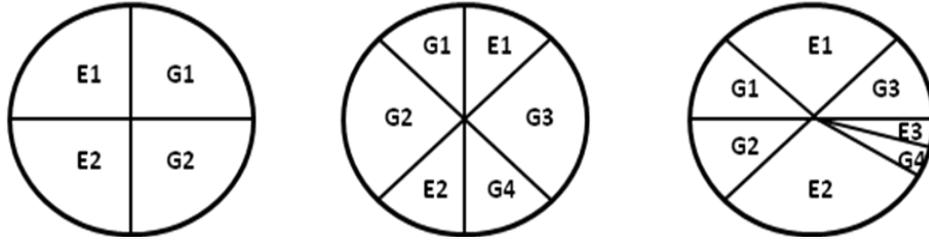


Figure 2. The three pies represent three individuals with the same disease. *E* denotes an environmental factor while *G* denotes genetic factors. If all factors are present in each respective individual disease will be developed. These factors are then referred to as a sufficient cause for the disease. The structure of the pies may look different in different individuals where some factors are unique for an individual while others are shared. In conclusion, single associated genetic and environmental factors are necessary but not sufficient to cause disease.

The degree of interaction may in turn be estimated by the use of three different measures; Relative excess risk due to interaction (RERI), the attributable proportion due to interaction (AP) and the synergy index (SI) [24].

The additive model is thought to explain both causal and statistical interaction. This is not the case with multiplicative interaction where interaction is only explained on a statistical level [23]. The multiplicative model which is based on a logistic regression model is more widely used compared to the additive model. In the multiplicative model, the joint effect of the product of the total effect of each individual factor is estimated. If the interaction term is significantly associated then statistical interaction between the tested factors is present.

Interaction can also be studied by using the *Multifactor dimensionality reduction (MDR)* and *Bayesian Networks (BN)* models. MDR is a non-parametric model where genotype data are divided into high risk and low risk individuals converting multidimensional variables into lower dimensional space. This step determines which combinations of risk factors predict affection status. It is then possible to determine how well the classification of risk factors predict affection status [25] (Figure 3).

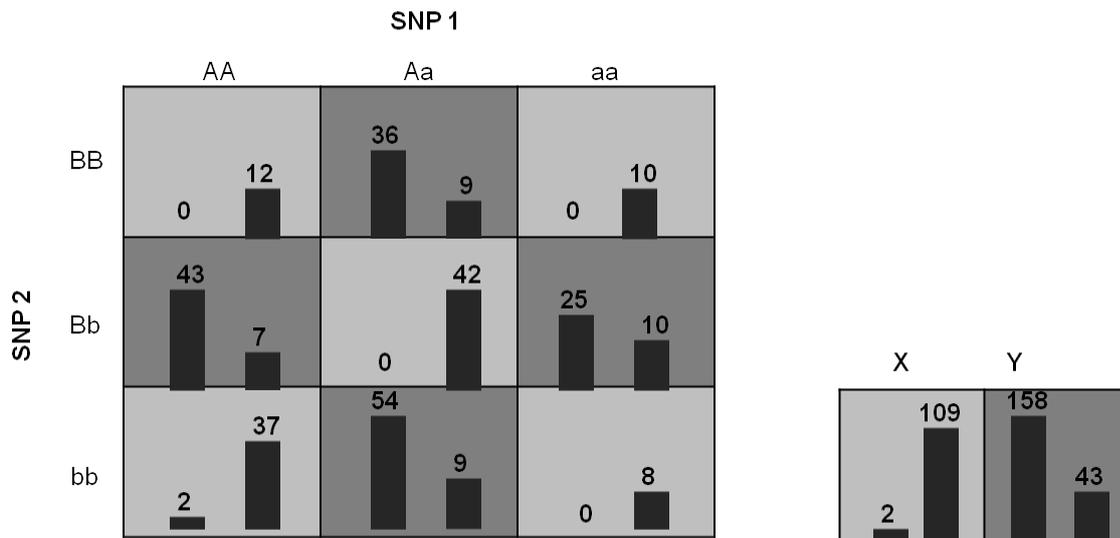


Figure 3. The figure illustrates a graphical model of MDR using two SNPs; SNP 1 and 2. The large table indicates the number of cases and controls for each genotype combination. The ratio of cases to controls for each genotype indicates whether a genotype combination is associated with risk (dark grey boxes) or protection (light grey boxes). The small table indicates the data being converted into a lower dimensional space including the total number of cases and controls carrying non risk and risk genotypes respectively (X and Y).

The BN model may be used for both causal as well as probabilistic interactions and is therefore ideal for including prior knowledge. In the BN model principles from graph and probability theories are combined with statistics and computer science. Each so called node in figure 4 is thought to represent a random variable. The edges between the variables are estimated by using computer sciences and statistics. The edges are thought to represent probabilistic dependencies between various studied variable. The BN model is believed to be particularly good since it does not assume any statistical model for interaction, it avoids data “over fitting” and can be used even when data is missing. BN is an ideal method to get an overview of possible causal interactions [26].

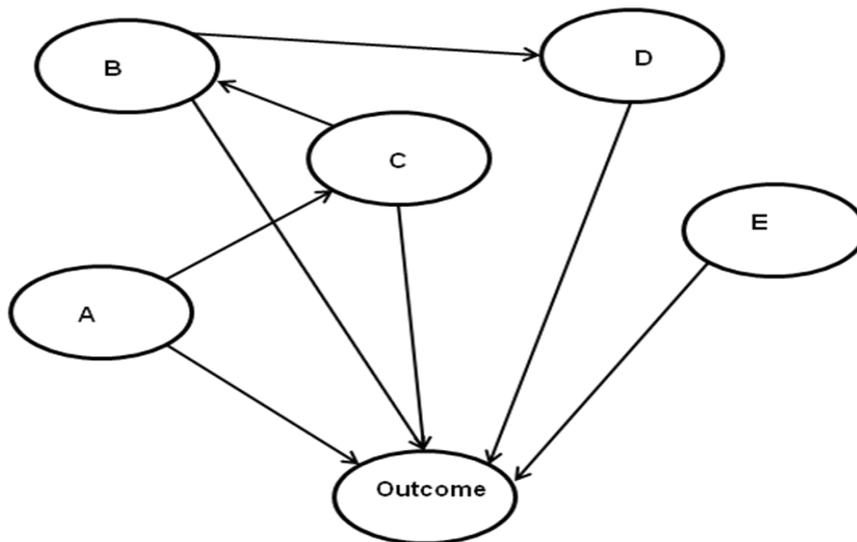


Figure 4. Example of a directed acyclic graph used in BN analysis where variables A-E represent different risk factors for the assumed outcome. All factors influence the outcome. However, A affects the outcome directly as well as indirectly both through risk factor B, C and D. E has no interaction with the other risk factors.

1.2 DIABETES

The number of diabetes cases is rapidly growing throughout the world. However, diabetes is not a “new” disease for humans. It was first mentioned in 1550BC in Egypt that a rare disease causes the patients to urinate frequently and to rapidly lose weight. Later, an ancient Greek physician named Aretaeus (30-90CE) noted a condition with symptoms such as frequent urination, excessive thirst and severe loss of weight. Aretaeus named this condition diabetes which basically means “flowing through” [27]. Years went by and until the 20th century patients suffering from diabetes rarely lived for more than a few years after being diagnosed with the disease.

In 1921 Fredrick Banting and Charles Best started isolating insulin from animals. The first bovine insulin treatment was given to a patient suffering from diabetes and it was seen that the patient’s condition improved dramatically. From that day the lives of diabetes patients changed and now, if treated right, it is no longer considered to be a deadly disease.

Science has come a long way after the discovery of insulin. It is now known that insulin is a hormone which is vital for the processing of glucose into energy. Diabetes is

classified as a chronic metabolic disease where patients have a change in insulin production caused by complex interactions of several different factors. In the pancreas cells called β -cells produce insulin. In turn insulin converts glucose into energy in the periphery. Energy is crucial for normal cell growth and cell survival. Inability to convert glucose into energy leads to high blood sugar levels (hyperglycaemia). There are two major types of diabetes; *type 1 diabetes* (T1D) and *type 2 diabetes* (T2D). T1D is unlike T2D an autoimmune disease and accounts for around 5-10% of all diabetes incidents. In general, T1D affects young people under the age of 35. At the time of diagnosis, T1D patients often lack the ability to produce insulin and suffer from ketoacidosis.

The vast majority of all diabetes incident cases (90%) are classified as T2D patients. These patients usually have a later disease onset and are able to produce insulin but are not able to respond to the insulin production (so called insulin resistance) and if not treated properly the production of insulin later declines.

The classic subdivision of diabetes with only two diabetes types however has been proven not to be entirely correct. Even older people have been diagnosed with type T1D. This form of diabetes is called latent autoimmune diabetes in adults (LADA). Further, more and more children are being diagnosed with T2D. This is most surely due to our modern lifestyle with increased obesity and physical inactivity.

There are at least four more additional forms of diabetes; Maturity onset diabetes of the young (MODY) affects around 2% of all diabetes patients and is inherited in an autosomal dominant fashion. MODY has an early onset (before 40 years of age) and is non-autoimmune [28,29]. Other forms of diabetes is gestational diabetes which affects around 3-10% of all pregnant women depending on the studied population [30] and neonatal diabetes which can be transient or permanent. Findings suggest that neonatal diabetes does not have the same etiology as type 1 diabetes and an unbalanced duplication of paternal chromosome 6 has described as the trigger of neonatal diabetes [31]. The final diabetes form is secondary diabetes which is caused by something other than genetic factors. It is usually caused by some kind of primary health problem such as inflammation of the pancreas (pancreatitis) or cystic fibrosis. Even some medicines may interfere with insulin production (i.e decrease levels of insulin production) and there by lead to secondary diabetes (www.pamf.org/health/healthinfo).

Diabetes and pre-diabetes can be diagnosed relatively easily by performing a fasting blood glucose test at two separate occasions where a blood glucose level of 6.1mmol/l or more indicates diabetes. Diagnosis can also be made by a non-fasting blood glucose

test where a patient with blood glucose levels of >11 mmol/l is classified with diabetes (WHO 1998).

1.2.1 Type 1 Diabetes

As mentioned earlier, around 5-10% of all diabetes patients are classified as type 1 diabetics. These patients require life-long insulin injections for survival.

T1D is the only form of diabetes which is classified as an autoimmune disease. It is manifested by the loss of insulin production due to destruction of the insulin producing β -cells in the pancreas. The development of T1D in a genetically predisposed individual may take months or even up to years. It is thought that exposure to an initiating event, such as a viral infection may trigger progressive β -cell destruction and the development of auto antibodies towards the pancreatic islets. This event may not necessarily lead to T1D but in case of more triggering events cell mediated destruction of β -cells may continue finally leading to fully developed T1D (Figure 5).

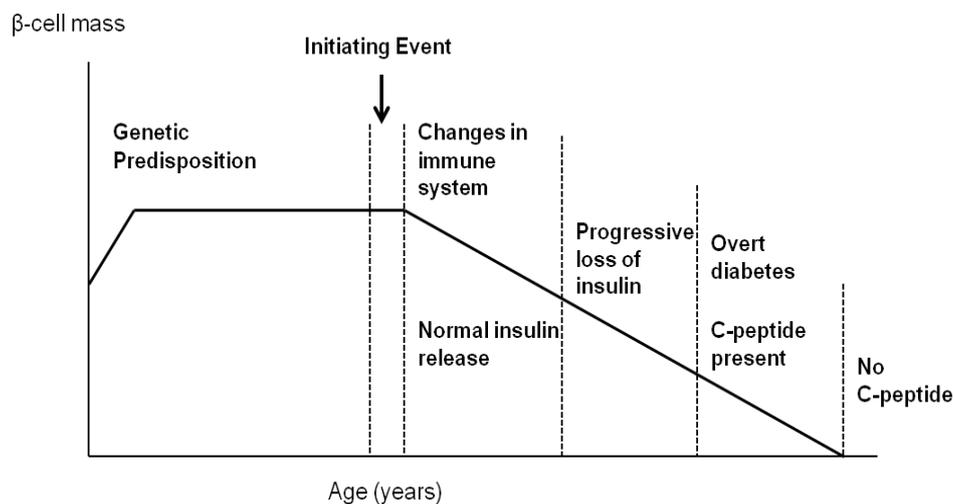


Figure 5. β -cell destruction and the stages in the development of T1D.

From Eisenberth, GS, New Engl J Med 1986; 314:1360-1368.

1.2.1.1 Symptoms and Complications

The disease progression varies greatly, sometimes ranging from just a few months to several years. Common classic symptoms are high blood sugar levels and high levels of sugar in the urine, increased appetite despite weight loss and frequent urination. Typical symptoms also include fatigue, problems in eyesight and poor healing of cuts and scrapes.

Due to greatly improved insulin treatment T1D is no longer considered to be a deadly disease. However, it has been observed that the life expectancy of T1D patients is shortened by 10-20 years due to complications. It is therefore of vital importance that all T1D patients lead a healthy life style which includes exercise, eating healthy meals and daily checks of blood sugar levels. Typical T1D complications include microvascular diseases such as; neuropathy (nerve damage throughout the body), retinopathy (damage of retina in eye), nephropathy (diabetic kidney disease) and cardiovascular damage; T1D patients may suffer from heart diseases and stroke due to high blood pressure.

1.2.1.2 Incidence and Epidemiology

The number of new cases occurring in a population during a given time period (i.e 100.000/year) is referred to as the *incidence* rate. Between 1990 and 1999 the DIAMOND project analyzed trends of new T1D cases in each continent. Results revealed that excluding Central America and the West Indies, between 1990-1999 T1D incidence cases are increasing by 2-5% world-wide. However, there are still huge variations in the incidence rates on the global scale which are thought to be due to exposure to different environmental factors as well as genetic heterogeneity. For example in China, the T1D incidence rate is around 0.1/100 000 while in Finland, which has the highest T1D incidence rate to date, it is as high as 40/100 000 [32]. Studies show that second to Finland and the island of Sardinia in Italy, Sweden has the highest T1D incidence ($\geq 20/100.000$ per year, Figure 6). Approx. 50.000 individuals in Sweden suffer from T1D today and each year, >800 new T1D cases are diagnosed [33]. Recent studies show that there is a dramatic increase in T1D cases in eastern European countries which have earlier had a rather low number of T1D cases. Poland and Romania have a yearly rise of new T1D cases of 9.3% and 8.4% respectively [34]. Furthermore, in almost all European countries is that more and more young children

between the ages 0 and 5 years are being affected by T1D. It is believed that between 2005 and 2020 there is going to be a doubling of new young T1D cases [35]. The increase in incidence rates seen in the whole world cannot be a consequence of solely genetic predisposition it must largely be due to changes in life style and environment. Gene-environment interaction seems to play a major role in the susceptibility to T1D [36] and it is these interactions that should be studied in more detail in order to understand the rise in new T1D cases.

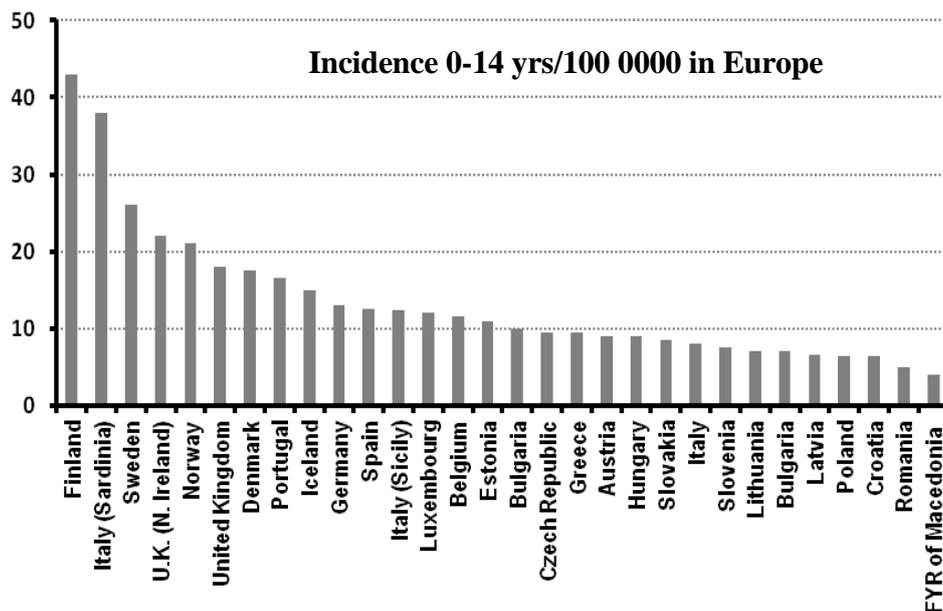


Figure 6. Geographic variation of T1D in Europe in 1989-1998 (Soltesz G et al., 2007)

1.2.1.3 Innate and Adaptive Immunity

The immune system has developed to protect the host from pathogens and other foreign substances. The early defense against foreign substances is the innate immunity and the main components include physical epithelial barriers, dendritic cells, natural killer (NK) cells and macrophages. Unlike the adaptive immune system the innate immune system recognizes structures common for various microorganisms and is therefore thought to be unspecific. Dendritic cells and macrophages also act as a link between the innate and adaptive immune system through antigen presentation to T-cells.

The adaptive immune response is antigen-specific and requires the recognition of specific “non-self” antigens during a process called antigen presentation. The adaptive

immunity also includes a “memory” that makes future response against a specific antigen more efficient. The adaptive immune system is composed of B-cells and T-cells. B-cells mature in the bone marrow and are involved in the humoral immune response through the formation of antibodies. T-cells on the other hand mature in the thymus and are involved in the regulation of the immune system. There are several types of T-cells each with a special purpose. CD4+ T-cells help B-cells in the production of antibodies and recognize peptide antigens in the context of MHC class II. CD4+ T-cells are also involved in activation of macrophages. CD8+ T-cells recognize peptide antigens presented in the context of MHC class I molecules and secrete granules containing chemicals that destroy a targeted cell and may also be involved in the activation of macrophages. Regulatory T-cells (Tregs) are subpopulations of CD4+ T-cells involved in the regulation of autoimmunity and suppression of immune response during infections. The most well characterized Tregs are those expressing CD4 and CD25 (IL2 receptor). Since activated normal CD4 T-cells also express CD25 it has been difficult to distinguish Tregs from activated T-cells. Recent research has shown that the regulatory T-cells can be defined by expression of the forkhead family transcription factor Foxp3 in addition to CD4 and CD25 [37,38]

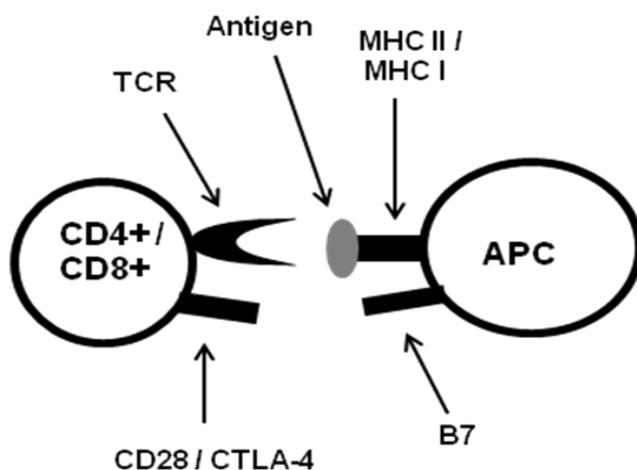


Figure 7. T-cells need two signals for activation; the binding of antigen which is presented by an antigen presenting cell (APC) on MHC I / II to the T-cell receptor (TCR) on CD8+ / CD4+ T-cells and the binding of co-stimulatory molecules B7 on APC and CD28/CTLA-4 on the T-cells. If both signals are present activation of T-cells may take place. Abnormalities in co-stimulatory molecules may lead to increased activity of auto reactive T-cells resulting in the development of T1D.

1.2.1.4 Immunologic Tolerance and Autoimmunity

The ability to discriminate self from non-self is a fundamental property of the immune system. A functional immune system requires the selection of T-cells expressing receptors that are tolerant to self-antigens. T-cell progenitors migrate from the bone marrow to the thymus where the T-cell maturation starts. Pro- T cells are called “double negative” since they express neither TCR nor the co-receptors CD4 or CD8. Some of these cells undergo rearrangement of the TCR gene segment to produce a functional TCR/CD3 complex and the cells expressing TCR develop into CD4+ and CD8+ cells (double positive cells). Those cells bearing receptors that recognize foreign peptides associated with self-MHC will be selected and allowed to mature (positive selection) and others will die by default. Next the cells become single positive. Among the positively selected cells some cells will recognize self antigens associated with self-MHC. In the following step of negative selection any cells with a high-affinity receptor for self-MHC molecules alone or self-antigen+ self-MHC are eliminated.

Unfortunately, sometimes T or B-cells manage to escape and become auto reactive cells. These cell types fail to see the difference between “self” and “foreign” substances leading to tissue damage and autoimmunity.

The presence of auto reactive T-cells and autoantibodies are typical characteristics of an autoimmune disease. There are two types of autoimmune diseases; Systemic - and Organ-specific autoimmune diseases.

Systemic Lupus Erythematosus (SLE) and Rheumatoid arthritis are two examples of Systemic autoimmune diseases. In these diseases the immune response is directed towards multiple organs and tissues with a broad range of autoantigens. Examples of organ specific autoimmune diseases include Addison’s disease, Graves’ disease and T1D [39].

In T1D the immune system attacks the insulin producing β -cells located in so called islets of Langerhans throughout the pancreas. The destruction of β -cells is thought to be caused by the infiltration of CD4+ and CD8+ T-cells and macrophages in the islets [40]. Exact details of the mechanism behind the β -cell destruction through this infiltration is however still unknown but it studies involving recent onset T1D patients and NOD mice indicate that once autoimmunity towards β -cells has been developed, β -cell autoantigens are presented to autoreactive CD4+ T-cells by macrophages, dendritic cells or B-cells in the periphery. The CD4+ T-cells then secrete cytokines which in turn activate β -cell specific CD8+ cytotoxic T-cells. The activated cytokine producing T-

cells are recruited to the pancreatic islets and further activate macrophages and T-cells which lead to β -cell apoptosis [41] (Figure 7).

Apart from auto reactive T-cells, autoantibodies are strong predictors of T1D. The first autoantibodies to be identified were the Islet cell antibodies (ICA). These autoantibodies are not antigen specific but are targeted against a variety of proteins in pancreatic islets. This discovery led to more extensive research where four distinct autoantibodies targeted against β -cell specific autoantigens were discovered; insulin (IAA) [42], glutamic acid decarboxylase 65 (GAD 65) [43], protein tyrosine phosphatase-like molecule (IA-2) [44], and ZnT8Ab [45]. It is now documented that 90% of patients with newly diagnosed T1D have autoantibodies for at least 1 autoantigen. Further, reports indicate that both IAA and IA-2 autoantibodies are found more frequently in young children [46] with a dramatic decrease during post T1D diagnosis [47]. GAD 65 autoantibodies however seem to be present for a long time period even after T1D diagnosis [48].

1.2.1.5 Disease Susceptibility Factors

More than 85% of all patients with T1D do not have a positive family history for the disease. Yet the mean prevalence (percentage of population with disease at given time) of T1D in siblings is around 6% while in the general population it is only around 0.4% indicating that there is significant familial clustering (λ) of T1D. The familial clustering for siblings (λ_s) is calculated as the disease prevalence in siblings divided by the prevalence in the general population ($6/0.4=15$). This means that siblings of T1D patients have a 15-fold higher risk of developing T1D as compared to the general population [49].

Understanding the role of environmental factors as well as genetic factors in the development of multifactorial diseases has not been easy. Twin studies have been important for distinguishing between hereditary and environmental factors in diseases such as T1D. Studies show that the concordance rate for T1D in MZ twins is between 30-50%. These are twins who have almost identical genetic information. In DZ twins who only share their genetic information up to 50% the concordance rate is only around 16% [50,51,52]. This is a clear indication that genetic predisposition has a major role in disease susceptibility. The concordance rate in MZ twins is not a 100%

and it is therefore believed that factors in the environment also play a major role in the development of T1D [50,53].

1.2.1.5.1 Environmental Factors

Environmental factors are thought to approximately account for 50% of the risk in T1D susceptibility. Due to the vast number of possible environmental factors involved in disease development, little progress has been made in identifying them. Many studies indicate that microbes, viruses, environmental toxins and dietary factors are all somehow involved in triggering T1D development. The “Hygiene hypothesis” however, still remains prominent. It suggests that in the modern society and developed countries the lack of viral as well as parasite infections early in life results in lower frequencies of protective antibodies. This in turn may lead to severe infections later thereby triggering autoreactive cells in the body [54].

Viruses; It has long been speculated that viral infections may be involved in triggering T1D. Congenital rubella infections were long considered to be the main viral infections being involved in T1D progression. Around 20% of all infants infected with congenital rubella infection, develop T1D later in life [55]. The increase of T1D incidences cannot be solely explained by the rubella virus since it has been eradicated in high incidence countries like Finland and Sweden [56,57].

Enterovirus infections have been implicated in early T1D development in children [58,59]. Traces of enterovirus RNA in sera of T1D patients and prediabetic children suggest that having enterovirus in the serum is a T1D risk factor [60].

An additional T1D associated virus is the rotavirus. Rotavirus infection is the main cause of gastroenteritis among children worldwide and it has been seen that blood antibodies directed against the virus is associated with the findings of islet cell antibodies [61].

According to above mentioned studies, it can be concluded that viral infections may be associated with T1D development. The β -cell destruction caused by these viral infections depends on the strain of the virus as well as host genetics.

There are two common hypotheses for β -cell destruction. Either the β -cells are destroyed in a direct manner through cytolysis [62] or by the involvement of the

immune system where during pancreatic tissue damage, β -cells release islet antigens that are presented to autoreactive T-cells (which in turn trigger T1D) [63]. Further, it is known that more patients are diagnosed with T1D during the winter than during the summer season [64,65,66]. This can be explained by the increased number of viral infections during the winter months. Viral infections lead to increased sugar level in the blood due to stress. This may cause extra burden on the already damaged β -cells leading to insufficient insulin production and diabetes symptoms.

Seasonal variation and Dietary products; An important factor considered to trigger the development of T1D is seasonal variation. Countries like Sweden and Finland have significantly less day light during winter as compared to the summer period leading to insufficient vitamin D production. Vitamin D is synthesized in the skin through exposure to sunlight. It has been suggested that vitamin D supplementation in infants and young children may reduce the number of T1D cases [67]. This is probably not the only explanation for high incidence rates for young children in countries like Sweden and Finland where a major part of infants are given oral vitamin supplementation daily. It cannot be excluded that vitamin D may have a protective role against T1D [68,69]. Low vitamin D levels may be part of the reason why there is a high prevalence of T1D among older children in the Nordic countries. Therefore it can be speculated that the concentration of vitamin D supplementation given in the Nordic countries should be increased in order to gain protection. The seasonal variation could also be due to variation in infections as mentioned above.

Several dietary products have also been suggested to be involved in triggering T1D. High correlations between high consumption of cow's milk and T1D incidence have been observed [70,71]. Although it is believed that this association may mainly be observed in genetically predisposed patients. Further, children that are breast fed for approximately a year have a significantly lower risk of developing T1D as compared to non breastfed children [72,73]. This suggests that breast feeding is protective against T1D and that an early exposure to foreign proteins affects the development of the immune system in such a way that autoimmunity may be favorable later in life. Additional dietary products which have been linked with T1D susceptibility are; gluten, coffee, tea, meat and sugar [74,75,76]. Also, obesity and rapid weight gain early in life

have been seen to be associated with T1D development and could in part explain the increased incidence of T1D [77].

1.2.1.5.2 Genetic Predisposition

Due to the complex nature of T1D it is impossible to identify only one single T1D affecting gene. Studies indicate that a number of genes are involved in the development of T1D directly as well as through interaction.

Researchers around the world have managed to identify and reconfirm the involvement of several genes and loci with T1D development.

HLA association; In the 1970's it was discovered that the *human leukocyte antigen (HLA) class I* locus, located on chromosome 6 is associated with T1D. It was however later seen that the *HLA class I* is in strong LD with *HLA class II* and the strongest association of type 1 diabetes was in fact to *HLA class II* [78,79]. This extremely complex locus including linked gene clusters which are highly polymorphic, is thought to account for almost 50% of the genetic risk for T1D [80,81]. The *HLA class II* molecules are located on the surface of antigen presenting cells (APC's) with the function to present foreign antigen peptides to CD4+ T-cells. The *HLA class II* locus is divided into three specific gene regions; HLA-DR, HLA-DQ and HLA-DP each showing high polymorphism. Further, studies have identified three distinct *HLA class II* haplotypes which are involved in the development of T1D [39]. The DR-DQ haplotypes that show the strongest T1D risk, accounting for 30-50% of all genetic risk to T1D, are *DR3-DQA1*05:01-DQB1*02:01* (DR3) and *DR4-DQA1*-03:01-DQB1*03:02* (DR4) [82]. In the general population around 40% carry one or two of the two high risk T1D haplotypes DR3 and DR4. On the other hand the DR3 and/- or DR4 haplotypes are found in 90% of all children affected with T1D [83]. Additionally, individuals carrying both DR3 and DR4 haplotypes have an even more increased risk of developing T1D. Around 30-40% of all T1D patients carry both DR3 and DR4 alleles whereas this combination is only found in 2.4% of the general population [84]. Children carrying both DR3 and DR4 usually have a very early T1D onset [85]. Conversely, the *DR15-DQA1*01:02-DQB1*06:02* (DR15) which is found in less than 2% of all T1D cases vs. 40% of general population, is dominantly protective against T1D [86]. The DR15 allele seems to be especially protective in young patients

suggesting that it protects from early onset of T1D [87]. How the different *HLA* molecules affect T1D is unknown but the hypothesis is that they bind more effectively to some antigens compared to others.

Insulin gene; Polymorphisms in the *insulin gene (INS)* area which is located on chromosome 11p15 have been studied thoroughly and its involvement in T1D susceptibility is widely accepted. How the associated polymorphisms exactly influence the etiology of T1D is however not yet understood. Studies show that *INS* contributes to T1D susceptibility by around 10% [88].

The *INS* gene has a locus of variable number of tandem repeats (*INS VNTR*) located 596bp upstream of the insulin gene translation initiation site [89]. The 14-15bp long consensus repeated sequence is; 5'-ACAGGGGTGTGGGG-3' and varies in numbers of times it is repeated [90]. The short VNTR class I form consisting of 28-44 repeats is believed to be associated with T1D susceptibility while the long VNTR class III form consisting of 138-159 repeats is associated with protection to T1D [88]. Studies indicate that the VNTR class III form is strongly associated with increased expression of thymic insulin mRNA. It is therefore speculated that during maturation of the T-cells and the immune system, the increased insulin levels leads to the deletion of insulin specific (autoreactive) T-cells and thereby protect against T1D development [91,92].

Additional T1D susceptibility genes; Excluding *HLA class II* and *INS* genes, researchers have managed to identify several more T1D susceptibility genes and gene regions (Table 1).

One important T1D susceptibility gene is the *cytotoxic T lymphocyte antigen 4 (CTLA4)* gene located on chromosome 2q33. Several other autoimmune diseases such as Graves' disease, Hashimoto's thyroiditis [93] and Addisons disease [94] show association to *CTLA4*. The *CTLA4* gene is expressed on the surface of activated T-cells and is homologues to CD28 molecules. *CTLA4* is thought to play an important role in immune regulation. Unlike with CD28 the binding of B7 to *CTLA4* leads to a down regulation of the immune response [95].

The *non-receptor type 22 (PTPN22)* gene located on chromosome 1p13 is in addition to T1D [96] also associated with rheumatoid arthritis [97], and systemic lupus erythematosus [98]. The lymphoid-specific phosphatase (LYP) is encoded by the

PTPN22 gene. LYP is believed to be involved in preventing T-cells to become spontaneously activated by dephosphorylating and by inactivating T-cell receptor-associated kinases [99,100].

The *interleukin-2 receptor α chain (IL2RA)* on chromosome 10p15 shows significant association to T1D [101]. The *IL-2* receptor complex has an α chain called CD25 and the receptor complex is expressed on activated T-cells and T-regulatory cells. The growth and survival of T-regulatory cells strongly depends on the expressed *IL2RA α* molecules [102]. It is thought that differences in circulating *IL2RA α* concentrations somehow leads to a functional defect in the T-regulatory cells leading to increased risk of getting various autoimmune diseases [103,104]. However, details of how *IL2RA* is associated with T1D are still unknown. Polymorphisms in *IL2RA* are also associated with Multiple Sclerosis (MS). It has been reported that there is at least one common SNP associated to both T1D and MS, while one SNP shows opposite association to both diseases and a third one only shows association to T1D [105].

Further, recently discovered T1D susceptibility genes include *IFIH1* on chromosome 2q24 [101,106] and *CLEC16A* on chromosome 5q14 [106,107]. Further, the *DLK1* gene located on an imprinting region on chromosome 14q32 has been seen to be associated with T1D [108]. Moreover, studies including genome wide association (GWAS) studies have located more than 40 additional areas in the genome which are thought to be associated with T1D susceptibility (Table 1).

The above mentioned genes are generally believed to be “true” T1D susceptibility genes since their association has been confirmed in multiple studies. Many more areas in the genome will probably be identified and confirmed as being involved in T1D development.

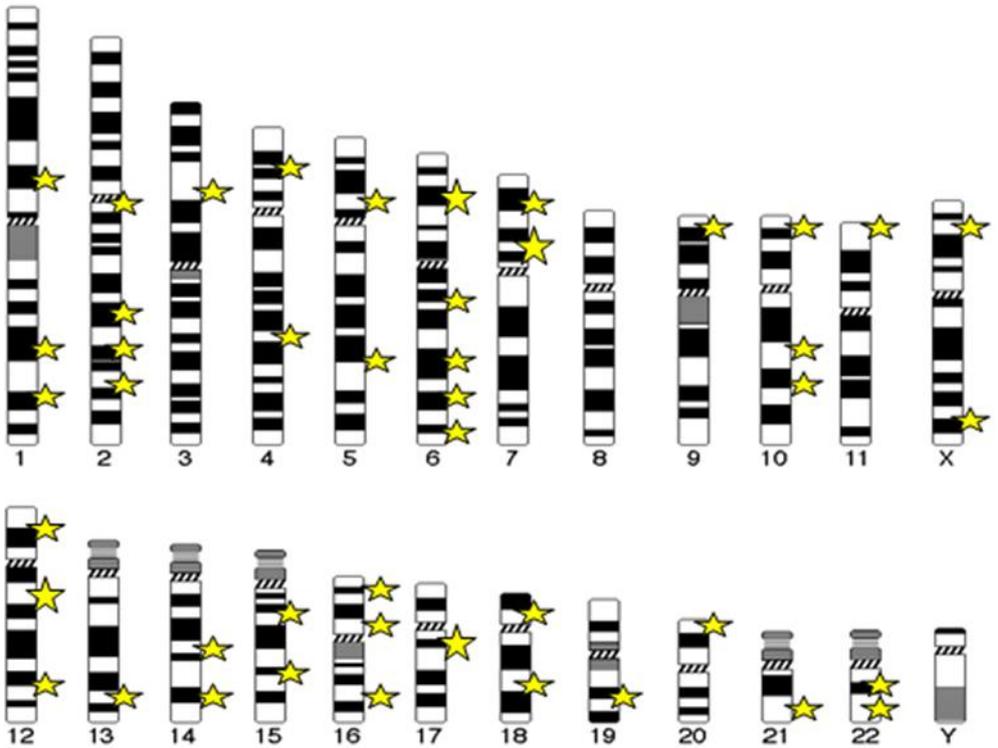


Figure 8. T1D susceptibility regions. Stars represent regions which show evidence of association to T1D.

Table 1. T1D susceptibility regions

Chromosome position	Gene name	Marker	OR (95% c.i.)	P-values	Reference
1p13	<i>PTPN22</i>	rs2476601	1.7	2.1 x 10 ^{-80*}	[107,109,110,111]
1q31.2	<i>RGS1</i>	rs2816316	0.9	3.1 x 10 ⁻⁵	[110,112]
1q32.1	<i>IL10</i>	rs3024505	0.8	2.2 x 10 ⁻⁶	[112]
2q11.2	<i>AFF3-LOC150577</i>	rs9653442, rs1160542	1.1	7.0 x 10 ⁻⁷ , 7.2 x 10 ⁻⁷	[107,112]
2q24.2	<i>IFIH</i>	rs1990760	0.9	1.8 x 10 ^{-11*}	[101,107,113]
2q32.2	<i>STAT4</i>	rs6752770	1.1	9.3 x 10 ⁻⁶	[112]
2q33	<i>CTLA-4</i>	rs3087243	0.9	7.4 x 10 ⁻⁴	[101,114,115]
3p21.31	<i>CCR5</i>	rs11711054	0.8	1.7 x 10 ⁻⁵	[112]
4p15.2		rs10517086	1.1	2.8 x 10 ⁻⁷	[112]
4q27	<i>Tenr-IL2-IL21</i>	rs17388568	1.1	2.9 x 10 ⁻⁴	[101,107]
5p13	<i>IL7R</i>	rs6897932	0.9	7.8 x 10 ^{-6*}	[101,107]
5q14	<i>KIAA0305</i>	Rs12708716	0.8	7.1 x 10 ⁻⁹	[107]
6q15	<i>BACH2</i>	rs11755527	1.1	5.4 x 10 ⁻⁸	[112,113]
6p21	<i>HLA-DRB1</i>	HLA	7.0	4.9 x 10 ⁻³²	[107],
6p21.3	<i>B*5701</i> <i>B*3906</i> <i>A*1101</i>	HLA	0.2 10.0 0.3	4 x 10 ⁻¹¹ 4 x 10 ⁻¹⁰ 5 x 10 ⁻⁸	[116]
6q22	<i>CENPW</i>	rs9388489	1.2	5.1 x 10 ⁻⁸	[112]
6q23	<i>TNFAIP3</i>	rs6920220	1.1	8.0 x 10 ⁻⁴	[112]
6q25	<i>TAGAP</i>	rs1738074	0.9	6.0 x 10 ⁻³	[112]
7p12.2	<i>IKZF1</i>	rs10272724	0.8	1.4 x 10 ⁻⁶	[112]
7p15.2	<i>C7orf71</i>	<u>rs7804356</u>	0.9	3.3 x 10 ⁻⁸	[112]
9p24	<i>GLIS3</i>	rs7020673	0.9	1.9 x 10 ⁻⁹	[112]
10p11	<i>NRP1</i>	rs2666236	1.1	9.8 x 10 ^{-6*}	[107],
10p15	<i>IL2RA</i> (<i>CD25</i>)	rs12251307	0.8	6.5 x 10 ⁻⁸	[107],
10q22	<i>ZMIZ1</i>	rs1250558	0.7	8.0 x 10 ⁻⁴	[112]
10q23	<i>RNLS</i>	rs10509540	0.6	6.9 x 10 ⁻⁹	[112]
11p15	<i>INS</i>	rs3741208/ rs689	2.0	7.4 x 10 ⁻⁴ / 3.8 x 10 ⁻³¹	[88,107,117]
12p13	<i>CLEC2D</i>	rs3764021	0.9	4.8 x 10 ⁻⁵	[101,112]
12q13	<i>ERBB3</i> <i>CYP27 B1</i>	rs2292239 rs10877012/ rs703842	 1.2	1.5 x 10 ⁻²⁰ 9.1 x 10 ^{-5*} / 9.5 x 10 ⁻³	[69,101,107,112]
12q24	<i>C12orf30</i>	rs17696736/ rs3184504	1.2	2.3 x 10 ⁻¹⁶ / 2.8 x 10 ⁻²⁶	[101,107,112]
13.23	<i>UBAC2</i>	rs9585056	1.2	2.1 x 10 ⁻⁵	[112]
14q24	<i>ZFP36L1</i>	rs1465788	0.9	1.4 x 10 ⁻⁸	[112]
14q32	<i>C14orf64</i>	rs4900384	1.1	1.1 x 10 ⁻⁶	[112]
15q14	<i>RASGRP1</i>	rs17574546	1.2	8.1 x 10 ⁻⁵	[112]
15q25	<i>CTSH</i>	rs3825932	0.9	7.7 x 10 ⁻⁸	[112]
16p13	<i>CLEC16A</i>	rs12708716	1.1	2.2 x 10 ⁻¹⁶	[112]
16p11	<i>IL27</i>	rs4788084	0.9	5.2 x 10 ⁻⁸	[112]
16q23	<i>CTRB1</i>	rs72082877	1.3	5.7 x 10 ⁻¹¹	[112]

17q21	<i>SMARCE1</i>	rs7221109	1.0	9.9×10^{-10}	[112]
18p11	<i>PTPN2</i>	rs2542151/ rs478582	1.3	$1.9 \times 10^{-6}/$ 2.2×10^{-12}	[101,107,112]
18q22	<i>CD226</i>	rs763361	1.2	1.3×10^{-5}	[112]
19q13	<i>PRKD2</i>	rs425105	0.9	1.5×10^{-7}	[112]
20p13	<i>SIRPG</i>	rs2281808	0.9	5.0×10^{-7}	[112]
21q22	<i>UBASH3A</i>	rs3788013	1.1	2.1×10^{-6}	[112,113]
22q12	<i>HORMAD2</i>	rs5753037	1.1	1.8×10^{-14}	[112]
22q13	<i>IL2RB</i>	rs3218253	1.0	2.5×10^{-5}	[112]
Xp13-p11		DXS1068	0.9	2.7×10^{-4}	[118]
Xp22	<i>TLR7</i>	rs5979785	0.8	6.7×10^{-6}	[112]
Xq28	<i>GAB3</i>	rs2664170	1.2	3.0×10^{-5}	[112]

Genomic regions and genes which are associated to T1D.

* = over-all p-values

1.2.1.6 Animal Models

When studying different diseases, animal models act as important tools in understanding biological mechanisms, dysfunctions caused by diseases and in the development of potential new therapeutics. For instance, the discovery of insulin therapy was first thoroughly tested on dogs.

The non-obese diabetic (NOD) mice and the Biobreeding rat (BB rat) are often used as animal models for T1D studies for many years. These models have similar pathogenic and genetic T1D conditions as humans and are therefore of great use in understanding the causes and progression to T1D [119]. The NOD mouse strain was first established by a Japanese research group over 20 years ago. Repetitive sister-brother mating lead to a mouse strain which spontaneously develops diabetes [120]. The BB rat is a rat strain which was similarly inbred in the laboratory and develops T1D spontaneously just like NOD mouse [121]. Among many genes, the *ctla4* gene region and *mhc II* genes have been identified as T1D susceptibility genes in the NOD mouse and BB rat respectively [122,123].

2 STUDY AIMS

- I. To study the association of the *Immunoreceptor PD-1 (PDCD1)* gene to T1D in a Swedish case-control material and families.
- II. To narrow the previously T1D linked region (5p13-q13) in Scandinavian materials. As well as identify new T1D susceptibility genes in this region.
- III. To test the hypothesis that the *MHC class II transactivator (CIITA)* is a T1D susceptibility gene. Further, to study how the association is affected by differences allele distribution depending on age of controls.
- IV. To evaluate four different statistical methods for studying interaction by determining model and prediction accuracy by applying them to the study of interaction between *HLA*, *INS* and *PTPN22* in Swedish case-control materials.

3 MATERIALS

3.1 SCANDINAVIAN FAMILIES

Family materials from the three Scandinavian countries Sweden, Norway and Denmark have been included in these studies.

The Swedish families (paper I and II) consist of 184 multiplex and 8 simplex families and were collected from two nationwide T1D incidence registries (the Swedish Childhood registry [124] and the Diabetes Incidence Study in Sweden [125]). In the multiplex families, at least 2 siblings were affected with T1D. The family material includes a total of 200 T1D affected sib-pairs.

The Norwegian families (paper II) were recruited through advertisements in the Journal of the Norwegian Diabetes Association. A total of 77 multiplex and 2 simplex families including 89 affected sib-pairs were collected.

The Danish families (paper II) consisting of 147 multiplex and 5 simplex families were collected through the Danish Study Group of Diabetes in Childhood (DSGD) and the Danish IDDM Epidemiology and Genetics Group (DIEGG). A total of 175 affected sib-pairs were included in the cohort [126].

Patients were classed as T1D affected based on two criteria 1) if they had an early onset of diabetes (diagnosed before the age of 15 years), required insulin treatment at the time of onset and remained on the treatment afterwards or 2) suffered from ketoacidosis or required insulin treatment from onset while being T1D diagnosed after the age of 15 years.

3.2 SWEDISH PATIENTS AND CONTROLS

The Diabetes Incidence study in Sweden 1 (DISS1) (paper III and IV) material was collected from the Diabetes Incidence Study in Sweden (DISS) registry [127]. The patients were between the ages of 15-34 years, having been diagnosed with T1D between 1987 and 1989. The cohort consists of 702 patients and 618 sex, age and residence matched controls [128]. The patients were classified into four separate groups; T1D, type 2 diabetes (T2D), unclassified patients and secondary diabetes using WHO criteria by the treating physician. In our studies (*Paper III and IV*) the patients diagnosed with T1D were included in genetic analysis.

The Diabetes Incidence study in Sweden 2 (DISS2) (paper I-IV) material was collected during 1992 - 1993 from the Swedish registry for newly diagnosed diabetes patients between the age of 15 and 34 years. Diagnosis was in the same way as in the DISS1 material. Additionally, patients were followed with yearly classifications and measurements of C-peptide for 6 years. A total of 778 T1D patients and 836 matched controls were collected [127].

The Swedish childhood Study (SV2) (paper III and IV) cohort consisting of 494 cases was collected between 1986 and 1987 from the Swedish Childhood registry [124] which is a registry for young children between the ages of 0-14 years diagnosed with T1D. For all cases above the age of 7 years, geographically, age and gender matched controls were collected. Ethical reasons did not allow us to contact control groups for the cases under the age of 7 years. Instead, 53 children under the age of 7 years who were being treated for other reasons than T1D were collected as matched controls [129].

The Better Diabetes Diagnosis Study (BDD) (paper II-IV) material was recruited from 40 pediatric clinics in Sweden between May 2005 and September 2009. The cohort consists of 2700 incident patients under the age of 18 years at the time of diagnosis. At follow up, 95% of all patients were classified as T1D by the treating physician. The remaining 5% were classified as “other” type [130]. Only individuals of self reported Scandinavian origin were included in the studies.

Diabetes Registry in Southern Sweden (DR) (paper III) material consists of 804 patients (436 males and 368 females) with T1D with an age range between 1 and 75 years from the Diabetes Registry in Southern Sweden, all enlisted at the Dep. of Endocrinology at Malmö University Hospital, Sweden and collected between 1996 and 2005. The material also consists of 2312 healthy controls between the ages of 45 and 75 years (1695 males and 617 females) [131]. All individuals of self reported non-Scandinavian origin were excluded.

Controls in Multiple Sclerosis (MS) study (paper II and III). 1215 healthy blood donors are included as a control group in the Multiple Sclerosis Swedish cohort 1. All blood donors originate from Sweden or the Nordic countries and were resident in the Stockholm region [132]. In the Multiple Sclerosis Swedish cohort 2, 663 age, sex and residential area matched controls in incident MS cases throughout Sweden were included as part of the EIMS study [133]. In our study, all controls of self reported non-Scandinavian origin were excluded. All MS controls were used in paper III while in paper II only 527 healthy controls were used.

Controls in Rheumatoid Arthritis (RA) study (paper III). The 1426 controls included as controls in the EIRA case control study of RA were matched to RA patients by residential area, gender and age as described earlier [134]. 97% of the controls were of self reported Caucasian origin. In *paper III*, 373 of the controls were used in the SNP tagging analysis.

Controls in Myocardial Infarction (MI) study (paper III). The control group included in the MI cohort consists of 387 age and sex matched healthy individuals between the ages of 40-60 years. All controls were of self-reported Caucasian origin [135].

Controls in Alzheimer's disease study (paper III). This group of controls consists of 463 healthy individuals between the ages of 56-91 years collected from the longitudinal study; The Swedish National Study on Aging and Care in Kungsholmen (SNACK), in Stockholm, Sweden [136].

Population based control cohorts from Osteoporosis study (paper III).

Between 1999 and 2003 1005 healthy women were selected from the Malmö city files to be included in the PEAK-25 study. All women were 25 years of age and of Swedish or North European ancestry [137].

Further, 1010 healthy controls were collected from the Malmö Osteoporosis Prospective Risk Assessment (OPRA) study. All controls had been randomly selected from the Malmö city files between 1995 and 1999. All controls were 75 years old and of Swedish or North European ancestry [138].

3.3 INDIVIDUALS FOR IMMUNOHISTOCHEMISTRY AND EXPRESSION STUDIES (PAPER II)

55 human donors of pancreatic islets were obtained from Lund University Diabetes Center (LUDC) Human Tissue Laboratory. No donors were positive for GAD antibodies nor were they diagnosed with T1D.

4 METHODS

4.1 FINE MAPPING OF CHROMOSOME 5 (*PAPER II*)

43 microsatellites selected from NCBI (<http://www.ncbi.nlm.nih.gov/>), GDB (<http://gdbwww.gdb.org/>), and Marshfield (<http://research.marshfieldclinic.org/genetics/>) were used for fine mapping of the chromosome 5 region. Amplification (PCR) of microsatellites using fluorescence labeled primers was first done on the PTC-225 thermocyclers (MJ Research). PCR products from 10-12 different primer pairs were pooled prior to separation. Separation was carried out on denaturing 6% polyacrylamide gels using the ABI373 and ABI377 (Applied Biosystems) respectively. GENESCAN (Applied Biosystems) version 1.2 or version 3.1 was used for final analysis (for further details about PCR conditions and separation conditions see article by *Nerup et al, 2001*).

4.2 SNP GENOTYPING

All used SNPs were selected from NCBI's dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), The SNP Consortium (<http://snp.cshl.org/>) or from previous publications. SNPs with minor allele frequencies of at least 0.3-0.5 were chosen for analysis. A lower allele frequency was only accepted for SNPs within genes for paper II. Average spacing between SNPs was 10-30kb. For paper II, all chosen SNPs were quality controlled, which included BLASTing of SNPs against the whole human genomes using the BLAST program in NCBI. Further quality controls included checking for primer dimers (which could prevent the sequencing process) and looping (which can cause sequencing of the primer itself).

Taqman (paper I-IV); For SNP analysis with the Taqman method, all primers and probes were obtained from Applied Biosystems (Applied Biosystems Inc., Sweden) and genotyped as described [139]. The fluorescence intensity in each well of the plate was read using the SDS 2.2.1 Sequence detection system program in the Taqman 7900HT machine (Applied Biosystems Inc., Sweden). The Taqman analysis consists of two probes with a fluorophore and quencher attached to it respectively. The quencher

inhibits a fluorescing signal from the fluorophore until it is released from the probe due to degradation by the Taq polymerase.

Pyrosequencing (paper I and II); For determining SNP genotypes run using the pyrosequencing method the standard protocol from the vendor was used. Final analyses was carried out using version 1.1 and 1.2 AQ of the PSQTM96 SNP Software.

Pyrosequencing involves sequencing of one strand and the synthesis of the complementary strand enzymatically. Nucleotides are then added one at the time and the “right” strand can be determined by emission of light.

Dynamic allele-specific hybridization (DASH, paper III); SNPs genotyped using the DASH method were typed according to manufacturers recommendations [140]. DASH is a method where the melting point of DNA is studied for determining a SNP. Through capturing a biotinylated DNA strand of interest by using streptavidin an allele specific oligonucleotide is added. When bound to another strand the nucleotide emits light which intensity is measured as the temperature is increased and the melting point can be determined. A lower than expected melting point will be observed for a SNP.

MassArray chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (Sequenom., San Diego, CA, USA) or IPLEX (Sequenom., San Diego, CA, USA, *paper IV*) were typed according to manufacturers recommendations [141,142] (*paper III*). The MassArray method is based on the fact that different strands including SNPs have different masses and can thereby be distinguished from each other. Up to 40 SNPs can be analyzed in one individual cocktail.

PCR-RFLP (paper I); For polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) studies, restriction enzymes *MspI* (TaKaRa Inc., Shiga, Japan) and *PstI* (Biolabs Bio Inc., Ipswich, MA, USA) were used. PCR-RFLP is carried out by amplification of the region of interest and then adding a restriction enzyme (on SNP location). The enzyme then cleaves the strand and results can be detected by running a gel electrophoresis. Three possible bands will appear on the gel depending on whether the restriction enzyme has cleaved the strand or not.

PCR conditions were used according to vendor’s recommendations.

4.3 SEQUENCING OF THE *HTR1A* GENE (*PAPER II*)

Sequencing of the *HTR1A* gene and along with its upstream and downstream regions was carried out using the ABI Prism 3730 Genetic Analyzer method (Applied Biosystems Inc., Sweden). Primer pairs for each ~800 -1000bp fragment (twelve primer pairs) were designed by the Primer express computer program. All primers were first optimized followed by PCR amplification. Sequencing was carried out according to ABI recommendations and results were analyzed in the SeqScape software version. 2.5 (Applied Biosystems Inc., Sweden). Sequencing in ABI Prism 3730 is carried out by a PCR reaction including a sequencing primer (forward and reverse separately in separate runs) and dNTPs with fluorescence molecules. Once a sequence has been established all fragments are separated on an acrylamide gel where the intensity of the individual fluorescence colors is read and the sequence of the fragment may be determined.

4.4 HLA TYPING (*PAPER III AND IV*)

HLA typing has been performed in six of our used materials (DISS1, DISS2, SV2, BDD, MS and RA);

For DR typing in DISS1 and SV2 restriction fragment-length polymorphism (RFLP) was used. Typing of DQB1, DRB1 and DQA1 and allelespecific amplification for DR7 and DR9 was carried out using PCR amplification followed by dot blot hybridizations [143].

DQA1, DQB1 and DRB1 in the DISS2 cohort were typed using PCR amplification followed by dot blot hybridizations and by RFLP as previously described [143]. For DRB1 alleles, allele-specific PCR amplification (PCR-SSP) was also used [144].

For HLA genotyping in the BDD cohort, a method which is based on an asymmetrical PCR followed by a subsequent hybridization of allelespecific probes was used [145]. In order to determinine DR genotypes for BDD individuals where only DQA1 and DQB1 were genotyped established haplotypes in the european population were used.

HLA typing of the MS and RA individuals was carried out using allele specific amplification as described previously [144].

4.5 IMPUTATION (PAPER II)

In paper II imputation of three SNPs among the MS controls was performed using the MaCH v. 1.0.16 analysis program with default settings [146]. Results from previous SNP genotyping in a MS Genome wide association (GWAS) study were used. Typed SNPs were quality checked for call-rates (per sample and per SNP), heterozygosity, recent shared ancestry, non-European ancestry, MAF and deviation from Hardy-Weinberg equilibrium [147].

4.6 QUANTITATIVE-PCR OF *HTR1A* AND *RNF180* MRNA FROM HUMAN ISLETS OF LANGERHANS (PAPER II)

A total of 10 healthy age and gender matched donors were used for preparing mRNA of human pancreatic islets (mRNA Easy Plus mini kit, Qiagen, Hilden, Germany). Reverse transcription was performed to obtain cDNA (Maxima™ First Strand cDNA Synthesis Kit, Fermentas, Thermo Scientific, Sweden). Quantification of mRNA levels was done using a Probe/ Rox Real-Time PCR (Maxima™ Probe/ROX qPCR Master Mix (2X), Fermentas, Thermo Scientific, Sweden) with an ABI PRISM 7900 (Applied Biosystems, Inc., Sweden), and assays-on-demand were employed for *HTR1A* (Hs 00265014) and *RNF180* (Hs 00400379) (Applied Biosystems Inc., Sweden). All samples were run as duplicates and normalization of the transcript quantity was done using the mRNA level of *cyclophilin A*, *polymerase 2* and *hypoxanthine guanine phosphoribosyl transferase* (Applied Biosystems).

4.7 TISSUE PREPARATION AND IMMUNOHISTOCHEMISTRY (PAPER II)

Immunohistochemistry is an excellent method for localizing certain proteins or markers and understanding their distribution in cells of various tissues. The basic principle is that one directs antibodies towards the specific antigens of interest in the tissue or cells.

We have in our study used the *indirect* form of *immunohistochemistry*. This method involves a primary antibody which reacts with tissue antigen and secondary antibody which in turn reacts with the primary antibody.

Isolation of human pancreatic islets was performed through fixation of human pancreas biopsies. The specificity of immunostaining was tested using primary antisera pre-absorbed with homologous antigen. Immunofluorescence was finally

examined in an epi-fluorescence microscope (Olympus, BX60). The location of different antibodies in double staining was determined by changing filters. Images were captured with a digital camera (Nikon DS-2Mv).

4.8 STATISTICAL ANALYSIS

Linkage analysis (paper I and paper II)

Mapping of microsatellites to the chromosome 5p13-q13 region was done using the *CRIMAP* program [148].

The *Allegro* program version 1.2 [149] designed for non-parametric linkage analysis was used for calculating linkage. For both single and multipoint analysis, the Exponential model with an equal weighing was used.

The *Linkage and Association Modeling in Pedigrees (LAMP)* software program was used to combine Linkage and association analysis. The program was also used for identifying SNPs responsible for linkage in the Danish and Swedish families in paper II.

Linkage Disequilibrium (LD, paper II and paper III)

In paper II the bioinformatic software program *Haploview version 4.2* [150] was used for identification of LD structure and tagging SNPs (htSNPs) using observed and HapMap genotype data.

In *paper III* identification of LD blocks and htSNPs was done using the *HapBlock analysis program* [151,152]. The common haplotype method was used for block partitioning and the capability of the method to identify all common haplotypes using the htSNPs was > 5%.

Association

The *Unphased* program (version 3.0.6) [153] was used for studying the association in cases and controls and the PDT test as implemented in Unphased was used for studying association in families (paper I and paper II). The Unphased program was also used for calculating conditional analysis and haplotype association (paper I, II, III and IV). When analyzing more than one cohort the study cohorts were used as confounders.

The *zGenStat 1.126* program (Henric Zazzi unpublished) was used to check mendelian inheritance of the genotypes on the multiplex/simplex families. Any inconsistencies were removed (paper I and II).

Power calculations assuming an unmatched retrospective design were performed using the PS software [154] program (paper I).

Meta analysis (paper I and III) is a method to exclude heterogeneity between separate studies. This is done by comparing and quantifying the detected effects in different studies. We have performed meta analysis with the Mantel–Haenszel summary method for estimating an overall odds ratio (OR) and Woolf’s method for testing for heterogeneity [155] in the *rmeta* package 2.14 in R [156].

Genotype association in materials including cases and controls were performed with standard Pearson's Chi-squared test.

Logistic regression analysis using generalized linear modeling was used to correct for effects of age, HLA and gender when allele frequencies between diabetic and non-diabetic individuals were compared. The analysis was carried out in the statistical computer program R (*R version 2.6.2, The R Foundation for Statistical Computing, <http://www.r-project.org/foundation>, paper III and IV*).

4.9 COMPUTATIONAL ANALYSIS (PAPER II)

For detecting potential transcription factor binding sites in the *HTR1A* gene we used the web based software program RAVEN (www.cisreg.ca/cgi-bin/RAVEN/a).

4.10 ADJUSTMENT FOR AGE

When correcting for age in logistic regression, the age-group 4 (15-19 years) (*paper III*) and the age group between 10-15 years (*paper IV*) were used as a reference group based on a large and equal number of individuals of both cases and controls in these groups.

4.11 IMPUTATION (PAPER II)

In paper II, MaCH v. 1.0.16 analysis program as implemented in BC₁SNPmax v. 3.5-191 was used for imputation of SNPs in the MS controls. All settings were set to default values used in MaCH v. 1.0.16.

4.12 INTERACTION STUDIES (PAPER IV)

Chi² and logistic regression using the multiplicative model was carried out in the R program version 2.12.2 (*The R Foundation for Statistical Computing*, <http://www.r-project.org/foundation>). A generalized log-linear model using allele and genotype results was used to study the interaction between *HLA* haplotypes, *PTPN22* and *INS* by assessing the significance of the association to T1D of the interaction terms while keeping the original terms in the model [23].

The R program was also used for logistic regression when calculating additive interaction (attributable proportion, AP) with a confidence interval (CI) of 95%. In *paper III and IV*, the `vcov` command was used to get the covariance matrix.

MDR was calculated using the MDR package in the R program [157].

The Bayesian Network was run using the Hugin version 7.6 (Linux 64 bit distribution) software [158]. Both the greedy search and score algorithm were used.

5 RESULTS AND DISCUSSION

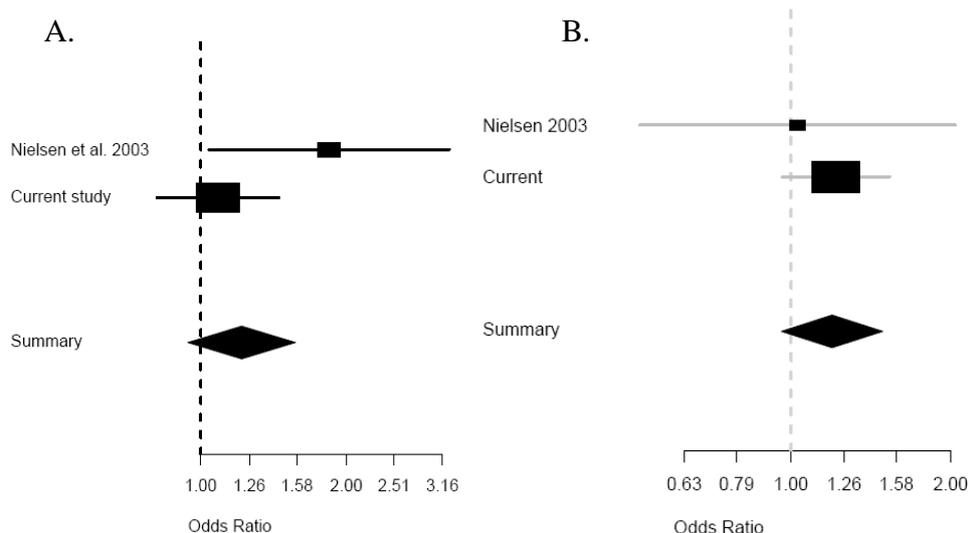
5.1 PAPER I

It has been suggested that the PD-1 7146 SNP in the *immunoreceptor PD-1 (PDCDI)* gene (chr 2) is involved in the pathogenesis of several autoimmune conditions such as SLE [159] and T1D [160]. *PDCDI* is believed to be involved in the maintenance of self tolerance and is therefore a natural candidate gene for autoimmune diseases such as T1D. We tested *PDCDI*'s linkage and association to T1D in a Swedish family material and association in a Swedish case-control (DISS2) material. Studying a gene in both family as well as case-control materials is a great advantage since both linkage and association can be measured. Linkage analysis can be carried out in families, which allows us to identify rare marker variants which have been accumulated in families rather than in the general population. On the other hand, case-control materials (which are easier to get hold of) represent SNP allele variants in the general population. Since LD analysis of the *PDCDI* gene has distinguished a block consisting of four haplotypes, we chose to genotype four tagSNPs; PD-1 7146, PD-1 7785 and PD-1 8738 (one SNP was discarded due to being non-polymorphic) in order to gain full gene information in the Swedish families and DISS2 cohort [159]. No single SNP or haplotype association was observed for any of the SNPs in the family material (Table 1, paper I). Further, when linkage analysis was carried out no linkage was observed neither in the singlepoint nor in the multipoint analysis ($LOD < 0.01$). In the DISS2 material, modest association was observed for the PD-1 7785 SNP. However, permutation tests showed that the association was insignificant after correction for multiple comparison. No association was observed for the two additional SNPs (Table 1, paper I).

There are several possible reasons to why our results did not follow the significant association that was found in the Danish material. We performed our analysis in both a family material as well as a case-control material with proper matching of controls. Also, unlike the Danish material which only consisted of two small cohorts (94 and 98 cases respectively and 155 unmatched healthy blood donors), both our materials were much larger meaning that there is a possibility that the Danish study may have got a significant association for the PD-1 7146 SNP purely by chance. Furthermore, we did

not observe any linkage to the *PDCDI* area indicating that it is not accumulated in multiplex T1D families. The *PDCDI* gene has not shown any linkage or association in large genome wide linkage and association studies. These results further support our results.

We have also performed a meta analysis including our case-control material and the previously published findings in the Danish T1D study. Meta analysis is a way of estimating the overall evidence for association in several studies and for identifying heterogeneity between two or more studies. The main downside is that there is a large publication bias due to unpublished negative data. This means that many studies which show no association may not have been published and therefore it is likely that the true association between *PDCDI* and T1D is even less strong than our meta analysis suggest. We observe no heterogeneity between our study and the Danish study for PD-1 7785 and PD-1 8738. Only slight heterogeneity was observed for the PD-1 7146 SNP. We have observed an error in the meta-analysis figures 1a-c in the published *paper I* and it should be noted that the figures in the thesis are the correct ones. The overall conclusions however, remain the same as in the published *paper I*.



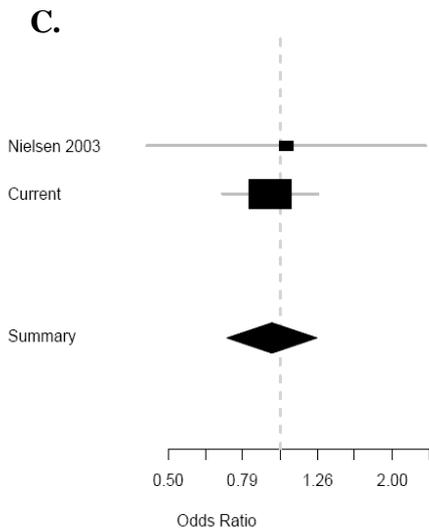


Figure 9. Meta-analysis of association of the (A) PD-1 7146G/A, (B) PD-1 7785C/T and (C) PD-1 8738C/T SNPs. The test was performed using the proportion of cases and control subjects having at least one A or T at each SNP position respectively in our current study (Swedish T1D material) and in the study by Nielsen *et al.*, (Danish T1D material)

Also, unknown gene-environment interactions and slight heterogeneity between the two populations may be the cause for the different results.

Further, it is widely understood that a significant number of T1D patients also have Celiac disease. Many of the underlying susceptibility genes are the same for the two diseases (i.e HLA-genes). It has been reported that there is no association of the PD-1 7146 SNP gene to Celiac disease in a Finnish study [161]. These results support our negative findings.

The differences between our study and the SLE [159] study may be explained by the fact that different underlying genetic factors are involved in SLE compared to T1D. Considering our results from the Swedish T1D study and comparing the results with the Danish study, we conclude that the *PDCDI* gene is not a major T1D susceptibility gene at least in the Swedish population.

5.2 PAPER II

Suggestive linkage to the chromosome 5 region (LOD<2.2) has been observed in a genome wide linkage study in Scandinavian T1D families [126]. In order to identify new T1D susceptibility genes in this area we have typed over 70 markers including microsatellites and SNPs in three separate Swedish materials as well as a Danish family material. Significant linkage (LOD 3.98 and LOD 3.66 respectively) was observed in the Swedish families for two SNPs (rs878567 and rs6295) located on either side of the *5-hydroxytryptamine (serotonin) receptor 1A (HTR1A)* gene. The rs6295 SNP also shows association to T1D ($p<0.01$). Furthermore a third SNP (rs356570) located upstream of *HTR1A* showed stronger association to T1D ($p<0.002$) than rs878567 and rs6295. *HTR1A* encodes a receptor for the neurotransmitter serotonin which has been seen to be expressed in the pancreas. Changes in serotonin signaling have been linked to several complex behaviors such as sleep, depression and suicide. It has also been reported that *HTR1A* is involved in the immune system and is involved in the regulation of T-cell activity and may thereby somehow influence the development of T1D.

In order to identify SNPs within the *HTR1A* gene, sequencing of the gene and its flanking upstream and downstream regions was carried out. No SNP in the gene was identified, however several known SNPs around the *HTR1A* gene were observed. In an analysis of LD we saw that the rs356570 SNP also tags for several other SNPs around the *HTR1A* gene as well as SNPs in the *ringfinger protein 180 (RNF180)* gene ($r^2 < 0.91$ between both genes). This discovery lead us to type three additional SNPs in and around *RNF180* in order to verify its involvement in T1D susceptibility. Positive association was observed for rs6880454 ($p<0.01$).

Further, haplotype association studies identified two haplotypes in the *HTR1A-RNF180* area (one positively associated and one protective, $p<0.002$ and $p<0.03$ respectively, Table 4, paper II). In order to confirm our findings we typed SNPs in the Swedish sporadic cases as well as Danish families. No single SNP or haplotype association was observed in the sporadic T1D cases. However, in the Danish families, positive association for rs356570 ($p<0.001$) was observed. Pooling of all four cohorts showed that both rs6295 and rs356570 were associated with T1D ($p<0.01$ and $p<0.003$ respectively, Table 3, paper II).

Moreover, in the Danish families, association was observed for the same haplotypes which were associated in the Swedish families ($p<0.05$ and $p<0.02$ respectively, Table 4, paper II).

Having detected both linkage as well as association in our Swedish families is a good indication that our finding is indeed a positive association. The lack of linkage in the Danish families may be explained by the fact that *HTR1A* only seems to have a small effect on T1D susceptibility and certain SNP variants have been accumulated more in the Swedish families. A major down side of linkage studies is that very large linked areas are identified making it a tedious task to identify susceptibility genes. The lack of association in the cases and controls which represent the general population, support the fact that *HTR1A* only has low effect on T1D development. The WTCCC study [101] and T1DGC study [112] did not detect any association of this region in their analysis. However, their studies did not include our most associated SNP, thereby making it hard to say whether or not a positive association might have been missed. The study included only a small number of families and sporadic cases were over represented. The consortium however does have a larger T1D family material and it would be interesting to run our associated SNPs in only them. Although having small effect on T1D susceptibility our immunohistochemistry results clearly indicate a presence of *HTR1A* in the pancreas. Our results from the conditional association analysis indicate that *HTR1A* is associated with T1D independently of *RNF180*. However, the effect on T1D as mentioned, is small suggesting that it interacts with additional factors. *RNF180* binds to zinc ions (Genecards homepage) and several studies indicate that zinc ions plays a major role in the synthesis, storage and secretion of insulin [162]. Additionally, it is suggested that *RNF180* could be involved in the ubiquitin-proteasome pathway, in which target proteins undergo degradation. Also, our expression analyses detect the presence of both *HTR1A* and *RNF180* in human pancreatic islets. Since *RNF180* has a rather clear effect on several important processes and is expressed in the pancreas it may be suggested that together with *HTR1A* it has a significant effect on the initiation of T1D. Since the associated SNP rs6295 located close to the *HTR1A* gene maps to binding sites of two transcriptions factors, functional studies should be carried out to see whether or not it has a clear effect on T1D initiation. This could be done by studying how the binding of transcription factors to different constructs with alternative SNP variants affect *HTR1A* expression.

In conclusion, I believe that since we were able to observe both suggestive linkage as well as association of *HTR1A* to T1D, the *HTR1A* gene is a T1D susceptibility gene at least in our T1D families. Further, we observe *HTR1A* expression in the pancreas which is the target organ in T1D. Although the effect on the risk of T1D is modest and it can

therefore not be ruled out that the associated SNPs in *HTR1A* affects transcription of nearby genes such as *RNF180*. Therefore expression and functional studies are an important next step in our study. By doing this we may get further insight to why and how these two genes are involved in the development of T1D.

5.3 PAPER III

When doing research studies, sometimes one discovers something unexpected and that is what happened to us in paper III. Our main aim of the paper was to try to study whether or not the class II transactivator (*CIITA*) protein located on chromosome 16 was associated with T1D. *CIITA* is the main control factor for *MHC class II* gene expression [163]. This makes it a natural T1D candidate gene since *MHC class II* haplotypes are involved in T1D susceptibility. Four independent cell specific promoters are in charge of the expression of *CIITA* (PI-PIV) [164]. Evidence of suggestive linkage to this region on chromosome 16 has been observed previously (LOD=2.8) in Scandinavian families [126]. Also, heterogeneity in linkage depending on *HLA DRB1* genotype was observed. Furthermore, *CIITA* has previously been demonstrated to be associated with diseases such as myocardial infarction (MI), rheumatoid arthritis (RA) and multiple sclerosis (MS) [135]. Recently, association of *CIITA* has also been observed in celiac disease [165]. Our first results indicated that the rs11074932 and rs3087456 SNPs are modestly associated with T1D (p=0.004 and p=0.001 respectively) in our DISS2 case control material. This association was not confirmed in our additional T1D materials. In order to exclude any possible genotyping errors we began to look at the different allele frequencies for the SNPs in different materials. We soon realized that there seemed to be a skewed distribution of the allele frequencies in the controls, indicating that the allele frequency of the major allele homozygotes is increasing with age (Figure 1, paper III). This observation was confirmed in an independent control material. We do not have a fully satisfactory explanation for this phenomena but our hypothesis is that older people being homozygote for the major allele are generally less ill and therefore tend to participate as controls in genetic studies at a higher rate than other people. The *CIITA* polymorphism probably does not affect health so much in younger individuals. We have with great caution, made sure that the finding is not due to genotyping, calculation errors or population stratification. All individuals of self proclaimed non-Scandinavian origin have been excluded from the

study. Further, both SNPs are in Hardy-Weinberg equilibrium. After the discovery of the change in allele frequencies in controls, we have in order to gain more power, pooled together five different Swedish T1D materials and adjusted for age in a logistic regression analysis. We observe modest association for the two SNPs even after adjustment for age and the association is independent of the nearby *CLEC16A* gene which is a previously identified T1D susceptibility gene [101].

This, together with the fact that we have previously observed significant linkage to the *CIITA* region, indicates that the gene is involved in the development of T1D but exactly how it acts is unclear. However, keeping in mind that the associated SNPs are located in the promoter I and III (PI and PIII) region (expressed in macrophages and activated B- and T-cells respectively) suggest that change in *CIITA* expression may convert protective macrophages into “cytotoxic” macrophages by changing the level of different interleukins. Also, the expression of *CIITA* may influence how quickly a patient develops T1D. Further, our interaction studies on the additive scale indicate that *CIITA* directly interacts with the major T1D protective HLA haplotype DR15 (Figure 5, paper III). The *CIITA* association is seen among DR15 negative T1D patients. This could indicate that the risk caused by *CIITA* is mostly important among young individuals as DR15 is absent among young T1D cases. In addition, it has been seen that increased expression of *CIITA* results in increased expression of MHC class II expression [135] which is the most significant risk gene for T1D. Therefore, any changes in *CIITA* expression may result in increased activation of the immune system. This may in turn lead to faster destruction of the pancreatic β -cells. In conclusion; we have identified *CIITA* as a T1D susceptibility gene. However, the main conclusion obtained from our study is that one should always try to carefully match controls to cases when performing association studies.

5.4 PAPER IV

Because of the vast number of genes involved in T1D susceptibility, it is almost impossible to identify the various pathways leading to disease just by studying one susceptibility gene at the time. It would also be naive to think that each risk gene “acts alone” in a disease pathway. Therefore, it is very interesting to study interaction between T1D risk genes. Although, since statistics simply deal with probability theories, it is often very difficult to decide which available statistical model one

should use especially when the interaction pattern are unknown (as is the case in T1D). This was the main reason why we chose to compare four interaction models (multiplicative, additive, MDR and BN) for investigating interaction between the most strongly T1D associated risk factors; *INS* gene, *PTPN22* and the *HLA* haplotypes *DR3-DQA1*05:01-DQB1*02:01* (DR3), *DR4-DQA1*-03:01-DQB1*03:02* (DR4) and *DR15-DQA1*01:02-DQB1*06:02* (DR15) using a total of 2466 cases and 1132 T1D controls. First we “built” a statistical model for each interaction model in a so called “test set” which consisted of 80% of the data and then testing the prediction accuracy of the affection status for each model in a smaller validation data set which consisted of the remaining 20% of the data. To determine the quality of each model we compared AIC and ROC values. The AIC value is a value which shows how good a model is statistically (taking factors such number of co-variants into consideration). It should be noted that the AIC value only gives you a goodness of fit value for the data which is being studied. On the other hand, the ROC value gives an estimate of how well your model predicts who is a case and control based on data on the studied factors (in our case, *HLA*, *INS* and *PTPN22* genotypes). We wanted to study all four models in the most similar way as possible. Since the BN and MDR models looks at interaction of all risk factors at the same time (unlike the additive model where 2x2 interactions are studied) and thereby gets regression co-efficient values “automatically”, we had to come up with ways to include all significant interactions on the additive scale in one single model. For this, we made so called “dummy variables” (all genotypes for each significant interaction) and included them in a “final” logistic regression model. In the multiplicative analysis, in order to avoid over fitting the model we first performed a 2x2 logistic regression analysis. Since none of the interactions were significant, only single significant risk factors were included in the “final model” (Table 4, paper IV). It should be noted that the “final model” is only used for obtaining ROC values using regression co-efficient values from the final model. According to Rothmans theories, the multiplicative and additive models go “hand in hand” meaning that if your data follows the multiplicative scale, interaction on the additive scale will be observed and vice versa [24]. The multiplicative model is based on a logistic regression model and is only thought to explain an interaction on the statistical scale while interaction on the additive scale is believed to explain interaction on both statistically and “causally”. We did not observe any significant interactions on the multiplicative model. On the other hand several interactions including all studied risk factors deviated from the

additive scale (Table 4, paper IV). This is in line with Rothmans theories and detecting interaction on the additive scale also indicates causal interaction. Hence these results indicate that these genetic risk factors are in combinations involved in some of the causes for disease.

The MDR, model albeit a non-parametric model where few assumptions of interaction are made, is according to me not a really desirable model to use for calculating interaction. The calculation does not give a direct indication of which risk factors interact. Instead it gives a complex results table where one has to interpret which risk factors seem to be present more in cases vs. controls. The model does not give you an AIC value (instead it uses something called balanced accuracy which measures the mean of the sensitivity and specificity to determine cases and controls) and therefore we had difficulties in understanding how to obtain an AIC value. The only way of obtaining an AIC value was by include the results from the predict MDR script (script which predicted cases from controls in the validation set using data from the test set) and running it in a logistic regression model. Since the AIC calculated in the logistic regression analysis only assumed presence of one variable in the model, we then included the obtained AIC value in the AIC formula; $AIC=2k-2\ln(L)$ where $k=5$ risk factors. The MDR model scored lower than the additive and multiplicative models both on AIC as well as ROC values (Table 7, paper IV).

The BN model is according to me the most difficult one to both understand and interpret. The model can be used using prior knowledge or “learn” from the data in different ways. When we first started, we used no prior knowledge about our data. This resulted in unexplainable results where edges were directed in wrong directions (e.g. nodes from affection status to HLA genes or from HLA genes to gender). Because of this we decided to start with a model with directed edges from each risk factor to affection status. This resulted in a final model where edges had also been added between genetic risk factors indicating interaction between them. The BN model scored the lowest in both AIC value and ROC value (Table 7, paper IV) indicating that perhaps it is not the most reliable model to study interaction with. Even though I am personally not a fan of the BN model yet, a huge positive thing about the model is that the interactions can be seen visually, making it easier to see how things interact with each other. Because of this, when studying large numbers of risk factors, perhaps the BN model could be a good first step just to see how things seem to interact. From the results, one may then pick out interesting interactions and study them on the additive scale. However, in my opinion, only people with very good statistical knowledge

should use the BN model especially if no prior knowledge is given to the network (because of the confusing results where one has to master probability theories). In conclusion, we observed no significant interactions on the multiplicative scale. On the additive scale however, several 2x2 interactions were observed. Also, complex interactions were observed in the BN and MDR models. Best AIC values and ROC values were observed for the multiplicative and additive models suggesting that these are the models that best predict case control status when interaction in the data set is present. In order to better understand our results, we plan to study all four interaction models in a synthetic data set where a number of predefined interactions on different scales are included. However, from our study so far, I believe that the additive model seems to be the most desirable one to use when studying interaction in diseases such as T1D since it showed one of the best AIC and ROC values and since it is thought to explain interaction on both statistical as well as causal level. Further, the additive model is also unlike, BN and MDR relatively easy to interpret. However, one should always remember that the additive model is only ideal to use when calculating interaction in relatively uncommon diseases such as T1D. Using the additive scale on e.g. T2D which is a fairly common disease, may if not used properly, lead to false interaction results since it calculates OR's which in turn is converted into relative risk ratios. It should be mentioned that we have used the AP value for calculating deviation from additivity. AP measures the increased proportion of cases due to the interaction of two risk factors among individuals who have been exposed to both risk factors. This measure is believed to be the most robust value when converting relative risk into OR's. Further, we see from our results that all of the studied risk factors seem to interact with each other in one way or the other. The interactions including HLA genes seem to be involved in the strongest interactions. It is likely that the interaction including all studied genes have an influence on autoimmunity in early development of the immune system. Our results remind us of exactly how complex the genetics behind T1D susceptibility really is that choosing different statistical models may give slightly different results.

6 CONCLUDING REMARKS

As mentioned earlier, T1D is a very complex disease where even small genetic differences may alter the expression of a specific gene leading to changes in the immune system. The immune response is an extremely complex and sophisticated system and trying to fit all our genetic findings into this system is not an easy task. In figure 9 I have tried to fit in the genes that showed significant association in my studies into the immune system and indicate where they may act to increase the risk for T1D. It may be suggested that *HTR1A* is involved in the regulation of T-regs leading to inadequate inactivation of auto reactive T-cells and thereby leading to destruction of cells in the pancreas. *RNF180* on the other hand may be involved in the destruction of β -cells through the ubiquitin-proteosome pathway as well as through interaction with *HTR1A*. The interaction may increase the negative effect of *HTR1A* on T-regs. The role of *CIITA* may be that certain polymorphisms alter the expression of MHC II which in turn may e.g. mean that auto reactive T-cells may escape the selection process in the thymus and break down central tolerance. Also, *CIITA* may somehow alter the cytokine production and activate macrophages which then destroy β -cells. *PTPN22* is known to be involved in the activation of T-cells in the periphery and it may be suggested that changes in *PTPN22* expression may inhibit or increase the activation of auto reactive T-cells. The *INS* gene is as mentioned earlier believed to play a vital role in the thymus where it is together with HLA haplotypes involved in the deletion of auto reactive T-cells. In the periphery, the insulin gene is presented by MHC as an antigen and recognized by auto reactive T-cells. Further, *INS* may also be presented to B-cells which in turn produce autoantibodies towards it.

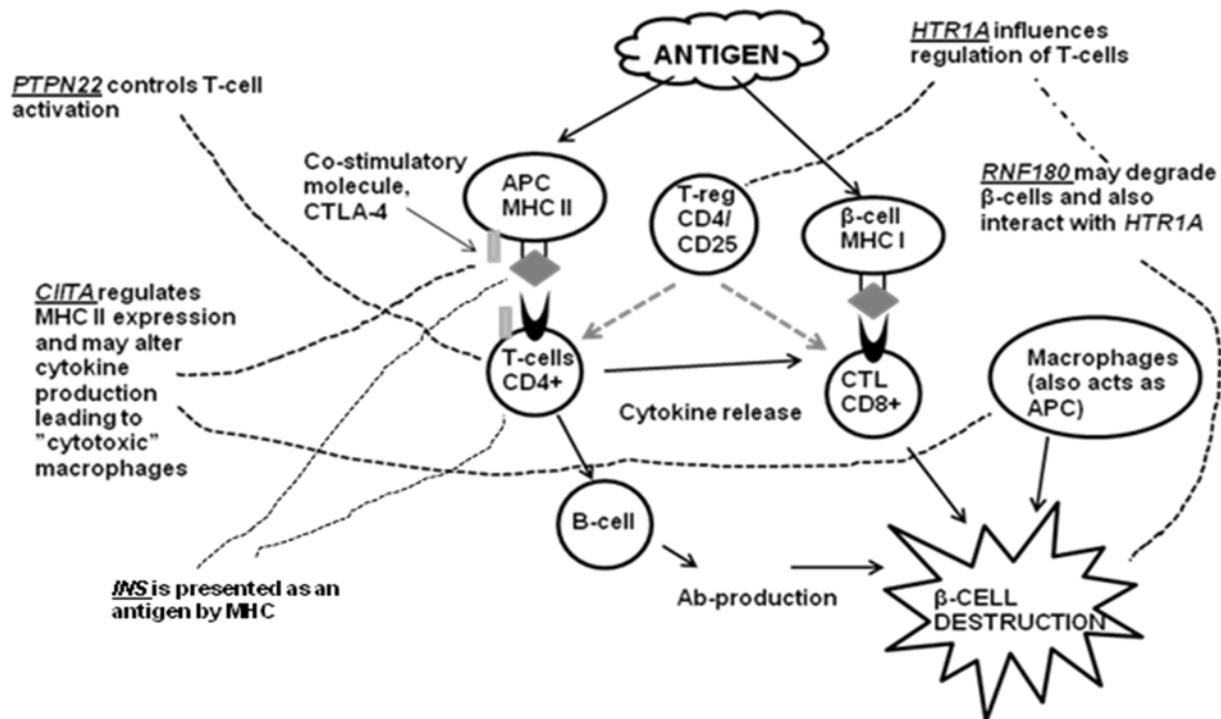


Figure 9. The figure represents the innate (in this case only macrophages) as well as the adaptive (cellular and humoral) immune systems. The black dashed lines represent possible interaction points for the studied genes which may initiate T1D development.

7 FUTURE PERSPECTIVES

The overall aim of my studies has been to try to identify and understand the underlying genetic factors which are involved in the development of T1D.

Multifactorial diseases such as T1D are characterized by the fact that both environment as well as genetics seem to be involved in disease development. Even though scientists have already identified the *HLA* gene as the T1D susceptibility gene with the strongest effect, it is crucial to identify other T1D risk genes with lower effect. Identifying as many genes as possible and understanding how they interact with each other will help us understand the etiology and pathology of T1D in more detail. This could lead to the discovery of new and refined T1D treatments.

However, identifying susceptibility genes in complex diseases is not easy. One reason for this is that many of the genes when being studied alone, have modest effect and therefore may be missed in massive GWAS scans. Also, even though it is understood that other variations such as CNV's may have great influence on the expression of a gene, no effective methods of studying CNV's are available. Further, gene-gene and gene-environment interaction as well as epigenetics are additional factors which make it complicated to identify common susceptibility genes. Therefore, one should always keep in mind that there will never be a "perfect" study which is ideally a cohort study or where e.g. all cases have at least two perfectly matched controls. Instead, one should try to look at the study from every possible angle making sure that one has thought about all possible interfering factors (e.g. correcting for all possible co-variants, making sure you have enough power, exclude possible errors in typing and calculation).

Studying and understanding the behavior of complex diseases is a difficult task.

However, keeping in mind the high increase in T1D incidence rates all over the world with patients being diagnosed with T1D at a much younger age than earlier and the complications caused by the disease, it is extremely important that we continue our research in order to understand T1D etiology and thereby develop better treatments, prevention strategies or even cures for the suffering patients.

Finally, our studies remind us of exactly how incredibly complex and amazing the human being actually is.

8 ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the people who have helped and encouraged me during the time of my PhD studies, especially:

Ingrid Kockum, my main supervisor. One of the kindest people I know. Your knowledge about science is overwhelming! Thank you for “forcing” me to dig deeper into the scary world of statistics and believing I could do it. Thank you for always so patiently answering my questions no matter how many times you have already explained them to me. You are the greatest supervisor a student could ask for. I wish you and your family love and happiness in life!

Pernilla Nikamo, my co-supervisor. You have helped me throughout the years with everything from understanding PCR and linkage when I first arrived to CMM to helping me finish my thesis. You have a great ability to explain even the most difficult things in a simple way which has helped me a lot! Thanks!

Professor **Tomas Olsson**, for letting me part of the neuro group.

Robert Harris, for helping me out with all the paper work and for having the patience to go through my LADOK a million times! ☺

Professor **Timo Koski** and **Henrik Källberg**, both of you have great knowledge in statistics and listening to both of you during our meetings have helped me put many pieces of the statistics puzzle into the right places.

Professor **Martin Cranage**, my supervisor during my MSc thesis at CAMR. Working with you and your group taught me many valuable things about science and made me realize how much fun it really is.

To all co-authors; thank you for all the feed-back and comments and good collaboration.

All patients and controls without whom this thesis would not exist.

My only Type 1 diabetes colleague **Alexandra Gyllenberg**, who has been with me from the very beginning. We have shared discussions about everything from science and tidy labs to weddings and “inskolning” at nursery. I’m glad it was with a kind and easy going co-worker like You I got to share my Type 1 Diabetes experience with. I wish you all the best in life and good luck with your thesis!

Since I only work part time and have small kids at home I rarely get the time to interact with everyone at work as much as I would like to and therefore I would like to give a special thanks to all the people on floor 5 and my old office space for

making me feel part of the group; **Emilie Sundqvist**, for always keeping me updated with the latest meetings and seminars and for always having a helping hand. **Magdalena Lindén**, you just ooze calmness and tranquility. Thank you for always having the time to stop by whenever we meet to chat about everything and nothing. **Magnus Lekman**, thank you for always having the time to help me out no matter how busy you are whenever I needed to understand what CNV's are or to use a program on the computer. **Cecilia Dominguez**, we go all the way back to Södertörn! What a coincidence that we ended up in the same lab. Thank you for all the nice discussions about life in general, the future and family. I wish you all the best in the future! **Izaura Lima Bomfim**, you seem to have your heart and mind at the right place. Thanks for making the lunches on floor 5 more enjoyable. **Tojo James**, the computer specialist and **Pernilla Stridh**, the focused and hard working woman, for shared office space. **Jenny Link**, we have known each other for many years. Thank you for introducing us with Sweden Rock and for all the enormous amount of work you put on helping me with my thesis writing. Your comments really made me try to think "outside the box". **Kerstin Imrell**, for nice company during lunch breaks.

The intelligent **Helga Westerlind**, thank you for sharing the world of Bayesian Networks with me. You make statistics look so simple! Good luck with your thesis!

The always so positive **Selim Sengul**, for always being there whenever one needs help on everything from Taqman to other practical matters.

Mikael Ström, Lina, and Nada Abdelmagid, for nice company in the L2 office.

Sahl, for showing me useful short cuts from the bus stop ☺

Everyone in the Neuro group, especially **Mohsen Khademi**, for everything from practical matters to finding the right folders and e-mail lists.

Everyone in the Rheuma group for shared office space.

Louise Sjöholm, for the chats in the corridors and giving me tips for the future.

Past and previous members of the Schalling group, especially **Anna Witasp, Malin Almgren, Jeanette Johansen** and **Sivonne Arvidsson**.

Old colleagues, especially **Elisabeth Ekelund, Priya Sakhtivel** and **Rita Nohra**.

All my friends and loved ones outside CMM;

Kiran Yunus my fellow molecular biologist, for being the good friend you are and always keeping in touch. It means a lot!

Jasmin Kaleem, Shahida Syed and **Sufia Wasim**, for all the unforgettable times we have shared together, both highs and lows. May there only be highs for all of us now Inshallah! And btw Shahida you are next! Good luck with your viva! 😊

My sweet cousin **Kiran Yousuf** and your whole family (especially our lovely **Isha**) for all the fun we have together making me forget the stress at work for a while. Thank you for being a loving “Kirni khala” for my kids!

My extended families in Pakistan and Finland. Especially my late grandmother in Pakistan and **Mummi** in Finland for always keeping all of us in your hearts and thoughts!

The **Asad’s**, especially **Ammi** and **Abbu** for encouraging me to study further and being proud of me!

My brother **Ilyas** for being a lovely brother! Keep working hard and success will be yours!

My loving parents; **Abba**, for always pushing me to aim higher and supporting me. **Äiti**, for always being the best MUMMO for my children, sharing science with me and being a caring and loving mother! I love you both!

My beloved husband and best friend **Fawwaz Asad**, You are one of the most patient people I know. Words are not enough to thank you for all the support and love you have given me throughout my studies and our marriage. My love for you grows stronger with each day. May Allah always bless our life together. Ameen!

Lastly but most importantly, my princess **Alina** and prince **Ramis**, for always making me smile and forget the troubles of the world. You remind me each and every day of the small things in life that really matter! May you be showered with Allah’s blessings! My love for you is unconditional always and forever....♥

9 REFERENCES

1. www.ncbi.nlm.nih.gov.
2. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, et al. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16: 1182-1190.
3. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. *Nature* 444: 444-454.
4. Beckmann JS, Sharp AJ, Antonarakis SE (2008) CNVs and genetic medicine (excitement and consequences of a rediscovery). *Cytogenet Genome Res* 123: 7-16.
5. Wu XM, X. (2009) Progress in the detection of human genome structural variations. *Sci China Ser C-Life Sci* 52: 560-567.
6. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, et al. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307: 1434-1440.
7. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, et al. (2005) NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 366: 649-659.
8. <http://www.iscid.org/encyclopedia/>.
9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
10. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58: 1347-1363.
11. Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265: 2037-2048.
12. Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57: 439-454.
13. Nyholt DR (2000) All LODs are not created equal. *Am J Hum Genet* 67: 282-288.
14. Rannala B, Slatkin M (1998) Likelihood analysis of disequilibrium mapping, and related problems. *Am J Hum Genet* 62: 459-473.
15. Koillinen H, Ollikainen V, Rautio J, Hukki J, Kere J (2003) Linkage and linkage disequilibrium searched for between non-syndromic cleft palate and four candidate loci. *J Med Genet* 40: 464-468.
16. Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, et al. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2: 204-211.
17. Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29: 311-322.
18. Spielman RSM, R.E. Ewens, W.J. (1994) The transmission/disequilibrium test detects cosegregation and linkage. *Am J Hum Gen* 54: 559-560.
19. Martin ER, Monks SA, Warren LL, Kaplan NL (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 67: 146-154.
20. Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11: 2463-2468.
21. Risch N, Ghosh S, Todd JA (1993) Statistical evaluation of multiple-locus linkage data in experimental species and its relevance to human studies: application to

- nonobese diabetic (NOD) mouse and human insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 53: 702-714.
22. Ahlbom A, Alfredsson L (2005) Interaction: A word with two meanings creates confusion. *Eur J Epidemiol* 20: 563-564.
 23. Kallberg H, Ahlbom A, Alfredsson L (2006) Calculating measures of biological interaction using R. *Eur J Epidemiol* 21: 571-573.
 24. Rothman K (2002) *Epidemiology. An Introduction* New York: Oxford University Press.
 25. Ritchie MD, Moutsinger AA (2005) Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies. *Pharmacogenomics* 6: 823-834.
 26. Ben-Gal I (2007) Bayesian Networks. Ruggeri F, Faltin F, and Kenett R, *Encyclopedia of Statistics in Quality and Reliability*, Wiley and Sons.
 27. <http://www.diabeteshealth.com/read/2008/12/17/715/the-history-of-diabetes/>.
 28. Byrne MM, Sturis J, Menzel S, Yamagata K, Fajans SS, et al. (1996) Altered insulin secretory responses to glucose in diabetic and nondiabetic subjects with mutations in the diabetes susceptibility gene MODY3 on chromosome 12. *Diabetes* 45: 1503-1510.
 29. Lawrence JM, Contreras R, Chen W, Sacks DA (2008) Trends in the prevalence of preexisting diabetes and gestational diabetes mellitus among a racially/ethnically diverse population of pregnant women, 1999-2005. *Diabetes Care* 31: 899-904.
 30. Bentley-Lewis R (2009) Gestational diabetes mellitus: an opportunity of a lifetime. *Lancet* 373: 1738-1740.
 31. Shield JP (2000) Neonatal diabetes: new insights into aetiology and implications. *Horm Res* 53 Suppl 1: 7-11.
 32. DIAMOND Project group (2006) Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. *Diabet Med* 23: 857-866.
 33. Gisela D (2008) Diabetes accelererar bland barn i Europa. *Läkartidningen* 26: 1917.
 34. Patterson CC, Dahlquist GG, Gyurus E, Green A, Soltesz G (2009) Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* 373: 2027-2033.
 35. Soltész GP, C. Dahlquist, C.C. (2006) Global trends in childhood type 1 diabetes. *Diabetes Atlas chapter 21 3rd edn International Diabetes Federation*: 153-190.
 36. Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, et al. (2000) Incidence of childhood type 1 diabetes worldwide. *Diabetes Mondiale (DiaMond) Project Group. Diabetes Care* 23: 1516-1526.
 37. Belkaid YT, K. (2009) Regulatory T-cells in the control of host-microorganism interactions. *Annu Rev Immunology* 27: 551-589.
 38. Wing KS, S. (2010) Regulatory T-cells exert checks and balances on self tolerance and autoimmunity. *Nat Immunology* 11.
 39. Goldsby RAK, T.J. Osborne, B.A. Kuby, J. (2003) *Immunology*, 5th edition.
 40. Foulis AK, McGill M, Farquharson MA (1991) Insulinitis in type 1 (insulin-dependent) diabetes mellitus in man--macrophages, lymphocytes, and interferon-gamma containing cells. *J Pathol* 165: 97-103.
 41. Yoon JW, Jun HS (2005) Autoimmune destruction of pancreatic beta cells. *Am J Ther* 12: 580-591.
 42. Becker F, Helmke K, Bretzel R, Seggewiss K, Sauer H, et al. (1988) [Islet cell (ICA), insulin (IAA) and other autoantibodies in type I diabetic patients and their families: prevalence and clinical correlation]. *Verh Dtsch Ges Inn Med* 94: 1-6.

43. Baekkeskov S, Warnock G, Christie M, Rajotte RV, Larsen PM, et al. (1989) Revelation of specificity of 64K autoantibodies in IDDM serums by high-resolution 2-D gel electrophoresis. Unambiguous identification of 64K target antigen. *Diabetes* 38: 1133-1141.
44. Lan MSW, C. Maclaren, N.K. Notkins, A.L. (1996) IA-2, a transmembrane protein of the protein tyrosine phosphatase family, is a major autoantigen in insulin-dependent diabetes mellitus. *Proc Natl Acad Sci USA* 93: 6367-6370.
45. Wenzlau JM, Hutton JC, Davidson HW (2008) New antigenic targets in type 1 diabetes. *Curr Opin Endocrinol Diabetes Obes* 15: 315-320.
46. Vandewalle CL, Decraene T, Schuit FC, De Leeuw IH, Pipeleers DG, et al. (1993) Insulin autoantibodies and high titre islet cell antibodies are preferentially associated with the HLA DQA1*0301-DQB1*0302 haplotype at clinical type 1 (insulin-dependent) diabetes mellitus before age 10 years, but not at onset between age 10 and 40 years. The Belgian Diabetes Registry. *Diabetologia* 36: 1155-1162.
47. Gorus FK, Goubert P, Semakula C, Vandewalle CL, De Schepper J, et al. (1997) IA-2-autoantibodies complement GAD65-autoantibodies in new-onset IDDM patients and help predict impending diabetes in their siblings. The Belgian Diabetes Registry. *Diabetologia* 40: 95-99.
48. Decochez K, Tits J, Coolens JL, Van Gaal L, Krzentowski G, et al. (2000) High frequency of persisting or increasing islet-specific autoantibody levels after diagnosis of type 1 diabetes presenting before 40 years of age. The Belgian Diabetes Registry. *Diabetes Care* 23: 838-844.
49. Redondo MJ, Eisenbarth GS (2002) Genetic control of autoimmunity in Type I diabetes and associated disorders. *Diabetologia* 45: 605-622.
50. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J (2003) Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 52: 1052-1055.
51. Hirschhorn JN (2003) Genetic epidemiology of type 1 diabetes. *Pediatr Diabetes* 4: 87-100.
52. Hawkes CH (1997) Twin studies in diabetes mellitus. *Diabet Med* 14: 347-352.
53. Redondo MJ, L. Hawa, M. Mackenzie, T. Pyke, DA. Eisenbarth, GS. Leslie, RDG. (2001) Heterogeneity of Type I diabetes: analysis of mono-zygotic twins in Great Britain and the United States. *Diabetologia* 44: 354-362.
54. Viskari H, Ludvigsson J, Uibo R, Salur L, Marciulionyte D, et al. (2005) Relationship between the incidence of type 1 diabetes and maternal enterovirus antibodies: time trends and geographical variation. *Diabetologia* 48: 1280-1287.
55. Forrest JM, Menser MA, Burgess JA (1971) High frequency of diabetes mellitus in young adults with congenital rubella. *Lancet* 2: 332-334.
56. Onkamo P, Vaananen S, Karvonen M, Tuomilehto J (1999) Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. *Diabetologia* 42: 1395-1403.
57. Peltola H, Davidkin I, Paunio M, Valle M, Leinikki P, et al. (2000) Mumps and rubella eliminated from Finland. *Jama* 284: 2643-2647.
58. Dahlquist GG, Ivarsson S, Lindberg B, Forsgren M (1995) Maternal enteroviral infection during pregnancy as a risk factor for childhood IDDM. A population-based case-control study. *Diabetes* 44: 408-413.
59. Hyoty H, Hiltunen M, Knip M, Laakkonen M, Vahasalo P, et al. (1995) A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes* 44: 652-657.

60. Nairn C, D. N. Galbraith, K. W. Taylor and G. B. Clements (1999) Enterovirus variants in the serum of children at the onset of Type 1 diabetes mellitus. *Diabetic Medicine* 16: 509-513.
61. Honeyman MCC, B.S. Stone, N.L. Gellert, S.A. Goldwater, P.N. Steele, C.E. Couper, J.J. Tait, B.D. Colman, P.G. Harrison, L.S. (2000) Association between rotavirus infection and pancreatic islet autoimmunity in children at risk of developing type 1 diabetes. *Diabetes* 48: 1319-1324.
62. Jun HS, Yoon JW (2001) The role of viruses in type I diabetes: two distinct cellular and molecular pathogenic mechanisms of virus-induced diabetes in animals. *Diabetologia* 44: 271-285.
63. Horwitz MS, Ilic A, Fine C, Rodriguez E, Sarvetnick N (2002) Presented antigen from damaged pancreatic beta cells activates autoreactive T cells in virus-mediated autoimmune diabetes. *J Clin Invest* 109: 79-87.
64. Karvonen M, Tuomilehto J, Virtala E, Pitkaniemi J, Reunanen A, et al. (1996) Seasonality in the clinical onset of insulin-dependent diabetes mellitus in Finnish children. Childhood Diabetes in Finland (DiMe) Study Group. *Am J Epidemiol* 143: 167-176.
65. Gamble DR, Taylor KW (1969) Seasonal incidence of diabetes mellitus. *Br Med J* 3: 631-633.
66. Padaiga Z, Tuomilehto J, Karvonen M, Dahlquist G, Podar T, et al. (1999) Seasonal variation in the incidence of Type 1 diabetes mellitus during 1983 to 1992 in the countries around the Baltic Sea. *Diabet Med* 16: 736-743.
67. Hyponen E, Laara E, Reunanen A, Jarvelin MR, Virtanen SM (2001) Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet* 358: 1500-1503.
68. Zipitis CS, Akobeng AK (2008) Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. *Arch Dis Child* 93: 512-517.
69. Bailey R, Cooper JD, Zeitels L, Smyth DJ, Yang JH, et al. (2007) Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes* 56: 2616-2621.
70. Scott F (1990) Cow milk and insulin-dependent diabetes mellitus: is there a relationship? *Am J Clin Nutr* 31: 489-491.
71. Fava D, Leslie RD, Pozzilli P (1994) Relationship between dairy product consumption and incidence of IDDM in childhood in Italy. *Diabetes Care* 17: 1488-1490.
72. Sadauskaitė-Kuehne V, Ludvigsson J, Padaiga Z, Jašinskienė E, Samuelsson U (2004) Longer breastfeeding is an independent protective factor against development of type 1 diabetes mellitus in childhood. *Diabetes Metab Res Rev* 20: 150–157.
73. Borch-Johnsen KJ, G. Mandrup-Poulsen, T. Christy, M. Zachau-Christiansen, B. Kastrup, K. Nerup, J. (1984) Relation between breast-feeding and incidence rates of insulin-dependent diabetes mellitus. A hypothesis. *Lancet* 10: 1083-1086.
74. Virtanen SM, Knip M (2003) Nutritional risk predictors of beta cell autoimmunity and type 1 diabetes at a young age. *Am J Clin Nutr* 78: 1053-1067.
75. Ziegler AG, Schmid S, Huber D, Hummel M, Bonifacio E (2003) Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. *Jama* 290: 1721-1728.
76. Norris JM, Barriga K, Klingensmith G, Hoffman M, Eisenbarth GS, et al. (2003) Timing of Initial Cereal Exposure in Infancy and Risk of Islet Autoimmunity. *Jama* 290: 1713-1720.

77. Hyponen E, Kenward MG, Virtanen SM, Piitulainen A, Virta-Autio P, et al. (1999) Infant feeding, early weight gain, and risk of type 1 diabetes. *Childhood Diabetes in Finland (DiMe) Study Group. Diabetes Care* 22: 1961-1965.
78. Nerup J, Platz P, Andersen OO, Christy M, Lyngsoe J, et al. (1974) HL-A antigens and diabetes mellitus. *Lancet* 2: 864-866.
79. Singal DP, Blajchman MA (1973) Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 22: 429-432.
80. Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 40: 1-14.
81. Todd J (1995) Genetic analysis of type 1 diabetes using whole genome approaches. *Proc Natl Acad Sci USA* 92: 8560-8565.
82. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, et al. (1996) The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* 59: 1134-1148.
83. Devendra D, Eisenbarth GS (2003) Immunologic endocrine disorders. *J Allergy Clin Immunol* 111: S624-636.
84. Steck AK, Rewers MJ (2011) Genetics of type 1 diabetes. *Clin Chem* 57: 176-185.
85. Caillat-Zucman S, Bach JF (1992) [Routine application of HLA class II oligotyping. Value of automatization]. *Presse Med* 21: 1962-1963.
86. Baisch JM, Weeks T, Giles R, Hoover M, Stastny P, et al. (1990) Analysis of HLA-DQ genotypes and susceptibility in insulin-dependent diabetes mellitus. *N Engl J Med* 322: 1836-1841.
87. Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, et al. (2002) Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. *Diabetes* 51: 1346-1355.
88. Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33: 176-183.
89. Bennett ST, Todd JA (1996) Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet* 30: 343-370.
90. Bell GS, MJ. Rutter, WJ. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* 295: 31-35.
91. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, et al. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15: 289-292.
92. Pugliese A, Zeller M, Fernandez A, Jr., Zalcborg LJ, Bartlett RJ, et al. (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* 15: 293-297.
93. Kristiansen OP, Larsen ZM, Pociot F (2000) CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? *Genes Immun* 1: 170-184.
94. Blomhoff A, Lie BA, Myhre AG, Kemp EH, Weetman AP, et al. (2004) Polymorphisms in the cytotoxic T lymphocyte antigen-4 gene region confer susceptibility to Addison's disease. *J Clin Endocrinol Metab* 89: 3474-3476.
95. Atabani SF, Thio CL, Divanovic S, Trompette A, Belkaid Y, et al. (2005) Association of CTLA4 polymorphism with regulatory T cell frequency. *Eur J Immunol* 35: 2157-2162.
96. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, et al. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 36: 337-338.

97. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75: 330-337.
98. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, et al. (2004) Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 75: 504-507.
99. Hill RJ, Zozulya S, Lu YL, Ward K, Gishizky M, et al. (2002) The lymphoid protein tyrosine phosphatase Lyp interacts with the adaptor molecule Grb2 and functions as a negative regulator of T-cell activation. *Exp Hematol* 30: 237-244.
100. Cloutier JF, Veillette A (1999) Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* 189: 111-121.
101. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-678.
102. Malek TR, Bayer AL (2004) Tolerance, not immunity, crucially depends on IL-2. *Nat Rev Immunol* 4: 665-674.
103. Brusko TM, Wasserfall CH, Clare-Salzler MJ, Schatz DA, Atkinson MA (2005) Functional defects and the influence of age on the frequency of CD4+ CD25+ T-cells in type 1 diabetes. *Diabetes* 54: 1407-1414.
104. Qu HQ, Verlaan DJ, Ge B, Lu Y, Lam KC, et al. (2009) A cis-acting regulatory variant in the IL2RA locus. *J Immunol* 183: 5158-5162.
105. Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, et al. (2009) IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 5: e1000322.
106. Concannon P, S. Onengut-Gumuscu, et al. (2008) "A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3." *Diabetes*.
107. Todd JA, N. M. Walker, et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature Genetics* 39: 857-864.
108. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, et al. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet* 42: 68-71.
109. Steck AK, Liu SY, McFann K, Barriga KJ, Babu SR, et al. (2006) Association of the PTPN22/LYP gene with type 1 diabetes. *Pediatr Diabetes* 7: 274-278.
110. Smyth D, Cooper JD, Collins JE, Heward JM, Franklyn JA, et al. (2004) Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes* 53: 3020-3023.
111. Bottini N, Vang T, Cucca F, Mustelin T (2006) Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 18: 207-213.
112. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, et al. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 41: 703-707.
113. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, et al. (2006) A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38: 617-619.
114. Donner H, Rau H, Walfish PG, Braun J, Siegmund T, et al. (1997) CTLA4 alanine-17 confers genetic susceptibility to Graves' disease and to type 1 diabetes mellitus. *J Clin Endocrinol Metab* 82: 143-146.

115. Marron MP, Raffel LJ, Garchon HJ, Jacob CO, Serrano-Rios M, et al. (1997) Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet* 6: 1275-1282.
116. Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes* 59: 2972-2979.
117. Howson JM, Walker NM, Smyth DJ, Todd JA (2009) Analysis of 19 genes for association with type I diabetes in the Type I Diabetes Genetics Consortium families. *Genes Immun* 10 Suppl 1: S74-84.
118. Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, et al. (1998) A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nat Genet* 19: 301-302.
119. Thayer TC, Wilson SB, Mathews CE, (2010) Use of nonobese diabetic mice to understand human type 1 diabetes. *Endocrinol Metab Clin North Am* 39: 541-561.
120. Makino S, Kunimoto K, Muraoka Y, Mizushima Y, Katagiri K, et al. (1980) Breeding of a non-obese, diabetic strain of mice. *Jikken Dobutsu* 29: 1-13.
121. Mordes JP, Desemone J, Rossini AA (1987) The BB rat. *Diabetes Metab Rev* 3: 725-750.
122. Wicker LS, Chamberlain G, Hunter K, Rainbow D, Howlett S, et al. (2004) Fine mapping, gene content, comparative sequencing, and expression analyses support *Ctla4* and *Nramp1* as candidates for *Idd5.1* and *Idd5.2* in the nonobese diabetic mouse. *J Immunol* 173: 164-173.
123. Wallis RH, Wang K, Marandi L, Hsieh E, Ning T, et al. (2009) Type 1 diabetes in the BB rat: a polygenic disease. *Diabetes* 58: 1007-1017.
124. Dahlquist G, Blom L, Holmgren G, Hagglof B, Larsson Y, et al. (1985) The epidemiology of diabetes in Swedish children 0-14 years--a six-year prospective study. *Diabetologia* 28: 802-808.
125. Ostman J, Arnqvist H, Blohme G, Lithner F, Littorin B, et al. (1986) Epidemiology of diabetes mellitus in Sweden. Results of the first year of a prospective study in the population age group 15-34 years. *Acta Med Scand* 220: 437-445.
126. Nerup JP, F. European Consortium for IDDM Studies. (2001) A genomewide scan for type 1-diabetes susceptibility in Scandinavian families: identification of new loci with evidence of interactions. *American Journal of Human Genetics* 69: 1301-1313.
127. Jensen RA, Gilliam LK, Torn C, Landin-Olsson M, Karlsson FA, et al. (2007) Multiple factors affect the loss of measurable C-peptide over 6 years in newly diagnosed 15- to 35-year-old diabetic subjects. *J Diabetes Complications* 21: 205-213.
128. Landin-Olsson M, Karlsson FA, Lernmark A, Sundkvist G (1992) Islet cell and thyrogastric antibodies in 633 consecutive 15- to 34-yr-old patients in the diabetes incidence study in Sweden. *Diabetes* 41: 1022-1027.
129. Dahlquist G, Blom L, Tuvemo T, Nystrom L, Sandstrom A, et al. (1989) The Swedish childhood diabetes study--results from a nine year case register and a one year case-referent study indicating that type 1 (insulin-dependent) diabetes mellitus is associated with both type 2 (non-insulin-dependent) diabetes mellitus and autoimmune disorders. *Diabetologia* 32: 2-6.
130. Carlsson A, Kockum I, Lindblad B, Engleson L, Nilsson A, 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. *Int J Obes (Lond)*.

131. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh CD (2001) Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol* 17: 983-989.
132. Roos IM, Kockum I, Hillert J (2008) The interleukin 23 receptor gene in multiple sclerosis: a case-control study. *J Neuroimmunol* 194: 173-180.
133. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L (2009) Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology* 73: 696-701.
134. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, et al. (2003) Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 62: 835-841.
135. Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, et al. (2005) MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat Genet* 37: 486-494.
136. Bjork BF, Katzov H, Kehoe P, Fratiglioni L, Winblad B, et al. (2007) Positive association between risk for late-onset Alzheimer disease and genetic variation in IDE. *Neurobiol Aging* 28: 1374-1380.
137. McGuigan FE, Larzenius E, Callreus M, Gerdhem P, Luthman H, et al. (2007) Variation in the BMP2 gene: bone mineral density and ultrasound in young adult and elderly women. *Calcif Tissue Int* 81: 254-262.
138. Gerdhem P, Isaksson A, Akesson K, Obrant KJ (2005) Increased bone density and decreased bone turnover, but no evident alteration of fracture susceptibility in elderly women with diabetes mellitus. *Osteoporos Int* 16: 1506-1512.
139. Ekelund E, Lieden A, Link J, Lee SP, D'Amato M, et al. (2008) Loss-of-function variants of the filaggrin gene are associated with atopic eczema and associated phenotypes in Swedish families. *Acta Derm Venereol* 88: 15-19.
140. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, et al. (2001) Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 11: 152-162.
141. Roach JC, Deutsch K, Li S, Siegel AF, Bekris LM, et al. (2006) Genetic mapping at 3-kilobase resolution reveals inositol 1,4,5-triphosphate receptor 3 as a risk factor for type 1 diabetes in Sweden. *Am J Hum Genet* 79: 614-627.
142. Maziarz M, Janer M, Roach JC, Hagopian W, Palmer JP, et al. The association between the PTPN22 1858C>T variant and type 1 diabetes depends on HLA risk and GAD65 autoantibodies. *Genes Immun* 11: 406-415.
143. Graham J, Kockum I, Sanjeevi CB, Landin-Olsson M, Nystrom L, et al. (1999) Negative association between type 1 diabetes and HLA DQB1*0602-DQA1*0102 is attenuated with age at onset. Swedish Childhood Diabetes Study Group. *Eur J Immunogenet* 26: 117-127.
144. Olerup O, Zetterquist H (1992) 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* 39: 225-235.
145. Kiviniemi M, Hermann R, Nurmi J, Ziegler AG, Knip M, et al. (2007) 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 Technol Ther* 9: 460-472.

146. Li YA, GR. (2006) Mach 1.0: rapid haplotype reconstruction and missing genotype inference [abstract 2290/C]. *American Journal of human genetics* S79: 416.
147. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, et al. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476: 214-219.
148. Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 13: 117-137.
149. Gudbjartsson DFJ, K. Frigge, M.L. Kong, A. (2000) Allegro, a new computer program for multipoint linkage analysis. *Nature Genetics* 25: 12-13.
150. Barrett JCF, M. Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *BIOINFORMATICS* 21: 263–265.
151. Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) A dynamic programming algorithm for haplotype block partitioning. *Proc Natl Acad Sci U S A* 99: 7335-7339.
152. Zhang K, Qin ZS, Liu JS, Chen T, Waterman MS, et al. (2004) Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. *Genome Res* 14: 908-916.
153. Dudbridge F (2008) Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 66: 87-98.
154. Dupont WaP, W. (1997) PS power and sample size program available for free on the Internet. *Control Clin Trials* 18: 274.
155. Woolf B (1955) On estimating the relation between blood group and disease. *Ann Hum Genet* 19: 251-253.
156. R Development Core Team (2006) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
157. Winham SJ, Motsinger-Reif AA An R package implementation of multifactor dimensionality reduction. *BioData Min* 4: 24.
158. Steck H (2001) Constrained-Based Structural Learning in Bayesian Networks Using Finite data Sets, PhD Thesis.
159. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, et al. (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* 32: 666-669.
160. Nielsen C, Hansen D, Husby S, Jacobsen BB, Lillevang ST (2003) Association of a putative regulatory polymorphism in the PD-1 gene with susceptibility to type 1 diabetes. *Tissue Antigens* 62: 492-497.
161. Haimila K, Smedberg T, Mustalahti K, Maki M, Partanen J, et al. (2004) Genetic association of coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33. *Genes Immun* 5: 85-92.
162. Chausmer AB (1998) Zinc, insulin and diabetes. *J Am Coll Nutr* 17: 109-115.
163. Steimle V, Siegrist CA, Mottet A, Lisowska-Grospierre B, Mach B (1994) Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* 265: 106-109.
164. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B (1997) Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. *Embo J* 16: 2851-2860.
165. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 43: 1193-1201.