TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. D OSA - TOM. 792
MEDICA – ODONTOLOGICA

GENES OUTSIDE THE HLA REGION AFFECTING SUSCEPTIBILITY TO TYPE 1 DIABETES – THE ROLE OF IDDM2 AND IDDM9 IN THE FINNISH POPULATION.

by

Antti-Pekka Laine

TURUN YLIOPISTO
Turku 2007

From the Department of Virology and Immunogenetics Laboratory, University of Turku, Turku, Finland

Supervised by

Professor Jorma Ilonen
Department of Clinical Microbiology
University of Kuopio
Kuopio, Finland
and
Immunogenetics laboratory
University of Turku
Turku, Finland

Reviewed by

Docent Per-Henrik Groop, MD DMSc
Folkhälsan Institute of Genetics
Folkhälsan Research Center
University of Helsinki
Helsinki, Finland

and

PhD Päivi Saavalainen
Haartman institute
Medical genetics
University of Helsinki
Helsinki, Finland

ABSTRACT

Antti-Pekka Laine
Genes outside the HLA region affecting susceptibility to type 1 diabetes – the role of IDDM2 and IDDM9 in the Finnish population
Immunogenetics laboratory, University of Turku, Turku, Finland
Annales Universitatis Turkuensis
Turku, Finland 2007

HLA region defines half of the genetic susceptibility to type 1 diabetes. Several susceptibility loci residing outside the HLA region have been reported and three loci have been confirmed to convey type 1 diabetes susceptibility.

In the present study twelve of these loci were tested for linkage in 107 Finnish type 1 diabetes multiplex families. In the follow-up study, 121 Finnish type 1 diabetes multiplex families were analyzed for linkage across the IDDM9 region at 3q21, and 384 Finnish type 1 diabetes simplex families were analyzed for association with two markers in the same region. Additionally, the IDDM2 region was analyzed for interaction with HLA haplotypes of varying disease risk in order to assess its potential regarding refinement of disease risk prediction. In an additional attempt to refine the IDDM2 region disease risk prediction, subtyping of the protective VNTR class III haplotypes and an association study was performed in the Finnish and Swedish populations.

No significant or suggestive linkage was found in the complete dataset outside the HLA region. The highest non-HLA LOD score was seen with the IDDM9 region marker D3S3576 (MLS=1.05). The follow-up study at the IDDM9 region revealed a stronger LOD score at the marker D3S1589 (MLS=3.4). Association with the disease was found close to D3S1589 with marker AFM203wd10 (TDT p=0.0002). Our linkage results did not confirm previous findings in other studies. The follow-up study revealed a LOD score close to genome-wide significance and a significant association at the 3q21 region which warrant a more comprehensive association study at this region. The predisposing MspI-2221 genotype CC was more common in diabetics with moderate or low HLA risk compared with the high risk group, and the genotype distribution between said groups also showed significant difference (p=0.05 and p=0.01 respectively). The VNTR analysis revealed a significantly stronger protective effect for the VNTR IIIA/IIIA genotype when compared to other class III genotypes.

These findings suggest that etiological heterogeneity between protective lineages of the IDDM2 region possibly exist and that the HLA region could modify the IDDM2 risk and further that deconstructing the IDDM2 heterogeneity has potential in enhancing the sensitivity and specificity of type 1 diabetes risk estimation.

Keywords: type 1 diabetes, IDDM2, IDDM9, linkage, association

TIIVISTELMÄ

Antti-Pekka Laine
Tyypin 1 diabeteksen perinnöllinen alttius Suomessa - HLA-alueen ulkopuolisten alttiuslokusten IDDM2 ja IDDM9 rooli taudin periytymisessä
Immunogenetiikan laboratorio, Turun Yliopisto, Turku, Suomi
Annales Universitatis Turkuensis
Turku, Suomi 2007

HLA-alue, joka sijaitsee kromosomissa 6p21.3, vastaa noin puolesta perinnöllisestä alttiudesta sairastua tyypin 1 diabetekseen. Myös HLA-alueen ulkopuolisten lokusten on todettu liittyvän sairausalttiuteen. Näistä kolmen lokuksen on varmistettu olevan todellisia alttiuslokuksia ja lisäksi useiden muiden, vielä varmistamattomien lokusten, on todettu liittyvän sairausalttiuteen.

Tässä tutkimuksessa 12:n HLA-alueen ulkopuolisen alttiuslokuksen kytkentä tyypin 1 diabetekseen tutkittiin käyttäen 107:aa suomalaista multiplex-perhettä. Jatkotutkimuksessa analysoitiin IDDM9-alueen kytkentä ja assosiaatio sairauteen laajennetuissa perhemateriaaleissa sekä IDDM2-alueen mahdollinen interaktio HLA-alueen kanssa sairauden muodostumisessa. Lisäksi suoritettiin IDDM2-alueen suojaavien haplotyyppien alatyypitys tarkoituksena tutkia eri haplotyyppien käyttökelpoisuutta sairastumisriskin tarkempaa ennustamista varten.

Ensimmäisessä kytkentätutkimuksessa ei löytynyt koko genomin tasolla merkitsevää tai viitteellistä kytkentää tutkituista HLA-alueen ulkopuolisista lokuksista. Voimakkain havaittu nimellisen merkitsevyyden tavoittava kytkentä nähtiin IDDM9-alueen markkerilla D3S3576 (MLS=1.05). Tutkimuksessa ei kyetty varmistamaan tai sulkemaan pois aiempia kytkentähavaintoja tutkituilla lokuksilla, mutta IDDM9-alueen jatkotutkimuksessa havaittu voimakas kytkentä (MLS=3.4) ja merkitsevä assosiaatio (TDT p=0.0002) viittaa vahvasti siihen, että 3q21-alueella sijaitsee todellinen tyypin 1 diabeteksen alttiusgeeni, jolloin alueen kattava assosiaatiotutkimus olisi perusteltu jatkotoimenpide. Sairauteen altistava IDDM2-alueen MspI-2221 genotyyppi CC oli nimellisesti yleisempi matalan tai kohtalaisen HLA-sairastumisriskin diabeetikoilla, verrattuna korkean HLA-riskin potilaisiin (p=0.05). Myös genotyyppijakauman vertailu osoitti merkitsevää eroa ryhmien välillä (p=0.01). VNTR-haplotyyppitutkimus osoitti, että IIIA/IIIA-homotsygootin sairaudelta suojaava vaikutus on merkitsevästi voimakkaampi kuin muiden luokka III:n genotyypeillä. Nämä tulokset viittaavat IDDM2-HLA -vuorovaikutukseen sekä siihen että IDDM2-alueen haplotyyppien välillä esiintyy etiologista heterogeniaa. Tämän johdosta IDDM2-alueen haplotyyppien tarkempi määrittäminen voisi tehostaa tyypin 1 diabeteksen riskiarviointia.

Avainsanat: tyypin 1 diabetes, IDDM2, IDDM9, kytkentä, assosiaatio

CONTENTS

ABBREVIATIONS

AFBAC	affected family-based artificial controls
ASP	affected sibling pair
Bp	base pairs of DNA nucleotides
CD	cluster of differentiation
CTLA4	cytotoxic T lymphocyte associated antigen 4
DIPP	diabetes prediction and prevention trial
DNA	deoxyribonucleic acid
GAD	glutamic acid decarboxylase
HLA	human leukocyte antigen
IA2A	insulinoma-associated protein 2 antigen
IAA	insulin autoantibody
IBD	identical by descent
IBS	identical by state
ICA	islet-cell autoantibody
IDDM	insulin-dependent diabetes mellitus
Kb	kilobase, thousands of base pairs DNA nucleotides
LD	linkage disequilibrium
LOD	logarithm of odds
Mb	megabase, millions of base pairs of DNA nucleotides
MHC	major histocompatibility complex
MLS	maximum LOD score
MMLS	multipoint maximum LOD score
mTEC	medullary thymic epithelial cells
OR	odds ratio
PCR	polymerase chain reaction
PIC	polymorphic information content
PTPN22	protein tyrosine phosphatase, non-receptor type 22
RA	rheumatoid arthritis

RFLP	restriction fragment length polymorphism
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
T1D	type 1 diabetes
TCR	T-cell receptor
TDT	transmission/disequilibrium test
VNTR	variable number of tandem repeats

LIST OF ORIGINAL PUBLICATIONS

I Laine AP, Nejentsev S, Veijola R, Korpinen E, Sjöroos M, Simell O, Knip M, Åkerblom HK, Ilonen J. A linkage study of 12 IDDM susceptibility loci in the Finnish population. Diabetes Metab Res Rev. 2004 Mar-Apr;20(2):144-9.

II Laine AP, Turpeinen H, Veijola R, Hermann R, Simell O, Knip M, Ilonen J. Evidence for linkage to and association with type 1 diabetes at the 3q21 region in the Finnish population. Genes Immun. 2006 Jan;7(1):69-72.

III Laine AP, Hermann R, Knip M, Simell O, Åkerblom HK, Ilonen J. The human leukocyte antigen genotype has a modest effect on the insulin gene polymorphism-associated susceptibility to type 1 diabetes in the Finnish population. Tissue Antigens. 2004 Jan;63(1):72-4.

IV Laine AP, Holmberg H, Nilsson A, Ortqvist E, Kiviniemi M, Vaarala O, Åkerblom HK, Simell O, Knip M, Ludvigsson J, Ivarsson SA, Larsson K, Lernmark A, Ilonen J; Finnish Paediatric Diabetes Registry. Two insulin gene single nucleotide polymorphisms associated with type 1 diabetes risk in the Finnish and Swedish populations. Disease Markers. 2007;23(3):139-45.

INTRODUCTION

Type 1 diabetes (T1D) is an autoimmune disorder with multifactorial inheritance where both genetic and environmental components contribute to the emergence of the disease. The genetic component explains approximately 50% of the overall risk of developing T1D. The human leukocyte antigen (HLA) locus, which is located at 6p21.3, is the main genetic component and accounts for ~50% of the genetic risk for T1D. While the majority of genetic risk for T1D can be attributed to the HLA region, other regions of importance have also been identified. Although there are many genomic regions with reported evidence of linkage to or association with T1D outside the HLA region, only three loci have consistently shown association in independent data sets and are generally regarded as true susceptibility loci. The best known and most studied of these three loci is IDDM2 at the insulin gene region (11p15). The other two loci, CTLA4 at 2q33 and PTPN22 at 1p13, have been identified recently and have not yet been studied as widely as the IDDM2 region.

The method of discovery for the HLA region, IDDM2 and PTPN22 has been the candidate gene approach using association methods. CTLA4 region is an exception as it was primarily detected in linkage studies (albeit only with suggestive evidence) and later confirmed in association studies.

The linkage method is based on detection of the intra-familial linkage disequilibrium between the disease allele and the parent-specific marker allele, unlike association methods, which are based on detecting population-wide linkage disequilibrium between a specific marker allele and a disease allele. The linkage method has less statistical power to detect susceptibility loci than association methods if a short distance (<5kb) is assumed between the marker and disease locus, but the linkage method can detect susceptibility loci from a distance up to 1 Mb. Association methods are limited by the extent of linkage disequilibrium at the studied region in the population (usually <10 kb). Linkage and association methods are both commonly used in mapping studies of complex disease genes, linkage for initial detection of an interesting region and association for fine mapping and confirming the susceptibility region.

1. REVIEW OF THE LITERATURE

1.1 The pathogenesis and etiology of type 1 diabetes

Progression to the clinical phase of T1D is preceded by an autoimmune process that typically takes years and leads to destruction of the insulin producing Langerhans islet cells (beta-cells) in the pancreas. The destruction is mainly due to cell-mediated immunity where T helper cells (CD4+) activate antigen-specific cytotoxic T cells (CD8+) which act specifically against the beta-cells.

Histological evidence of the autoimmune destruction is visible during the onset of T1D. The size of the pancreatic islets is reduced considerably due to the fact that the largely destroyed beta-cells comprise ~70% of the number of cells present in the islets. Another histological feature found in the islets at the onset of T1D is the chronic inflammatory infiltrate known as insulitis that consists of CD8 and CD4 positive T cells, B lymphocytes, macrophages and natural killer cells. The expression of HLA class I and class II genes is also increased in the islet cells during insulitis (Bottazzo et al. 1985; Foulis et al. 1987; Foulis et al. 1991; Hanninen et al. 1992).

The clinical symptoms of T1D (commonly caused by hyperglycemia and ketoacidosis) emerge abruptly and only when most of the beta-cells have been destroyed. Progression to the clinical phase of the disease is usually preceded by the production of several autoantibodies against beta-cells during the autoimmune process. Some of these autoantibodies are routinely detected and used as predictors for the onset of the clinical phase of T1D (Christie 1996) and also as a means for diagnosis and disease classification (Bottazzo et al. 1974). Autoantibodies against alpha, beta, delta and pancreatic polypeptide islet cells (ICA) were the first discovered in 1974 (Bottazzo et al. 1974), followed by the insulin autoantibodies (IAA) in 1983 (Palmer et al. 1983). Since then, several other autoantibodies have been described, including the enzymes glutamic acid decarboxylase (GAD) (Baekkeskov et al. 1990) and protein tyrosine phosphatase insulinoma associated antigen-2 (IA-2) which has some similarity with the protein tyrosine phosphatase but no demonstrated enzyme activity (Rabin et al. 1994; Lan et al. 1996). GAD and IA2A, in addition to IAA, are currently exhibiting the best combined sensitivity and specificity, i.e. probability of autoantibodies to predict morbidity in affected individuals when

present (sensitivity) and protection against disease in non-affected individuals when absent (specificity). Combined sensitivity and specificity for these autoantibodies (2≥ of the IAA, GADA or IA-2A present) at the onset of T1D are 92% and 98% respectively (Kukko et al. 2005).

The etiology of type 1 diabetes has a strong genetic background, but the role of environmental factors is of equal importance. Studies comparing T1D concordance rates between monozygotic and dizygotic twins have consistently shown significantly stronger disease concordance for monozygotic twins, which attests the existence of strong genetic factors in the etiology of T1D (Kaprio et al. 1992; Hyttinen et al. 2003). On the other hand, reports of significant discordance between monozygotic twins in the emergence of T1D demonstrate the significance of environmental factors (Redondo et al. 1999), although the randomly determined characters of the immune system, which are not caused by any singular constant environmental factor, also play an important part in the development of discordance between monozygotic twins. In addition, the monozygotic twins who have a high genetic risk of developing T1D are more concordant than those with a lower genetic risk, thus implying a smaller-than-assumed effect for environmental factors in the etiology of T1D.

Although the significance of environmental factors in the etiology of T1D is undeniable, no individual exogenous factor or factors that trigger beta-cell autoimmunity have been definitely identified. The most heavily studied candidates so far are virus infections and dietary factors. The emergence of autoimmunity towards beta-cells in very early life follows a seasonal pattern where most new cases appear during the cold months in the fall and winter. This, in addition to the observed annual variation in the emergence of T1D, points to enterovirus infections as a very plausible trigger for beta-cell autoimmunity (Knip et al. 2005). Enterovirus infections fit well with the observed pattern of T1D emergence since the number of enterovirus infections follows a temporally variable seasonal pattern with a peak at the start of the cold season.

In addition to viral triggers, considerable effort has been expended in assessing the role of very early exposure to cow's milk in the emergence of T1D. It is hypothesized that in children who are exposed to cow's milk and develop immunity towards bovine insulin, the cross-reactivity between bovine and human insulin, which differ only by three amino-acids, could allow autoimmunity to develop towards human insulin. There are data from population-based case-control studies that suggest a connection with higher cow's milk intake and the emergence of T1D or appearance of autoantibodies (Verge et al. 1994; Virtanen et al. 1998; Virtanen et al. 2000), but conflicting

data have also been published (Dahlquist et al. 1990). It is still unclear whether or not an exposure to cow's milk in early life is a diabetogenic risk factor or not (Knip et al. 2005). More comprehensive studies are required to assess this.

1.2 Epidemiology of type 1 diabetes

The incidence of T1D varies considerably across geographical regions (1988; 1991). The differences are apparent both between and within populations across geographical distance, which could imply that regional clustering of genetic or environmental effects or their interaction is causing the observed difference. The highest incidence rates are found among white European populations, especially in Northern Europe. The lowest incidence rates are found in Asian and South-American populations.

The country with the highest prevalence of T1D in the world is Finland, where incidence for children under 15 years in 2005 was 62 per 100,000 per year (A. Reunanen, personal communication). Incidence is also among the highest in Sweden and Norway but not in Estonia, Latvia and Lithuania, and in fact, a marked north-south and west-east gradient of decreasing incidence in Europe is seen (Green et al. 1992). A notable exception is Sardinia where T1D incidence is the second highest in Europe after Finland. Other regions of Italy show 3.1-4.3-fold smaller incidence (Green et al. 2001) and similar regional variation of incidence has been shown in other populations, including Finland. Variation in Finland was not as pronounced as that in the studied regions of Italy, but 2-fold differences in incidence rates among different geographical regions have been described (Rytkonen et al. 2001). The regions that showed persistently high incidence of T1D throughout the two study periods (1987-1991, 1992-1996) in Finland were rural heartland areas (Rytkonen et al. 2001; Rytkonen et al. 2003). The possibility that the urban environment protects one from T1D has been studied to some extent, but no consistent proof that confirms or denies the hypothesis has emerged (Rytkonen et al. 2003). The T1D incidence has been rapidly increasing, particularly in Western Europe. An annual increase of 3-4% in incidence rates has been observed in the EURODIAB study (2000). The phenomenon is age dependent, the increase in incidence rates is 2.6 times bigger in the age group of 0-4 years than in the age group of 10-14 years (Atkinson et al. 2001). In Finland, a 4.5-fold increase in T1D incidence has taken place during the last 50 years (Gale 2002). This observed increase of incidence rate, as well as the geographical variation of

incidence, is most logically explained by transient environmental causes. Genetic factors could participate in the regional variation, but not be the principal cause of either variation or the increasing incidence rate.

1.3 Genetics of type 1 diabetes

1.3.1 Genetic mapping of the complex trait susceptibility loci

1.3.1.1 Linkage

The proportion of expected recombinants between two unlinked loci is 50%. This proportion is called the recombination fraction (θ=theta), which reaches its maximum probability 0.5 only if the loci are in different chromosomes or so far apart in the same chromosome that crossing-over is certain to occur between them. If the distance between the two loci is small enough to reduce the probability of crossing-over to <1.0, the proportion of observed recombinants is less than 50% ($\theta < 0.5$) and the loci are said to be genetically linked. Genetic linkage is used in the mapping of multifactorial diseases in two basic approaches. In the linkage analysis approach, families with parents and pairs of affected siblings (affected sibling pair=ASP) are collected and genetic markers are genotyped in candidate genes or across the whole genome. Since the trait is inheritable, the phenotype (=disease) is assumed to arise from a genetic locus/loci and therefore linkage between markers and underlying disease genes can be examined. In the association approach, polymorphisms are genotyped in candidate genes or in the whole genome using cohorts of affected cases and a control population. If a marker is linked to the disease gene, marker alleles are expected to co-segregate with the disease gene alleles (phenotype) and thus be more prevalent in the case cohort than in the control cohort.

1.3.1.2 Statistical linkage analysis

Linked loci produce less than the expected 50% of recombinant offspring. In parametric linkage analysis, the amount of reduction in the number of recombinants is detected, and the statistical significance of the degree of deviation is tested using the LOD score (logarithm of odds) method. Parametric linkage analysis is poorly suited to calculating linkage in multifactorial diseases since it requires inheritance parameters such as penetrance, mode of inheritance and disease allele frequencies, which are not known. To overcome this problem, non-parametric methods can be used to analyze linkage. Instead of calculating the linkage by estimating the proportion of recombinants, non-parametric methods compare how much of the parental genome affected individuals share at the region of interest. The most commonly used non-parametric method for mapping multifactorial traits is the affected sibling pair (ASP) method, in which at least two affected children and both parents are analyzed. If the marker studied is linked to the disease gene in the parent, both affected children are likely to receive the same parental marker allele because it co-segregates with the disease causing allele. The ASP method calculates the number of marker alleles affected sibling pairs share identical by descent (IBD). Sibling pairs share an allele IBD when both siblings have inherited the same allele from the same parental chromosome. Since there are four parental chromosomes, siblings can share 0, 1 or 2 alleles IBD in their genotypes. In the case of no linkage, the expected sharing proportions for sibling pair genotypes in a collection of families are 25%, 50% and 25% respectively. In other words, 50% of parental alleles in sibling pairs are shared IBD if the loci are not linked. Now, if a certain marker allele is linked to the disease gene allele, that marker allele is more likely to be transmitted to both affected siblings and consequently enlarge the proportion of sibling pairs that share one or two alleles IBD and reduce the proportion of siblings that share no alleles IBD. When a large number of unrelated ASP families is collected and analyzed, linkage can be detected by testing the significance of deviation between expected 50% IBD-sharing (no linkage) and observed IBD-sharing, using the chi-square test.

1.3.1.3 Linkage disequilibrium

Two bi-allelic loci that are located in the same chromosome (such as SNP markers) can theoretically form four distinct haplotypes (or gametes) if the markers combine freely, i.e. a recombination event is certain to happen between the loci. In this case, the expected haplotype frequencies are simply a product of allele frequencies of the markers. Since the probability of a recombination event decreases as the physical distance between the loci decreases, closely-spaced bi-allelic loci do not combine freely and thus haplotype frequencies differ from the expected, or one or two haplotypes are missing entirely. If this is the case, the loci are said to be in linkage disequilibrium (LD), where a specific allele of one locus co-segregates more often than expected with a specific allele of another locus and thus predicts the genotype status of the other marker (if LD is complete).

The most common statistical measures of LD used in the context of gene mapping are D' and r^2. Both measures denote perfect LD with the value 1 and complete lack of LD with 0. If no recombination event has taken place between the loci during the history of the (studied) population, the D' is 1, which means that one or two of the four possible haplotypes (gametes) is missing. The existence of three haplotypes between two SNPs can be explained by consecutive mutations when the two SNPs were formed, the occurrence of historical recombinations is not required. On the other hand, the existence of four haplotypes in the population between two SNPs can only be explained by historical recombination between the loci (D'<1) (Mueller 2004). The statistical properties of D' are problematic when it is applied in association studies as the measure of LD. A small sample size and rare alleles can cause inflated estimates of D' and it is not comparable between populations since it is a measure of historical recombination events in the population (sample set) studied. In addition, the relative properties of D' are unclear when D'<1 (Pritchard et al. 2001; Carlson et al. 2004; Mueller 2004; Wang et al. 2005). The human genome consists of blocks of limited genetic diversity where LD and consequently also haplotypes extend over several kilobases (usually 1-10kb). These "haplotype blocks" of extended LD are interrupted by break-points caused by recombination hotspots. Since D' is a measure of historical recombination events between loci, it can be used as a tool for detecting recombination hotspots and accordingly also as a way of defining a particular haplotype block by showing the break points around it.

Another commonly used measure of LD is r^2, which is the correlation between the alleles of two loci. If two loci are not in perfect LD and the first locus is analyzed as a surrogate marker for the second locus, the factor for increasing the sample size to gain the same power as analyzing the second locus directly is calculated as $1/r^2$. For example, if r^2 between the first and the second locus is 0.5, twice as many samples would be needed for the first locus to gain the same statistical power as analyzing the second locus directly (Pritchard et al. 2001). These properties of r^2 enable its use as a measure of tagging efficiency in association mapping studies which increasingly rely on analyzing tagging SNPs instead of all available SNPs.

1.3.2 Complex inheritance

In contrast to monogenic traits, where a single gene accounts for the phenotype in a Mendelian manner, complex traits have several genes acting jointly in constituting a phenotype. Genetically complex traits, such as common diseases, are typically multifactorial; not only multiple genes but also multiple environmental factors play a vital part in their emergence.

Multifactorial diseases show gene-environment interactions and gene-gene interactions in addition to having typically low-penetrance susceptibility genes and locus and allele heterogeneity. Due to these aspects, multifactorial disease loci have low phenotype-genotype equivalency (heritability). The individual effect each multifactorial disease gene mutation conveys to the overall risk is small, which suggests that their effect on the function or structure of the protein the gene codes for is also small. They are more likely to cause minor variation in the regulation of the gene expression or splicing than completely block the function of the gene. The efficiency of linkage and association studies relies to a great degree on the assumption that a certain phenotype equals a certain genotype. Therefore, in the gene mapping studies of multifactorial diseases where equivalency of the two is low the statistical power to find an individual disease polymorphism is also low.

The amount of heritability of a trait such as a complex disease is measured in the amount of familial clustering the trait demonstrates. If the disease is heritable, the proband's relatives have a higher risk for the disease than a person whose relatives are healthy because relatives share a larger amount of genome than non-relatives. In the context of gene mapping using a linkage

method, such as ASP, the measure which is commonly used for familial clustering is a sibling risk/population prevalence ratio, λ_s (lambda-s) (Risch 1990). For example, since the risk for T1D in the general population is ~0.4% and the risk for a sibling of an affected individual is ~6%, the λ_s for T1D is ~15. This number comprises the total familial clustering of all T1D susceptibility genes; the locus specific numbers are obviously much smaller (λ_s=~3 for the HLA region (Concannon et al. 2005) and $\lambda_s \leq 1.4$ for the rest of the loci (Cox et al. 2001)). Lambda-s operates as an indicator of and measure for a heritable trait, but it is also used as a factor in power calculations to predict probabilities of detecting susceptibility loci in linkage studies.

The power to detect linkage in traits with multiple loci does not rely only on the amount of heritability of a trait studied. There are several confounding factors involved with detecting a linkage to particular susceptibility locus. The model of inheritance has a strong effect on the expected power of the ASP method (Risch 1990). A particular susceptibility locus of a complex trait with multiple loci could have a dominant, recessive or additive model of inheritance or the loci (alleles) can act epistatically (multiplicative model). Other confounding factors include the degree of marker polymorphism (Risch 1990) and the recombination fraction θ (i.e. the genetic distance between the marker studied and the susceptibility polymorphism), although the effect of these factors can be notably alleviated by the use of multipoint linkage methods (Risch 1990).

In the case of association studies, similar functionality for power calculations and heritability quantification is gained using different methods. Association studies are commonly performed using cohorts of cases and controls (case-control studies) or analyzing simplex families with TDT. The measure of heritability used in case-control studies is the odds ratio, which is the ratio of the odds of finding a certain genetic variant in cases versus finding the same variant in controls. If the odds ratio is larger than one, the genetic variant has a predisposing effect, and if the odds ratio is smaller than one, the genetic variant has a protective effect. The odds ratio can also be presented as corresponding lambda-s value by mathematical conversion. In the TDT studies, the proportion of transmitted or non-transmitted alleles differing from the expected 50% transmission gives a comparable and mathematically adaptable measure of heritability across different studies.

1.3.3 T1D susceptibility genes

Type 1 diabetes is regarded as a complex disease with multiple genetic and multiple environmental factors contributing to the overall risk, although rare monogenic and oligogenic forms of T1D exist (Nagamine et al. 1997; Verge et al. 1998). A comparison of the concordance rates between pairs of monozygotic and dizygotic T1D twins provides an estimate of the contribution the genetic factors independently have to the overall disease susceptibility (Boomsma et al. 2002). Based on such twin studies, a contribution of 66-72% to the overall risk has been estimated for the genetic factors (Kaprio et al. 1992; Kyvik et al. 1995; Hirschhorn 2003).

The etiology of T1D has an unambiguous genetic basis, but the disease does not follow any apparent mode of inheritance in families. Type 1 diabetes is most likely an outcome of several genetic factors interacting with each other and with environmental factors.

There are currently four separate loci that are regarded as confirmed T1D susceptibility loci (the HLA region, the insulin gene region, the CTLA4-region and PTPN22), but over twenty genomic regions have been implicated in linkage and association studies over the years (Table 1.). Plausible evidence for four additional loci (12q24, 12q13, 16p13 and 18p11) has been obtained recently in large association studies (Todd et al. 2007; (WTCCC) 2007), but further evidence is required before these loci can be considered confirmed T1D susceptibility loci.

Table 1. The loci with reported linkage or association evidence for T1D

Loci	Location	Reference
IDDM1	6p21.3	(Singal et al. 1973)
IDDM2	11p15.5	(Bennett et al. 1995)
CTLA4-region (IDDM12)	2q33	(Nistico et al. 1996)
PTPN22	1p13	(Bottini et al. 2004)
IDDM3	15q26	(Field et al. 1994)
IDDM4	11q13	(Hashimoto et al. 1994)
IDDM5	6q25	(Concannon et al. 1998)
IDDM6	18q22	(Merriman et al. 1997)
IDDM7	2q31	(Copeman et al. 1995)
IDDM8	6q25-q27	(Luo et al. 1996)
IDDM9	3q21	(Cordell et al. 1995)
IDDM10	10p15.1	(Davies et al. 1994)
IDDM11	14q24-q31	(Field et al. 1996)
IDDM13	2q34	(Morahan et al. 1996)
IDDM15	6q21	(Delepine et al. 1997)
IDDM17	10q25	(Verge et al. 1998)
IDDM18	5q33	(Morahan et al. 2001)
IDDM19	2q24	(Smyth et al. 2006)
GCK	7p15-p13	(Rowe et al. 1995)
	1q42	(Ewens et al. 2002)
	16q22-q24	(Mein et al. 1998)
	Xp11	(Davies et al. 1994)
IFNG	12q13	(Pociot et al. 1997)
	12q24	((WTCCC) 2007)
	16p13	((WTCCC) 2007)
	18p11	((WTCCC) 2007)

1.3.3.1 The HLA region (IDDM1)

The HLA region at 6p21.3 was the first locus found to be associated with T1D (Singal et al. 1973; Nerup et al. 1974; Cudworth et al. 1975; Cudworth et al. 1975) and was designated the first T1D susceptibility locus (IDDM1). The strong association with and linkage to IDDM1, which has been confirmed consistently in numerous studies over the years, has corroborated the status of HLA region as the primary T1D susceptibility locus. Based on linkage studies that have generally excluded the existence of T1D susceptibility loci with λ_s higher than 1.4 (Concannon et al. 2005), the individual contribution of IDDM1 to overall genetic risk of T1D has been estimated to be around 50% with λ_s=2.5-3.6 (Risch 1987; Cox et al. 2001).

The HLA region is known to harbor more than one locus affecting T1D susceptibility. The strong LD throughout the HLA region has greatly diminished the resolution of causal variant mapping, making it an example of the difficulties involved in the positional cloning of complex trait genes. The associations were originally described to serologically defined molecules, HLA-DR3 and HLA-DR4 were associated with disease risk. Molecular biology based methods have allowed more exact mapping of disease genes and sequence based nomenclature is used for the alleles. The HLA class II locus DQB1 has shown the strongest association with T1D in both protective and predisposing haplotypes and was the first locus that was detected (Todd et al. 1987). DQB1 encodes the etiologically relevant DQ heterodimer molecule in concert with the DQA1 and, therefore, polymorphisms in DQA1 also show major effects as expected. In addition, other class II loci also affect the disease risk (*DR, DP*). In fact, the haplotypes of DRB1-DQA1-DQB1 are currently the basis of the T1D risk assessment at the HLA region (Table 2) (Todd et al. 1987; Noble et al. 1996; Undlien et al. 1997). Additionally, evidence for susceptibility loci apart from DRB1, DQA1 or DQB1 is steadily accumulating (Lie et al. 1999; Nejentsev et al. 2000; Zavattari et al. 2001; Johansson et al. 2003; Gombos et al. 2006; Van Autreve et al. 2006).

Table 2. The major risk and protective haplotypes in the HLA region in the Finnish population (Hermann et al. 2003).

Risk HLA haplotypes	**Odds ratio**
DRB1*0401-DQB1*0302	5.64
DQA1*05-DQB1*02	3.40
DRB1*0404-DQB1*0302	3.07
Protective HLA haplotypes	**Odds ratio**
DQA1*0201-DQB1*0303	0.02
DQB1*0503	0.04
DRB1*0403-DQB1*0302	0.14
DQB1*0602	0.07
DQB1*0603	0.19
DQA1*05-DQB1*0301	0.28
DQB1*0501	0.53

1.3.3.2 The Insulin gene region (IDDM2)

IDDM2 has shown consistent association with T1D susceptibility in numerous studies performed in different populations. A variable number of tandem repeat (VNTR) polymorphism located 596 bp upstream of the first codon of the insulin-encoding *INS* gene is frequently proposed as the functional polymorphism of the IDDM2 locus, although two SNP markers (HphI-23, +1140A/C) cannot be ruled out as functional polymorphisms due to their complete LD with VNTR classes I and III (Barratt et al. 2004). The VNTR consists of repeats that are 14-15 bp long with a consensus motif

ACAGGGGTGTGGGG (Lucassen et al. 1993). The VNTR alleles are divided into three classes according to the number of repeats they entail. The shorter class I alleles contain 28-44 repeats (population frequency app. 70%) and longer class III alleles contain 138-159 repeats (population frequency app. 30%). Class II alleles are very rare in white Europeans. Homozygosity for the class I allele is associated with the T1D risk while the class III alleles convey a dominantly protective effect, although haplotypes that disprove this theory are found in both classes (Stead et al. 2000; Vafiadis et al. 2001).

The VNTR classes are associated with insulin transcription levels in the pancreatic islets and in the thymus (Kennedy et al. 1995; Pugliese et al. 1997; Vafiadis et al. 2001). Insulin is transcribed at 1.15-1.25-fold higher levels in the pancreas when VNTR class I instead of class III alleles are present (Bennett et al. 1995; Bennett et al. 1996; Vafiadis et al. 1996). Conversely to the pancreatic expression of insulin, the VNTR class III alleles are associated with significantly (2.4-2.7-fold) higher insulin transcription levels in medullary thymic epithelial cells (mTEC) than are the shorter class I alleles. The crucial step of the induction of the central tolerance occurs in the mTEC through negative selection of potentially autoreactive thymocytes (Pugliese et al. 1997; Vafiadis et al. 1997; Chentoufi et al. 2002). This has raised the hypothesis that the induction of self-tolerance to insulin in the thymus is the biological basis for the IDDM2 locus (Vafiadis et al. 1997). Consequently, VNTR would be the functional polymorphism since it allegedly regulates insulin transcription levels in the mTEC.

The mechanism by which VNTR supposedly regulates the thymic expression of insulin is unknown, leaving the possibility that other *cis*-acting polymorphisms or even unknown *trans*-acting factors could have a functional effect on the IDDM2 locus. The existence of such factors is consolidated to some extent by the observed disparity in the levels of insulin transcription among certain VNTR class III alleles which actually silence the insulin transcription in the mTEC and thus have a predisposing effect on T1D risk (Vafiadis et al. 2001). The two silencing alleles (S1, S2) are middle-sized class III alleles (app. 150 repeats) and do not differ strikingly from other class III alleles or from each other. Whether this disparity is explained by structural differences between the silencing alleles and other class III alleles or by linkage disequilibrium to other *cis* acting polymorphisms with a silencing effect, is unknown.

1.3.3.3 The CTLA4 region (IDDM12)

The cytotoxic T lymphocyte-associated antigen 4 (CTLA4) is expressed in the activated T cells in a similar manner to CD28. They both bind antigen-presenting cell receptors CD80 and CD86 and are involved in co-stimulatory events in an inhibitory (CTLA4) and stimulatory (CD28) manner (Magistrelli et al. 1999).

The IDDM12 region is located at 2q33 and was first reported in a linkage study of 48 Italian T1D sibling pair families (Nistico et al. 1996). Linkage evidence was restricted to a ~1 cM region defined by microsatellite markers D2S72-CTLA4(AT)n-D2S116. Nistico et al. also observed nominal evidence for increased transmission of one of the CTLA4(AT)n alleles to affected offspring. Since the microsatellite in question was located at the 3' end of a plausible candidate gene (CTLA4), they decided to perform an association study by carrying out a TDT for one non-synonymous coding SNP in five different populations. The cSNP, denoted as CTLA4 +49, is located in the first exon of the CTLA4 gene and it showed significant association with T1D in the combined dataset (p=0.002) (Nistico et al. 1996). Since that original report, linkage and association in the IDDM12 region has been replicated in several independent linkage and association studies in different populations (Marron et al. 1997; Van der Auwera et al. 1997; Lowe et al. 2000; Marron et al. 2000; Ueda et al. 2003) and the CTLA4 region has been established as a confirmed T1D susceptibility locus. The 2q33 region has also been implicated in other autoimmune diseases, such as systemic lupus erythematosus (SLE), autoimmune thyroid diseases (AITD), celiac disease (CD) and rheumatoid arthritis (RA) (Zhernakova et al. 2005) implying that the region harbors a general autoimmune locus.

The IDDM12 locus has been mapped to a 6.1 kb region at the 3' end of the CTLA4 gene in a comprehensive fine-mapping study with 108 SNPs (Ueda et al. 2003). The most associated marker in that region, CT60 (renamed as A6230G), presented a fairly low odds ratio of 1.51. The weak contribution of IDDM12 to the overall risk of T1D combined with a very strong LD throughout the associated region has hampered attempts at disclosing the actual disease polymorphism, and it still remains unconfirmed; although A6230G is a likely candidate (Anjos et al. 2004; Anjos et al. 2006).

1.3.3.4 PTPN22

PTPN22 is located at 1p13 and it encodes the lymphoid tyrosine phosphatase, LYP, which is an important inhibitor of the T cell activation (Cloutier et al. 1999; Gjorloff-Wingren et al. 1999). Association of PTPN22 with T1D was first discovered by Bottini et al., who investigated PTPN22's potential role as a susceptibility gene for autoimmune diseases. They described and analyzed a non-synonymous SNP at nucleotide 1858 of codon 620 which encodes arginine (1858 allele C) or tryptophan (1858 allele T). The 1858C-T SNP was tested for association in 294 North American T1D cases and 395 controls of the same origin, and it was found that the 1858T allele was significantly more common in the patients than in controls (odds ratio=1.83) (Bottini et al. 2004). This finding was replicated in a Sardinian dataset of 174 cases and 214 controls (odds ratio=2.31) (Bottini et al. 2004) and in 341 North American T1D multiplex families using TDT (Ladner et al. 2005). Since then, association of the 1858C-T polymorphism with T1D has been replicated in several studies (Smyth et al. 2004; Qu et al. 2005; Zheng et al. 2005; Hermann et al. 2006), and consequently, PTPN22 is considered to be the fourth confirmed T1D susceptibility locus to date.

The effect of PTPN22 is not T1D specific. It seems to affect autoimmunity on a more general level because, in addition to T1D, the 1858C-T polymorphism has shown association with many other autoimmune diseases, such as RA (Begovich et al. 2004; Gregersen et al. 2005), SLE (Kyogoku et al. 2004; Reddy et al. 2005), Grave's disease (Smyth et al. 2004; Velaga et al. 2004) and generalized vitiligo(Canton et al. 2005).

The 1858C-T polymorphism is currently the best candidate for a causal variant in the PTPN22 region. The efforts to resequence the PTPN22 gene have not discovered polymorphisms that are more strongly associated with T1D or RA than 1858C-T (Begovich et al. 2004; Bottini et al. 2006), but there is evidence for association with a non-1858T haplotype in RA (Carlton et al. 2005).

The disease-predisposing 1858T allele encodes Trp (W) at codon 620 (LYP*W620 protein), and it is a gain-of-function variant. LYP*W620 suppresses T cell receptor (TCR) signaling much more efficiently than LYP*R620 due to its significantly better functional capacity as a phosphatase (Vang et al. 2005). The more efficient suppression of TCR signaling that the W620 variant mediates has been hypothesized to result in increased positive

selection of T cells with moderate self-affinity towards MHC in the thymus and thus a higher disease risk (Bottini et al. 2006).

1.3.3.5 IDDM9

Linkage to the 3q21-q25 region was first detected in genome-wide linkage studies (Davies et al. 1994; Mein et al. 1998) and it was given the IDDM9 designation. The strongest linkage evidence found in these studies, albeit only suggestive, was seen in sample sets stratified by HLA or sex. A LOD score of 1.9 was seen in sibling pairs sharing two HLA alleles IBD (Davies et al. 1994) and a LOD score of 2.4 in 126 HLA-identical (DR3/DR4) UK sibling pairs (Mein et al. 1998). Furthermore, a LOD score of 2.0 was detected in the female sibling pairs of the aforementioned UK dataset (Paterson et al. 1999). The published genome-wide association scans for type 1 diabetes have not reported an association reaching moderate or genome-wide significance level at the 3q21-q25 region (Hakonarson et al. 2007; (WTCCC) 2007).

In addition to T1D, linkage evidence to the IDDM9 region has been found in other autoimmune diseases. A linkage scan of 308 European Caucasian sibling pairs with rheumatoid arthritis (RA) found linkage in the 3q21 region with marker D3S3576 (p=0.0001) (Cornelis et al. 1998). Additionally, a scan of 229 multiplex systemic lupus erythematosus (SLE) families stratified by different SLE autoantibodies found linkage to region 3q21 in families positive for anti-La/SSB (p=1.9E-6) (Ramos et al. 2006). The 3q21 region has also been implicated in linkage scans of psoriasis (Enlund et al. 1999), asthma and atopic dermatitis (Lee et al. 2000; Kurz et al. 2005). While these are not autoimmune diseases as such, they are common disorders with complex inheritance and erratic function of the immune system and potentially could share susceptibility genes with autoimmune diseases. These findings suggest that the 3q21 region could harbor a gene or genes involved in the production of autoimmunity and modulation of the immune system but the proof of its involvement with T1D susceptibility is still only suggestive.

1.3.3.6 Other susceptibility loci

In addition to the four confirmed T1D susceptibility loci, linkage evidence at or near the genome-wide significance level have been accumulated for at least three other loci (2q31-q33, 10p14-q11 and 16q22-q24) (Concannon et al. 2005). The IDDM7-IDDM12 region at 2q31-q33 contains the confirmed locus IDDM12 with association to the CTLA4 gene, but this region is hypothesized to contain an additional susceptibility locus or loci. This assertion is based on the stronger familial clustering observed by regional linkage analysis (λ_s~1.19) than the CTLA4 gene association alone (λ_s~1.01, predicted from the odds ratios 1.1-1.2) (Concannon et al. 2005).

The IDDM10 region at 10p14-q11 has shown suggestive linkage in several studies, although some of the datasets used in these studies are partially overlapping (Davies et al. 1994; Reed et al. 1997; Mein et al. 1998) (Concannon et al. 1998; Cox et al. 2001). The same theory mostly applies also to the 16q22-q24 region, although significant proof has been reported for this region (Mein et al. 1998; Cox et al. 2001; Concannon et al. 2005). Few follow-up studies have been performed at these regions even though consistent linkage evidence suggests that the IDDM10 and 16q22-q24 regions harbor a true T1D susceptibility gene.

The other previously reported non-HLA T1D susceptibility loci that are not mentioned above (Table 1) have not reportedly shown consistent evidence for linkage in follow-up studies (Concannon et al. 2005). This is probably due to initially false positive linkages or perhaps the alleged susceptibility loci at these regions have very small effect on the overall risk and have therefore remained undetected in the linkage scans.

2. AIMS OF THE STUDY

To find genetic loci outside the HLA region that affect T1D susceptibility in the Finnish population. This will be achieved by analyzing selected, previously suggested, non-HLA susceptibility loci for linkage to T1D in the Finnish population.

To further analyze the regions that in the first stage showed nominal or stronger linkage to T1D by narrowing the linked region sufficiently for association studies to be feasible.

To examine alleged interaction between confirmed and suggested T1D susceptibility loci and utilize this to decrease genetic heterogeneity in the sample set in order to increase statistical power for linkage and association analysis.

3. SUBJECTS, MATERIALS AND METHODS

3.1 Subjects

The multiplex T1D families with multiple affected siblings and parents (when available), as well as the simplex families with one affected child and parents, were collected within the project called Genetics of Type 1 Diabetes in Finland and The Finnish Pediatric Diabetes Registry.

The index cases of these families as well as the sporadic T1D subjects used in publication III were diagnosed with the typical ketosis-prone T1D and collected from all university hospitals and several central and regional hospitals in Finland. Samples representing the background population were taken from consecutively born children participating in the Finnish Type 1 Diabetes Prediction and Prevention Study (DIPP) at Turku, Oulu and Tampere university hospitals. All subjects with T1D were diagnosed under the age of 15 years, according to the WHO criteria. The studies were approved by the Ethics Committees of the participating hospitals and informed consent was obtained from the study subjects and/or their parents.

The 107 multiplex T1D families used in the study reported in the publication I were included in the 121 families used in the 3q21 region mapping (publication II). In addition, the 348 simplex families used in publication II were included in the 612 simplex families used in the publication IV.

3.2 Materials

Genomic DNA for family samples and for some of the case-control cohort samples were extracted from the whole blood using the salting-out method (Miller et al. 1988). A circle of dried blood spot on a filter paper was used directly in the PCR as a genomic DNA template for genotyping some of the insulin gene region case-control samples. The blood spot was first incubated with sodium hydroxide and then neutralized with a Tris-Cl buffer. 20 µl of extract containing genomic DNA was transferred into the amplification mixture.

3.3 Methods

3.3.1 Microsatellite genotyping

The primer sequences for the 23 microsatellite markers used in publication I and the 22 microsatellite markers used in the publication II were obtained from the original papers or electronic databases. Microsatellite markers were amplified using fluorescence-labelled PCR primers. Fluorescent labels FAM-6, HEX and TET representing the respective pseudo-colors blue, yellow and green were attached to the 5' end of the forward primer. The PCR reactions were performed using Hybaid Touchdown and Bio-Rad Tetrad thermal cyclers in 96-well plates (Robbins CyclePlate) in a 10 µl volume containing 0.2 U of the Dynazyme II DNA polymerase (Finnzymes Oy).

Automated sequencers ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and MegaBace 1000 (Amersham Biosciences, Buckinghamshire, UK) were used for electrophoresis and the GENOTYPER (Reed et al. 1994) software for ABI and the GENETIC PROFILER software for MegaBace were used for scoring the genotypes.

3.3.2 SNP genotyping

Genotyping for the SNP markers MspI-2221, HphI-23 and CTLA4+49 was performed using various methods. Genotyping of MspI-2221 in publications I, III and IV and genotyping of CTLA4+49 in publication I were performed using Europium-labelled oligonucleotide probes specific for both alleles in each SNP. The allele-specific probes were hybridised with a PCR-amplified gene segment in separate wells of a microtitration plate and detected with time-resolved fluorometry using a Victor 1420 fluorometer (PerkinElmer Wallac Oy, Turku, Finland) as described earlier (Sjöroos et al. 1995).

With the exception of the Swedish samples from Malmö in publication IV, which were analysed using the PCR-RFLP method, all HphI-23 genotyping

was performed using allele-specific Europium-labeled and Terbium-labeled oligonucleotide probes that were hybridised with a PCR-amplified gene segment in the same well of a microtitration plate. The detection of the alleles was performed in the same manner as in the MspI-2221 method.

3.3.3 Statistical analysis and software

Two-point and multipoint allele sharing and the corresponding maximum LOD scores given by the likelihood-ratio test in publications I and II were calculated using the MAPMAKER/SIBS program (Kruglyak et al. 1995). Dominance variance was allowed and the sharing probabilities were restricted to a possible triangle to increase the power of the likelihood-ratio test (Faraway 1993; Holmans 1993). λ_S-values for each marker were obtained by dividing the expected proportion of 0-sharing siblings in the case of no linkage (0.25) with the observed proportion (Risch 1987).

In publications I and II, a transmission/disequilibrium test (TDT) was used for detecting association in the multiplex and simplex families. In publication I, the ETDT software was used for calculating global corrected allele-wise p-values for each marker (Sham et al. 1995). In publication II, the TDTPHASE program of the UNPHASED software package (Dudbridge 2003) was used. Only unambiguous transmissions of alleles were calculated, no EM estimation of uncertain allele transmissions was used.

Power calculations for ASP and TDT analysis used in publication II were based on articles by Risch (Risch 1990) (ASP) and Risch and Merikangas (Risch et al. 1996) (TDT). A spreadsheet by Dave Curtis (POWTEST, http://www.smd.qmul.ac.uk/statgen/dcurtis/software.html) using formulas presented in the aforementioned publications was used for obtaining the specific power to detect linkage or association in our data set. POWTEST requires the user to set only lambda-s and the number of samples in ASP, and the genotype relative risk (GRR), population frequency of the disease allele and sample size in TDT. For the ASP calculations, a fully informative marker and $\theta=0$ was assumed. In the TDT calculations, the actual disease polymorphism or a marker with complete LD to the disease gene allele was assumed to be analyzed.

The Powermarker software (Liu et al. 2005) was used for assessing multi-allelic D' LD values (Zaykin et al. 1995) used in publication II.

In order to create affected-family-based artificial controls (AFBAC) (Thomson 1995) for the Finnish dataset in publication IV, the 612 Finnish T1D simplex families were haplotyped using the GENEHUNTER 2.0 software (Kruglyak et al. 1996). AFBACs were created so that in each family, the parental haplotypes not transmitted to the affected child were considered to belong to the AFBAC group and the transmitted haplotypes were considered as cases. The families were tested for inconsistencies between the parents and offspring with the PEDCHECK (O'Connell et al. 1998) program. Families showing possible non-paternity were removed from the AFBAC analysis.

3.3.4 Stratification

Genetic variability of disease etiology in complex diseases (locus and allele heterogeneity) causes loss of power in linkage and association analyses. In an attempt to compensate for this problem, sample sets were stratified according to their risk or protective genotypes in two known T1D susceptibility loci: the HLA and IDDM2 regions. Different HLA or IDDM2 susceptibility alleles or haplotypes could be associated with different disease mechanisms behind T1D etiology and thus stratifying samples according to them could increase statistical power to detect linkage or association.

Publication I:
Sibling pair families were stratified according to the HLA-DQB1 risk alleles DQB1*02 and DQB1*0302. Due to the limited number of families available in the studies (107), conditioning was restricted to two strata, where a reasonable number of families were available (>30), affected siblings sharing genotypes DQB1*0302/x (x≠DQB1*02) and DQB1*02/DQB1*0302. Families were also stratified according to sharing of alleles at the HLA-DQB1 locus to reveal non-allele-specific HLA-dependent linkages at other susceptibility loci. Families sharing two HLA-DQB1 alleles identical by state (IBS2, n=66) and families sharing one or zero HLA-DQB1 alleles IBS (IBS1,0, n=37) were analyzed separately.

Publication II:

Both data sets (multiplex and simplex) were stratified into six groups according to the HLA-DQB1 and INS *HphI* -23 genotypes. Affected children were genotype-identical in each group within and between families.

HLA groups: HLA-DQB1*02/HLA-DQB1*0302, HLA-DQB1*02/X (X≠DQB1*302), HLA-DQB1*0302/X (X≠DQB1*201) and sibling pairs with other genotypes (non-DQB1*02/non-DQB1*0302).

INS groups: those carrying the risk genotype AA and those carrying the protective genotypes AT or TT.

Publication III:

HLA genotypes were divided into three groups according to the disease risk they convey:

High risk genotypes (OR>5.0) included

*DQA1*05-DQB1*02/DRB1*0401/2/5-DQB1*0302* and

*DRB1*0401-DQB1*0302/x (x ≠ DQB1*02, *0301, *0602, *0603).*

Moderate risk (OR between 1.0 and 5.0) genotypes included

*DQA1*0201-DQB1*02/DRB1*0401/4-DQB1*0302,*

*DQA1*05-DQB1*02/DRB1*0402/3/8-DQB1*0302,*

*DQB1*0301/DRB1*0401/5-DQB1*0302,*

*DRB1*0401-DQB1*0302/DRB1*0404-DQB1*0302,*

*DRB1*0401-DQB1*0302/DRB1*0402-DQB1*0302,*

*DRB1*0404-DQB1*0302/x (x ≠ DQB1*02, DQB1*0301,*

*DQB1*0602, DQB1*0603),*

*DRB1*0401/5-DQB1*0302/DQB1*0603,*

*DQA1*05-DQB1*02/x (x ≠ DQA1*0201-DQB1*02, DQB1*0301, DQB1*0602, DQB1*0603).*

All other genotypes were associated with OR < 1.0 and considered to be low risk genotypes.

4. RESULTS AND DISCUSSION

4.1 *Several non-HLA loci show nominal evidence for linkage to type 1 diabetes in the Finnish sample set (Publication I). The 3q21 region shows evidence for linkage to and association with type 1 diabetes (Publication II).*

The Finnish multiplex dataset (107 T1D families) was analyzed using 24 genetic markers in 12 previously reported susceptibility regions (Table 3.). With the exception of the HLA region, no significant or suggestive linkage evidence (LOD 3.6 and 2.2, respectively (Lander et al. 1995)) was seen in the non-HLA-stratified dataset. Marker D3S3576 in the IDDM9 region was the only non-HLA marker that reached the nominal interval-wide significance level of p≈0.01 (Lander et al. 1995) (MLS=1.05) in the non-stratified dataset.

The lack of significant or suggestive linkage findings was not unexpected since the T1D susceptibility loci outside the HLA region are likely to have recurrence risk ratios for siblings (λ_s) smaller than 1.4 (Cox et al. 2001) and the expected maximum LOD score for 100 families is 0.96 if $\lambda s=1.4$ is assumed (Hauser et al. 1996). The statistical power in our 107 families should not differ significantly from that expectation. In fact, to gain even suggestive MLS values (2.2) with 100 families, a λ_s of ~2.0 or higher has to be assumed for the studied susceptibility locus (Hauser et al. 1996). Since the majority of T1D loci outside the HLA region most likely have λ_s smaller than 2.0, we should not expect to find suggestive evidence for linkage in our data set. Also the single-marker strategy in some of the studied loci makes the analysis more vulnerable to variation in marker informativity and subsequent loss of power.

One potential way to overcome the problem of limited statistical power caused by a small sample set is to decrease the genetic heterogeneity of the sample set by stratifying it according to known sources of genetic heterogeneity.

Using HLA-stratified datasets, MLS reaching nominal 1.1 over the unstratified LOD scores were seen in three loci. The multipoint maximum LOD score (MMLS) increased from 0.6 to 2.0 at CTLA4 (AT)n when siblings identical by state for two HLA-DQB1 alleles (IBS2) were analyzed. Additionally, MMLS increased from 0.7 to 1.1 at D18S64 and from 0.4 to 1.3 at D12S313 in the DQB1*302/x dataset (Table 3).

Four loci showed linkage in the presence of association (TDT performed in ASP families with all available siblings) in the unstratified data set and also when stratified by HLA (Table 4.). The highest increase in significance when stratified datasets were compared with the complete dataset was detected at IDDM12 in microsatellite CTLA4 (AT)n, where a global p-value (combined significance of all alleles) of 0.0006 was detected in the IBS2 stratified dataset. MspI –2221 at IDDM2 had a p=0.0094 in the 302/x stratified dataset, and p=0.01 in the 2/302 stratified dataset (p=0.03 in the complete dataset). Finally, the IDDM6 marker D18S64 gave a p-value of 0.0037 in the IBS1,0 dataset where siblings identical by state for one or zero HLA-DQB1 alleles were studied (p=0.46 in the complete dataset).

In addition to the already established linkage and association of type I diabetes with HLA and insulin gene regions in Finnish population, we found evidence for linkage in four additional suggested T1D susceptibility loci: the 12q12-q15 region, IDDM6, IDDM12 and IDDM9.

The proof for the interferon gamma (IFNG) region at 12q12-q15 has been accumulated from candidate gene association studies (Pociot et al. 1997; Jahromi et al. 2000; Tegoshi et al. 2002). We cannot say if the nominal linkage we observed for marker D12S313 in the 302/x stratified group is a false positive or reflects an actual finding. Recent type 1 diabetes genome-wide association study ((WTCCC) 2007) has found an association reaching genome-wide significance level at 12q13 in gene *ERBB3*. This finding could account for the linkage evidence found in our study.

IDDM6 has been mapped to 18q12-q21. The region around microsatellite marker D18S487 has shown varying levels of linkage and association to T1D in different populations (Merriman et al. 1997), including a sample of 104 simplex families from Finland (Merriman et al. 1998). The marker studied here, D18S64, was chosen based on a whole genome scan of 96 UK families (Davies et al. 1994). As D18S64 is ~10cM away from the D18S487 region, it is not optimal for confirming the association in the IDDM6 region, but it shows nominal linkage (MLS=1.1) to the 18q12-q21 region in the 302/x data set. In a genome scan of 356 UK ASPs, a LOD score of 3.2 was obtained with the DR4/x stratified group, while an unstratified cohort gave a LOD score of 1.2, thus suggesting DR4-dependent HLA interaction at IDDM6 (Lander et al. 1995). Our results are concordant with these findings since HLA-DQB1*0302 is found only in the DR4 haplotype, and more than 90% of DR4 haplotypes found in diabetic children are HLA-DQB1*0302-positive (Reijonen et al. 1990). A recent meta-analysis of several linkage studies in the UK, US and Scandinavian populations that combined data from 1435 T1D families

(Concannon et al. 2005) did not find significant or suggestive linkage proof for the IDDM6 region. They concluded that the reason for the lack of reproduction of previous findings was that they were false positives or that the alleged susceptibility locus that the IDDM6 region harbors has a very small effect on the overall risk and thus is undetectable due to etiologic heterogeneity in different populations.

IDDM12 is currently considered a confirmed T1D susceptibility locus which has been mapped to a 6.1 kb region at the 3' end of the CTLA4 gene (Ueda et al. 2003). We found relatively strong linkage to and association with T1D with the CTLA4(AT)n marker in the IBS2 group. We added four microsatellites and one SNP to a 3.6 cM region around the CTLA4(AT)n but the CTLA4(AT)n remained the most strongly linked (Figure 1) and associated (Table 4) marker in the region, which is consistent with the fact that it is located close to the associated region at the 3' end of the CTLA4 gene. Whether the linkage and association observed solely in the IBS2 group reflects possible interaction with the HLA locus is unclear. A chance event in unintentionally selecting the linked families to the IBS2 group could explain the results.

Since IDDM9 was the only region in addition to the HLA region that showed significant linkage in the unstratified dataset, three additional markers around D3S3576 were genotyped. The MMLS curve reveals a broad peak which tops between D3S3576 and the CD86 gene (Figure 2). CD86 is involved in T cell co-stimulation with CTLA4 and therefore is a good T1D candidate gene. An association study of two *CD86* SNPs in the Finnish population showed no association with T1D (Turpeinen et al. 2002) implying that *CD86* is not involved in T1D etiology. On the other hand these two markers cover only a fraction of genetic variation present at the *CD86* gene.

The variation in MMLS at the admittedly short span around CD86 that these markers cover is very small and therefore a larger linkage study with 22 microsatellite markers covering a 60 cM region around the reported IDDM9-region was performed. The 121 Finnish T1D multiplex families used comprised the original 107 families.

Table 3. Single and multipoint maximum LOD scores in total and stratified data sets (LOD scores ≥ 1.0 are bolded). Data for 2/302 and IBS1,0 are not shown due to the lack of significant results. A multipoint test was used to derive LOD scores from IDDM2, 9 and 12.

Locus	Location	Marker				MLS		HLA-DQB1 302/x		IBS 2 HLA-share
			%IBD	PIC	λs	n=107	%IBD	n=34	%IBD	n=66
IDDM1	6p21.3	DQB1	76.8	0.71	3.7	**10.8**				
		D6S273	77.2	0.74	3.14	**8.8**				
		TNFa	72.2	0.34	2.19	**8.2**				
		TNFb	68.1	0.78	1.92	**3.3**				
		TNFc	67.8	0.68	3.48	**1.7**				
IDDM2	11p15.5	TH	50	0.75		0				
		MspI –2221	52.3	0.14	1.1	0				
IDDM4	11q13	FGF3	50	0.73		0				
IDDM5	6q25	ESR	50	0.84		0				
IDDM6	18q21	D18S64	57.2	0.73	1.29	0.7	65.7	**1.1**		
IDDM9	3q21-q25	D3S1303	56.8	0.80	1.28	0.8				
		D3S3620	57.2	0.71	1.21	**1.0**				
		D3S3576	57.6	0.63	1.27	**1.0**				
		D3S1269	57.2	0.81	1.30	0.8				
IDDM10	10p11-q11	D10S193	57.2	0.77	1.41	0.7				
IDDM12	2q33	D2S2392	52.4	0.83	1.10	0.17				
		D2S116	53.5	0.76	1.16	0.15			58.7	**1.0**
		CTLA4 +49	54.0	0.37	1.13	0.32			61.0	**1.6**
		CTLA4 (AT)n	55.7	0.73	1.29	0.56			62.4	**2.0**
		D2S2189	54.7	0.54	1.23	0.27			62.0	**1.8**
		D2S1384	54.4	0.72	1.17	0.23			62.1	**1.8**
IFNG	12q15	D12S313	54.3	0.81	1.09	0.4	64.5	**1.3**		
Chr 1	1q42	D1S1617	53.2	0.80	1.07	0.2				
Chr X	Xp11.21	DXS991		0.78		0.6				

Table 4. Global ETDT p-values using an allele-wise test in complete and stratified data sets. (p-values ≤ 0.05 are presented in bold)

Locus	Location	Marker	ETDT p-value n=107	HLA-DQB1 302/x p-value n=34	HLA-DQB1 2/302 p-value n=39	IBS 1,0 HLA-share p-value n=37	IBS 2 HLA-share p-value n=66
IDDM1	6p21.3	DQB1	**<1.0E-7**				
		D6S273	**1.3E-4**				
		TNFa	**2.9E-4**				
		TNFb	0.11				
		TNFc	0.15				
IDDM2	11p15.5	TH	0.35				
		MspI	**0.03**	**0.0094**	**0.01**		
IDDM4	11q13	FGF3	0.53				
IDDM5	6q25	ESR	**0.03**				
IDDM6	18q21	D18S64	0.46			**0.0037**	
IDDM8	6q27	D6S264	0.51				
IDDM9	3q21-q25	D3S1303	0.14				
		D3S3620	0.53				
		D3S3576	0.13				
		D3S1269	**0.02**		**0.02**		
IDDM10	10p11-q11	D10S193	0.43				
IDDM12	2q33	D2S2392	0.78				
		D2S116	0.21		**0.05**		
		CTLA4 +49	0.74				
		CTLA4$(AT)_n$	**0.03**				**0.0006**
		D2S2189	0.30				
		D2S1384	0.14				**0.05**
IFNG	12q15	D12S313	0.21				
Chr 1	1q42	D1S1617	0.93				
Chr X	Xp11.21	DXS991	0.07				

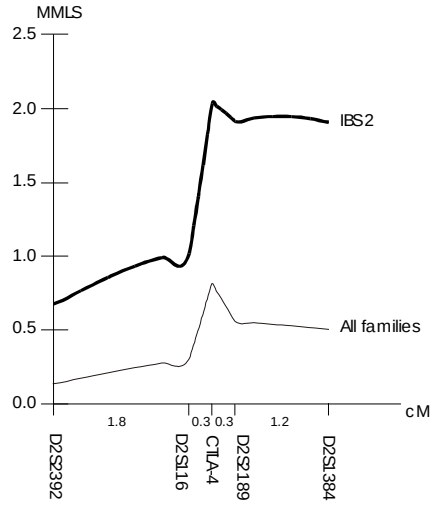

Figure 1. Plot of multipoint maximum LOD scores at the CTLA-4 region. The distance between markers across the map on the y axis is shown in cM.

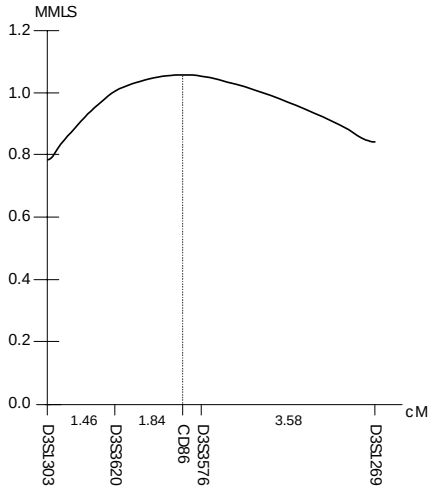

Figure 2. Plot of multipoint maximum LOD scores at the IDDM9 region. The distance between markers across the map on the y axis is shown in cM. Location of the CD86 gene is indicated by a vertical dashed line.

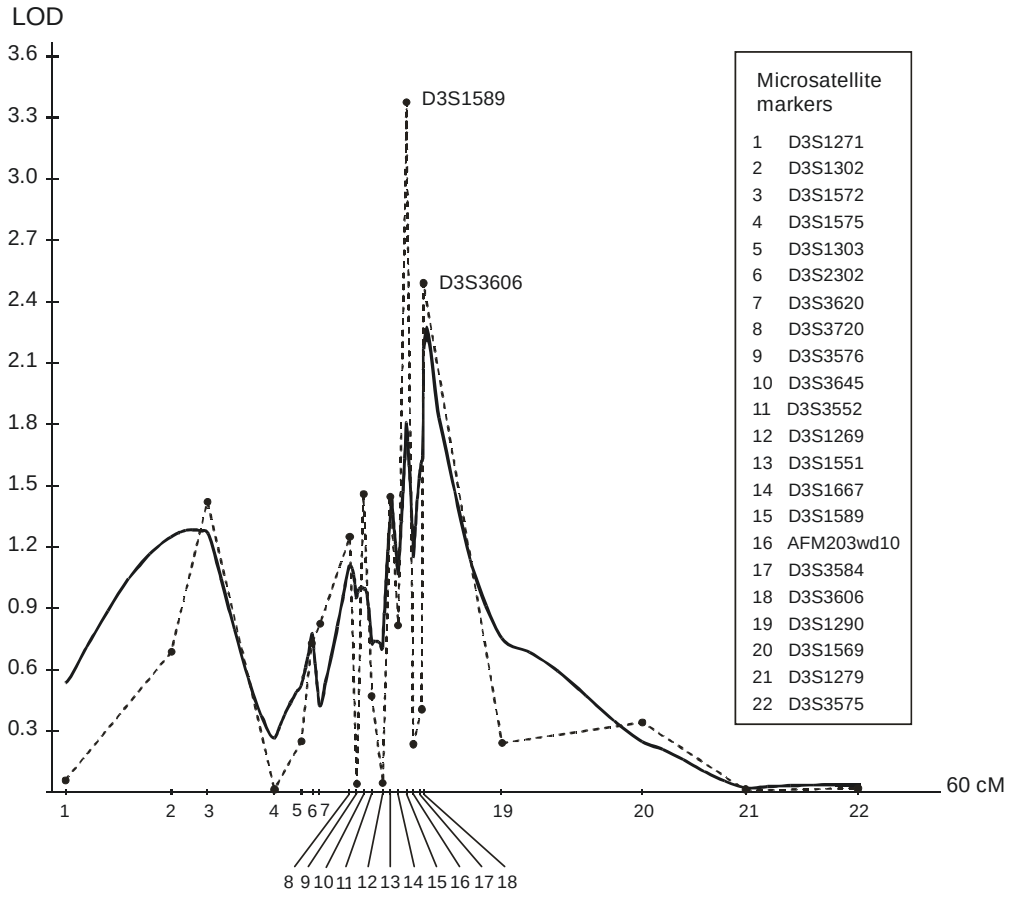

Figure 3.

Plot of multipoint (bolded line) and two point (dashed line) maximum LOD scores in the complete data set of 121 Finnish T1D multiplex families.

Table 5. Transmission of AFM203wd10 alleles in simplex families with affected and healthy children and after stratification for the insulin gene *HphI* -23 AA and AT/TT genotypes.

Affected sibs					
Allele	Transmitted	Not transmitted	Transmission %	chisq	p-value
R15	69	114	37.7	11.18	0.0008
R16	116	66	63.7	13.91	0.0002
R17	2	7	22.2	2.94	ns

Healthy sibs					
Allele	Transmitted	Not transmitted	Transmission %	chisq	p-value
R15	62	67	48.1	0.19	ns
R16	65	63	50.8	0.03	ns
R17	4	1	80	1.93	ns

HphI -23 AA					
Allele	Transmitted	Not transmitted	Transmission %	chisq	p-value
R15	48	88	35.3	11.94	0.0005
R16	91	46	66.4	15.06	0.0001
R17	1	6	14.3	3.96	0.05

HphI -23 AT/TT					
Allele	Transmitted	Not transmitted	Transmission %	chisq	p-value
R15	19	23	45.2	0.38	ns
R16	22	18	55	0.40	ns
R17	1	1	50	0	ns

Two extensive and distinct regions of above nominal linkage (MMLS≥1.2) were found. The centromeric region was defined by markers D3S1302 and D3S1572 with the peak two-point LOD score at D3S1572 (MLS=1.4) and the peak multipoint LOD score between markers D3S1302 and D3S1572 (MMLS=1.3). The distal region was defined by markers D3S1551 and D3S3606, with the highest two-point LOD scores at D3S1589 (MLS=3.3) and D3S3606 (MLS=2.5), and the highest multipoint LOD scores at D3S3606 (MMLS=2.2) and D3S1589 (MMLS=1.8). None of the HLA or *HphI*-23 stratified datasets revealed stronger linkage than the complete dataset.

The linkage findings reported here provide the strongest linkage evidence discovered thus far for T1D susceptibility at the 3q region. Suggestive LOD scores approaching genome-wide significance were observed with markers separated by 1.2 Mb (D3S1589, D3S3606). The drastic drop in LOD scores between D3S1589 and D3S3606 can be explained by the low heterogeneity of intervening markers AFM203wd10 and D3S3584, which have polymorphic information content (PIC) values of 0.37 and 0.57, respectively, compared to the PIC values of 0.68 and 0.83 for D3S1589 and D3S3606, respectively. These differences in marker informativity can cause 1.5-4.5-fold differences in the power to detect linkage (Risch 1990). In contrast to earlier studies, we could not find any clear evidence for HLA-dependent linkage in our data. Even though this does not rule out the possibility of interaction between the loci, it implies that HLA has no major modifying effect on the susceptibility conferred by the alleged 3q21 T1D locus.

Additional evidence for this interval was found in the TDT analysis using an independent sample set of 348 Finnish simplex T1D families. The microsatellite marker D3S3584 has only three alleles in the Finnish population and AFM203wd10 has effectively only two alleles which make them better suited for the TDT analysis than the more multiallelic microsatellites used in the linkage studies. These markers were specifically chosen for the TDT analysis because of their location within the strongest linkage peak.

D3S3584 did not show association with T1D, but significant association was found with AFM203wd10. Allele-wise significance for AFM203wd10 alleles R15 and R16 were p=0.0008 and p=0.0002 respectively using the TDTPHASE program (Table 5). Data sets stratified by the HLA genotypes did not show significantly increased or decreased transmission when compared to the complete data set or each other. The data set stratified by the *HphI* -23 T1D risk genotype AA showed a somewhat increased association when compared to the complete dataset (Table 5), but the difference was not significant. In

contrast, the combined dataset of the protective *HphI* -23 AT/TT genotypes did not show significant transmission disequilibrium (Table 5).

The associated marker AFM203wd10 is located 530kb distal to the most strongly linked marker D3S1589, and D3S3584 is located 700kb further downstream. D3S3584 did not show association, thus providing a suggestive distal limit for the associated region. Since D3S1589 is not well suited for association analysis due to its larger number of alleles, a set of SNPs around AFM203wd10 should be analyzed in an attempt to provide the centromeric limit for the associated region and to confirm the initial finding with AFM203wd10.

Association was not present in families stratified by the dominantly protective effect of the IDDM2-marker *HphI* -23 T allele positive haplotypes. In contrast, in the families homozygous for the predisposing A allele, an association of the same magnitude as in the complete data set was seen. This finding would suggest interaction between the 3q21 region and the insulin gene region in T1D etiology although the *HphI* -23 AT/TT stratified data set is fairly small and is therefore lacking the statistical power to make definite claims about possible interaction between these two loci.

Although our sibling pair dataset is underpowered to detect linkage in loci with weak or modest effect, the fact that we found an association in an independent simplex dataset with sufficient power to detect moderate to weak effects in TDT with observed significance argues against arbitrary causes explaining the linkage found at 3q21.

A spurious association could result from segregation distortion (Spielman et al. 1993) where the transmission ratio is distorted from the null hypothesis of 50% at the whole-population level. However, we performed the TDT with 212 healthy siblings of the affected children used in the TDT analysis, and no association was present, which excludes transmission ratio distortion as the source of association.

Before any definitive claims can be made about the role of the 3q21 locus in T1D etiology, more replication studies are needed. Due to the power limitations attached to the affected relative pair linkage methods, future efforts in confirming the finding and subsequently fine mapping the alleged locus should be directed primarily to the association methods using SNP markers.

4.2 Evidence for IDDM2 disease risk modulation by VNTR subclasses and various HLA genotypes (Publications IV and III).

The possible effect of the HLA class II genotypes upon the IDDM2 risk was analyzed by genotyping the VNTR class I and III associated *MspI*-2221 INS gene polymorphism in a Finnish case-control sample set of 1331 T1D children and 2222 controls and stratifying the cohorts according to the HLA risk they confer (Publication III).

As expected, there was a significant difference in the distribution of the *MspI*–2221 genotypes between patients and controls (Chi square 79.119, df=2, P<0.0001, OR=2.1 for the predisposing CC genotype) (Table 6). The VNTR class I homozygosity-associated CC genotype was slightly more common among diabetic children in the combined group of those with moderate or low HLA-associated risk than among diabetic children with high HLA-associated risk (Chi square 4.957, P=0.048) (Table 6). The distribution of the *MspI*-2221 genotypes differed significantly between the affected children with high-risk HLA genotypes and the combined group of affected children carrying low or moderate-risk HLA genotypes (Chi square 8.257, p=0.0126) (Table 6). No significant differences in the *MspI*-2221 genotypes were seen when healthy children with various HLA genotypes were compared.

We also compared the frequency of the *INS* gene genotypes in children with specific HLA risk alleles. Although the CC genotype was slightly more common among diabetic children with (DR3)-DQA1*05-DQB1*02 as the only risk-conferring haplotype (107/122, 86.3%) than in children with DRB1*04-DQB1*0302 as the only risk conferring haplotype (500/616, 81.2%), the difference remained non-significant.

Our data based on large numbers of children with T1D and healthy control children confirm the effect of an *INS* gene polymorphism on the risk for T1D but fail to suggest any specific interactions between the HLA risk alleles and the *INS* gene alleles. The effect of the *MspI*–2221 polymorphism was clear in all major HLA risk genotypes, although the effect was slightly stronger in children with low or moderate HLA-associated risk. According to the results presented here, the *MspI*–2221 INS polymorphism could be used as an additional risk predicting marker in subjects irrespective of their HLA genotypes. On the other hand since the *MspI*-2221 represents only poorly the multitude of haplotypes present at the IDDM2 region and its alleles are not in complete LD with VNTR classes I and III, possible interaction between the

HLA region and IDDM2 will have to be elucidated in a larger study, taking into account the diversity of the IDDM2 risk and protective haplotypes.

For this task, another insulin gene polymorphism strongly associated with T1D risk was analyzed. We developed a high-throughput test for the detection of the HphI-23 polymorphism and analyzed it together with MspI-2221 in the Finnish and Swedish datasets (Publication IV). The VNTR class I associated MspI-2221 C allele is also present in the protective class III haplotype, but HphI-23 alleles A and T are effectively in complete LD with VNTR classes I and III respectively. This allows more efficient use of HphI-23 as a VNTR surrogate marker. Analyzing HphI-23 instead of MspI-2221 could enhance the sensitivity and specificity of IDDM2 in the T1D risk estimation. In addition, since the VNTR class III subclasses IIIA and IIIB can be distinguished by respective MspI-2221 – HphI-23 haplotypes T-T and C-T, and if disparity in the levels of the protective effect from class III subclasses IIIA and IIIB exists, analyzing both markers would enable subclass separation and further enhance the accuracy of the T1D risk estimation.

As expected, MspI-2221 and HphI-23 were both associated with diabetes risk in Finland and Sweden (Table 7). The Finnish and Swedish populations did not show significant disparity in the odd ratios (OR) associated with the protective genotypes of the two SNPs. The VNTR class III homozygote (TT for both SNPs) showed a stronger protective effect than the heterozygote in both SNPs. This was true for both populations. The differences between the protective effects for the heterozygote and homozygote were significant in the combined data sets of the Finnish and Swedish samples ($p=0.01$, $OR=2.53$ for the MspI-2221, $p=0.02$, $OR=1.87$ for the HphI-23).

The risk genotypes of the two SNPs were significantly more common in Finland than in Sweden in both patients and controls, although the difference was slightly more pronounced among the controls ($p=0.003$ and 0.0003 for patients and controls respectively in MspI-2221, $p=0.0003$ and 0.0001 respectively in HphI-23). The combined homozygous protective genotype was twice as common in Sweden.

The combined MspI-2221 – HphI-23 genotypes and derived VNTR genotypes for patients and controls are shown in Table 8. As expected, the only genotype showing increased risk for T1D was VNTR class I/I genotype CC-AA. The VNTR subclass IIIA homozygote showed the strongest protection for T1D in both populations; the odds ratios were 0.21 and 0.18 respectively for Finland and Sweden. The IIIA/IIIA homozygote demonstrated a noticeably stronger protective effect than the I/IIIA heterozygote in both populations and a

significantly stronger effect when the combined dataset was analyzed (Table 9). The IIIB/IIIB genotype was rare in both populations and therefore the difference between the effects of the subclass IIIB heterozygote and subclass IIIB homozygote could not be tested reliably, but the comparison in the combined datasets seems to suggest that subclass IIIB conveys a weaker protective effect than subclass IIIA (Table 8, Table 9). Further proof of a divergence between the protective effects of IIIA and IIIB is the contrast between the significant difference observed in the protective effects between I/IIIA and IIIA/IIIA and the non-significant and close to neutral difference observed between the protective effects of I/IIIA and IIIA/IIIB. Yet more evidence for this discrepancy is provided by the clear (although non-significant) difference observed between the effects of IIIA/IIIB and IIIA/IIIA (OR=1.83) (Table 9.).

The common IDDM2 region lineages containing VNTR subclasses IIIA and IIIB are sometimes referred to as protective (PH) or very protective (VPH) haplotypes, respectively. Naturally, the haplotypes are named for the differences they have been observed to confer upon T1D protection (Bennett et al. 1995). Conversely to these previous findings, our data suggest a stronger protective effect for the more common subclass IIIA (PH) instead of the IIIB (VPH). The effect appears to be genotype-specific since the significance is due to the strong protective effect of the IIIA/IIIA genotype when it is compared to other genotypes containing the class III haplotypes. No significant or otherwise striking difference is found when the other class III genotypes are compared with each other. The significance of the differences seen in the protective effects here can still be considered nominal, especially if the p-values are corrected for multiple testing. A comprehensive study combining data from five populations did not find significant differences between the effects of PH and VPH, but it is not clear from the article if the IIIA/IIIA (PH/PH) homozygote was compared with the other genotype combinations (Barratt et al. 2004). Regardless of this, the fact that the same magnitude and direction of protective effects was seen in two independent populations could indicate genuine effects and warrants further studies using bigger cohorts and better-defined IDDM2 haplotypes. The observed variation in the protective effects of the different class III genotypes revealed by the subtyping of the class III lineages suggests that genotyping both -23HphI and -2221MspI could have potential for added specificity and sensitivity in T1D risk estimation in the insulin gene region.

The risk associated with the MspI-2221 and HphI-23 genotypes was very similar in Finland and Sweden even though frequencies of the risk and

protective genotypes and alleles differed significantly between the two populations. This difference in allele frequencies might contribute to the known difference in the incidence of T1D between the two populations (43.9 for Finland, 25.7 for Sweden) (Green et al. 2001), although a better understanding of the magnitude of the effect the IDDM2 locus conveys upon T1D morbidity should be gained before definite claims can be made.

Table 6. Distribution of the *MspI*-2221 genotypes among diabetic children and healthy controls (consecutively born infants in three cities in Finland). (χ^2=79.119, df=2, p<0.0001 for the difference in the distribution between diabetic children and control genotypes in the combined groups, p=0.0484 for the difference in the *MspI*-2221 CC genotype and p=0.0126 for the difference in the distribution of the *MspI*-2221 genotypes in diabetic children with various HLA-associated risk groups).

	All diabetic children and controls			
	T1D		Controls	
INS gene genotype	N	%	N	%
CC	1108	83.2	1557	70.1
CT	211	15.9	607	27.3
TT	12	0.9	58	2.6
	1331		2222	

	High-risk *HLA* genotypes (OR >5)			
	T1D		Controls	
INS gene genotype	N	%	N	%
CC	535	81.1	110	69.2
CT	115	17.4	46	28.9
TT	10	1.5	3	1.9
	660		159	
	Moderate-risk *HLA* genotypes (5> OR >1)			
	T1D		Controls	
INS gene genotype	N	%	N	%
CC	375	85.4	274	72.7
CT	64	14.6	94	24.9
TT	0	0.0	9	2.4
	439		377	
	Low-risk *HLA* genotypes (OR < 1)			
	T1D		Controls	
INS gene genotype	N	%	N	%
CC	198	85.3	1173	69.6
CT	32	13.8	467	27.7
TT	2	0.9	46	2.7
	232		1686	

Table 7. Genotype distributions of two single nucleotide polymorphisms in the insulin gene region among Finnish and Swedish children with type 1 diabetes (T1D) and controls (cont), and respective VNTR genotypes. P values are calculated for the difference in genotype distribution between cases and controls using the predisposing CC or AA genotypes as a reference (OR = Odds ratio. 95% CI = 95% confidence interval).

Finland								
MspI-2221	VNTR genotype	T1D	%	cont	%	p-value	OR	95% CI
CC	I/I or I/III	515	84.8	418	68.9		1.0	
CT	I/III or III/III	88	14.5	174	28.7	<1.0E-6	0.41	0.30-0.55
TT	III/III	4	0.7	15	2.5	0.006	0.22	0.06-0.70
sum		607		607				
HphI-23		T1D	%	cont	%	p-value	OR	95% CI
AA	I/I	483	79.6	362	59.6		1.0	
AT	I/III	115	18.9	218	35.9	<1.0E-6	0.40	0.30-0.52
TT	III/III	9	1.5	27	4.4	0.0003	0.25	0.11-0.56
sum		607		607				
Sweden								
MspI-2221		T1D	%	cont	%	p-value	OR	95% CI
CC	I/I or I/III	272	76.2	454	59.8		1.0	
CT	I/III or III/III	80	22.4	261	34.4	7.8E-6	0.51	0.38-0.69
TT	III/III	5	1.4	44	5.8	0.0002	0.19	0.07-0.51
sum		357		759				
HphI-23		T1D	%	cont	%	p-value	OR	95% CI
AA	I/I	244	68.3	380	50.1		1.0	
AT	I/III	101	28.3	310	40.8	1.3E-6	0.51	0.38-0.68
TT	III/III	12	3.4	69	9.1	3.2E-5	0.27	0.14-0.53
sum		357		759				
Finland+Sweden								
MspI-2221		T1D	%	cont	%	p-value	OR	95% CI
CC	I/I or I/III	787	81.7	872	63.8		1.0	
CT	I/III or III/III	168	17.4	435	31.9	<1.0E-6	0.43	0.35-0.53
TT	III/III	9	0.9	59	4.3	<1.0E-6	0.17	0.08-0.36
sum		964		1366				
HphI-23		T1D	%	cont	%	p-value	OR	95% CI
AA	I/I	727	75.4	742	54.3		1.0	
AT	I/III	216	22.4	528	38.7	<1.0E-6	0.42	0.34-0.51
TT	III/III	21	2.2	96	7.0	<1.0E-6	0.22	0.13-0.37
sum		964		1366				

Table 8. Combined genotypes for MspI-2221 and HphI-23 and their effect on diabetes risk in the Finnish, Swedish and combined populations. The VNTR genotypes are deduced from the MspI-2221 – HphI-23 haplotypes. P values are calculated for the difference in genotype distribution between the T1D patients and controls using the predisposing CC-AA genotype as a reference (OR = Odds ratio. 95% CI = 95% confidence interval. ns = non-significant).

Finland								
Genotype	VNTR genot.	T1D	%	Controls	%	p-value	OR	95% CI
CC-AA	I/I	482	79.4	362	59.6		1.00	
CC-AT	I/IIIB	32	5.3	54	8.9	0.0006	0.44	0.27-0.72
CC-TT	IIIB/IIIB	1	0.2	2	0.3			
CT-AA		1	0.2	0	0			
CT-AT	I/IIIA	83	13.7	163	26.9	<1.0E-6	0.38	0.28-0.52
CT-TT	IIIA/IIIB	4	0.7	11	1.8	0.04	0.27	0.07-0.94
TT-AA		0	0	0	0			
TT-AT		0	0	1	0.2			
TT-TT	IIIA/IIIA	4	0.7	14	2.3	0.007	0.21	0.06-0.70
sum		607		607				
Sweden								
CC-AA	I/I	243	68.1	378	49.8		1.00	
CC-AT	I/IIIB	27	7.6	70	9.2	0.04	0.60	0.36-0.99
CC-TT	IIIB/IIIB	2	0.6	6	0.8	ns	0.52	
CT-AA		1	0.3	2	0.3			
CT-AT	I/IIIA	74	20.7	239	31.5	3.3E-6	0.48	0.35-0.66
CT-TT	IIIA/IIIB	5	1.4	20	2.6	ns	0.39	0.12-1.11
TT-AA		0	0	0	0.0			
TT-AT		0	0	1	0.1			
TT-TT	IIIA/IIIA	5	1.4	43	5.7	0.0001	0.18	0.06-0.49
sum		357		759				
Fin+Swe								
CC-AA	I/I	725	75.2	740	54.2		1.00	
CC-AT	I/IIIB	59	6.1	124	9.1	1.5E-5	0.49	0.34-0.68
CC-TT	IIIB/IIIB	3	0.3	8	0.6	ns	0.38	0.08-1.59
CT-AA		2	0.2	2	0.1			
CT-AT	I/IIIA	157	16.3	402	29.4	<1.0E-6	0.40	0.32-0.49
CT-TT	IIIA/IIIB	9	0.9	31	2.3	0.001	0.30	0.13-0.65
TT-AA		0	0.0	0	0.0			
TT-AT		0	0.0	2	0.1			
TT-TT	IIIA/IIIA	9	0.9	57	4.2	<1.0E-6	0.16	0.07-0.34
sum		964		1366				

Table 9. Statistical significance of differences between the protective effects of the class III VNTR genotypes in the combined Finnish and Swedish dataset. P values are calculated using a 2x2 contingency table (OR = Odds ratio. 95% CI = 95% confidence interval).

Fin+Swe			
VNTR genotype	p-value	OR	95% CI
I/IIIA vs IIIA/IIIA	0.02	2.47	1.15-5.50
I/IIIA vs IIIA/IIIB	ns	1.35	0.60-3.12
I/IIIB vs IIIB/IIIB	ns	1.27	
I/IIIB vs IIIA/IIIA	0.006	3.01	1.33-7.02
IIIA/IIIB vs IIIA/IIIA	ns	1.83	0.59-5.73

SUMMARY

The results presented in this thesis study demonstrate the difficulties involved with the identification of complex diseases genes and in resolving the heterogeneity of a known susceptibility locus. Gaining significant proof for identifying or dissecting a susceptibility region requires a large number of samples in many independent study sets from several populations. Fortunately, results from different studies in one region can generally be combined so that the effort can be distributed into manageable portions which could justify carrying out studies with limited statistical power. This study could not indisputably authenticate the previous findings it aimed to resolve but did provide enough indicative evidence to warrant further studies on two regions:

1. The combined and independent evidence of linkage and association in the 3q21 region strongly suggests the existence of true T1D susceptibility locus/loci. However, previous evidence for this locus is weak and the results presented here could simply be a false positive caused by random sampling. Furthermore, even were this region to harbor true susceptibility locus/loci, their effect on the overall risk is likely to be weak which would require extensive sample sets using numerous SNP markers over a large region in the follow-up association studies. The cost of said study would be fairly high, but the evidence appears sufficiently strong to warrant it.

2. The evidence for HLA interaction or risk modification in the IDDM2 region found in this study is only suggestive, but an investigation using more rigorously characterized risk and protective IDDM2 haplotypes could reduce the heterogeneity enough to reveal the alleged interaction between the two loci on a significant level. The genotype specific variation found in the protective effects of the VNTR class III subtypes where the IIIA/IIIA homozygote confers significantly stronger protection than other class III genotypes in two independent populations certainly warrants follow-up studies in bigger cohorts and additional populations. The studies presented here demonstrate that dissecting IDDM2 region haplotypes is potentially beneficial for enhancing the type 1 diabetes risk prediction in the Finnish population.

ACKNOWLEDGEMENTS

This work was carried out at the Immunogenetics laboratory, Medicity, University of Turku.

I wish to express my gratitude towards my supervisor, Prof. Jorma Ilonen for providing me both scientific and financial means to complete my work uninterrupted.

I would also like to thank all my past and present co-workers for sharing the work and everyday life at the laboratory.

REFERENCES

Anjos, S. and C. Polychronakos. Mechanisms of genetic susceptibility to type I diabetes: beyond HLA. *Mol Genet Metab* 81: 187-95, 2004.

Anjos, S. M. and C. Polychronakos. Functional evaluation of the autoimmunity-associated CTLA4 gene: the effect of the (AT) repeat in the 3'untranslated region (UTR). *J Autoimmun* 27: 105-9, 2006.

Atkinson, M. A. and G. S. Eisenbarth. Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet* 358: 221-9, 2001.

Baekkeskov, S., H. J. Aanstoot, S. Christgau, A. Reetz, M. Solimena, M. Cascalho, F. Folli, H. Richter-Olesen and P. De Camilli. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature* 347: 151-6, 1990.

Barratt, B. J., F. Payne, C. E. Lowe, R. Hermann, B. C. Healy, D. Harold, P. Concannon, N. Gharani, M. I. McCarthy, M. G. Olavesen, R. McCormack, C. Guja, C. Ionescu-Tirgoviste, D. E. Undlien, K. S. Ronningen, K. M. Gillespie, E. Tuomilehto-Wolf, J. Tuomilehto, S. T. Bennett, D. G. Clayton, H. J. Cordell and J. A. Todd. Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes* 53: 1884-9, 2004.

Begovich, A. B., V. E. Carlton, L. A. Honigberg, S. J. Schrodi, A. P. Chokkalingam, H. C. Alexander, K. G. Ardlie, Q. Huang, A. M. Smith, J. M. Spoerke, M. T. Conn, M. Chang, S. Y. Chang, R. K. Saiki, J. J. Catanese, D. U. Leong, V. E. Garcia, L. B. McAllister, D. A. Jeffery, A. T. Lee, F. Batliwalla, E. Remmers, L. A. Criswell, M. F. Seldin, D. L. Kastner, C. I. Amos, J. J. Sninsky and P. K. Gregersen. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75: 330-7, 2004.

Bennett, S. T., A. M. Lucassen, S. C. Gough, E. E. Powell, D. E. Undlien, L. E. Pritchard, M. E. Merriman, Y. Kawaguchi, M. J. Dronsfield, F. Pociot

and et al. (1995). Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus [see comments]. Nat Genet. **9:** 284-92.

Bennett, S. T., A. J. Wilson, F. Cucca, J. Nerup, F. Pociot, P. A. McKinney, A. H. Barnett, S. C. Bain and J. A. Todd. IDDM2-VNTR-encoded susceptibility to type 1 diabetes: dominant protection and parental transmission of alleles of the insulin gene- linked minisatellite locus. *J Autoimmun* 9: 415-21., 1996.

Boomsma, D., A. Busjahn and L. Peltonen. Classical twin studies and beyond. *Nat Rev Genet* 3: 872-82, 2002.

Bottazzo, G. F., B. M. Dean, J. M. McNally, E. H. MacKay, P. G. Swift and D. R. Gamble. In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulitis. *N Engl J Med* 313: 353-60, 1985.

Bottazzo, G. F., A. Florin-Christensen and D. Doniach. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. *Lancet* 2: 1279-83, 1974.

Bottini, N., L. Musumeci, A. Alonso, S. Rahmouni, K. Nika, M. Rostamkhani, J. MacMurray, G. F. Meloni, P. Lucarelli, M. Pellecchia, G. S. Eisenbarth, D. Comings and T. Mustelin. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 36: 337-8, 2004.

Bottini, N., T. Vang, F. Cucca and T. Mustelin. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 18: 207-13, 2006.

Canton, I., S. Akhtar, N. G. Gavalas, D. J. Gawkrodger, A. Blomhoff, P. F. Watson, A. P. Weetman and E. H. Kemp. A single-nucleotide polymorphism in the gene encoding lymphoid protein tyrosine phosphatase (PTPN22) confers susceptibility to generalised vitiligo. *Genes Immun* 6: 584-7, 2005.

Carlson, C. S., M. A. Eberle, M. J. Rieder, Q. Yi, L. Kruglyak and D. A. Nickerson. Selecting a maximally informative set of single-nucleotide

polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74: 106-20, 2004.

Carlton, V. E., X. Hu, A. P. Chokkalingam, S. J. Schrodi, R. Brandon, H. C. Alexander, M. Chang, J. J. Catanese, D. U. Leong, K. G. Ardlie, D. L. Kastner, M. F. Seldin, L. A. Criswell, P. K. Gregersen, E. Beasley, G. Thomson, C. I. Amos and A. B. Begovich. PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *Am J Hum Genet* 77: 567-81, 2005.

Chentoufi, A. A. and C. Polychronakos. Insulin expression levels in the thymus modulate insulin-specific autoreactive T-cell tolerance: the mechanism by which the IDDM2 locus may predispose to diabetes. *Diabetes* 51: 1383-90, 2002.

Christie, M. R. Islet cell autoantigens in type 1 diabetes. *Eur J Clin Invest* 26: 827-38, 1996.

Cloutier, J. F. and A. Veillette. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* 189: 111-21, 1999.

Concannon, P., H. A. Erlich, C. Julier, G. Morahan, J. Nerup, F. Pociot, J. A. Todd and S. S. Rich. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 54: 2995-3001, 2005.

Concannon, P., K. J. Gogolin Ewens, D. A. Hinds, B. Wapelhorst, V. A. Morrison, B. Stirling, M. Mitra, J. Farmer, S. R. Williams, N. J. Cox, G. I. Bell, N. Risch and R. S. Spielman. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus [see comments]. *Nat Genet* 19: 292-6, 1998.

Copeman, J. B., F. Cucca, C. M. Hearne, R. J. Cornall, P. W. Reed, K. S. Ronningen, D. E. Undlien, L. Nistico, R. Buzzetti, R. Tosi and et al. Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. *Nat Genet* 9: 80-5, 1995.

Cordell, H. J. and J. A. Todd. Multifactorial inheritance in type 1 diabetes. *Trends Genet* 11: 499-504, 1995.

Cornelis, F., S. Faure, M. Martinez, J. F. Prud'homme, P. Fritz, C. Dib, H. Alves, P. Barrera, N. de Vries, A. Balsa, D. Pascual-Salcedo, K. Maenaut, R. Westhovens, P. Migliorini, T. H. Tran, A. Delaye, N. Prince, C. Lefevre, G. Thomas, M. Poirier, S. Soubigou, O. Alibert, S. Lasbleiz, S. Fouix, J. Weissenbach and et al. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci U S A* 95: 10746-50, 1998.

Cox, N. J., B. Wapelhorst, V. A. Morrison, L. Johnson, L. Pinchuk, R. S. Spielman, J. A. Todd and P. Concannon. Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. *Am J Hum Genet* 69: 820-30., 2001.

Cudworth, A. G. and J. C. Woodrow. Evidence for HL-A-linked genes in "juvenile" diabetes mellitus. *Br Med J* 3: 133-5, 1975.

Cudworth, A. G. and J. C. Woodrow. HL-A system and diabetes mellitus. *Diabetes* 24: 345-9, 1975.

Dahlquist, G. G., L. G. Blom, L. A. Persson, A. I. Sandstrom and S. G. Wall. Dietary factors and the risk of developing insulin dependent diabetes in childhood. *Bmj* 300: 1302-6, 1990.

Davies, J. L., Y. Kawaguchi, S. T. Bennett, J. B. Copeman, H. J. Cordell, L. E. Pritchard, P. W. Reed, S. C. Gough, S. C. Jenkins, S. M. Palmer and et al. A genome-wide search for human type 1 diabetes susceptibility genes [see comments]. *Nature* 371: 130-6, 1994.

Delepine, M., F. Pociot, C. Habita, L. Hashimoto, P. Froguel, J. Rotter, A. Cambon Thomsen, I. Deschamps, S. Djoulah, J. Weissenbach, J. Nerup, M. Lathrop and C. Julier. Evidence of a non-MHC susceptibility locus in type I diabetes linked to HLA on chromosome 6. *Am J Hum Genet* 60: 174-87, 1997.

DiabetesEpidemiologyResearchInternationalGroup. Geographic patterns of childhood insulin-dependent diabetes mellitus. Diabetes Epidemiology Research International Group. *Diabetes* 37: 1113-9, 1988.

Dudbridge, F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 25: 115-21, 2003.

Enlund, F., L. Samuelsson, C. Enerback, A. Inerot, J. Wahlstrom, M. Yhr, A. Torinsson, J. Riley, G. Swanbeck and T. Martinsson. Psoriasis susceptibility locus in chromosome region 3q21 identified in patients from southwest Sweden. *Eur J Hum Genet* 7: 783-90, 1999.

EURODIAB_ACEStudyGroup. Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. *Lancet* 355: 873-6, 2000.

Ewens, K. G., L. N. Johnson, B. Wapelhorst, K. O'Brien, S. Gutin, V. A. Morrison, C. Street, S. G. Gregory, R. S. Spielman and P. Concannon. Linkage and association with type 1 diabetes on chromosome 1q42. *Diabetes* 51: 3318-25, 2002.

Faraway, J. J. Improved sib-pair linkage test for disease susceptibility loci [published erratum appears in Genet Epidemiol 1994;11(1):99]. *Genet Epidemiol* 10: 225-33, 1993.

Field, L. L., R. Tobias and T. Magnus. A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus [see comments]. *Nat Genet* 8: 189-94, 1994.

Field, L. L., R. Tobias, G. Thomson and S. Plon. Susceptibility to insulin-dependent diabetes mellitus maps to a locus (IDDM11) on human chromosome 14q24.3-q31. *Genomics* 33: 1-8, 1996.

Foulis, A. K., M. A. Farquharson and A. Meager. Immunoreactive alpha-interferon in insulin-secreting beta cells in type 1 diabetes mellitus. *Lancet* 2: 1423-7, 1987.

Foulis, A. K., M. McGill and M. A. Farquharson. Insulitis in type 1 (insulin-dependent) diabetes mellitus in man--macrophages, lymphocytes, and interferon-gamma containing cells. *J Pathol* 165: 97-103, 1991.

Gale, E. A. The rise of childhood type 1 diabetes in the 20th century. *Diabetes* 51: 3353-61, 2002.

Gjorloff-Wingren, A., M. Saxena, S. Williams, D. Hammi and T. Mustelin. Characterization of TCR-induced receptor-proximal signaling events negatively regulated by the protein tyrosine phosphatase PEP. *Eur J Immunol* 29: 3845-54, 1999.

Gombos, Z., J. Wachowicz, R. Veijola, H. K. Åkerblom, O. Simell, M. Knip, J. Ilonen and R. Hermann. Human leukocyte antigen non-class II determinants for type 1 diabetes in the Finnish population. *Hum Immunol* 67: 714-21, 2006.

Green, A., E. A. Gale and C. C. Patterson. Incidence of childhood-onset insulin-dependent diabetes mellitus: the EURODIAB ACE Study. *Lancet* 339: 905-9, 1992.

Green, A. and C. C. Patterson. Trends in the incidence of childhood-onset diabetes in Europe 1989-1998. *Diabetologia* 44 Suppl 3: B3-8, 2001.

Gregersen, P. K. and F. Batliwalla. PTPN22 and rheumatoid arthritis: gratifying replication. *Arthritis Rheum* 52: 1952-5, 2005.

Hakonarson, H., S. F. Grant, J. P. Bradfield, L. Marchand, C. E. Kim, J. T. Glessner, R. Grabs, T. Casalunovo, S. P. Taback, E. C. Frackelton, M. L. Lawson, L. J. Robinson, R. Skraban, Y. Lu, R. M. Chiavacci, C. A. Stanley, S. E. Kirsch, E. F. Rappaport, J. S. Orange, D. S. Monos, M. Devoto, H. Q. Qu and C. Polychronakos. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448: 591-4, 2007.

Hanninen, A., S. Jalkanen, M. Salmi, S. Toikkanen, G. Nikolakaros and O. Simell. Macrophages, T cell receptor usage, and endothelial cell activation in the pancreas at the onset of insulin-dependent diabetes mellitus. *J Clin Invest* 90: 1901-10, 1992.

Hashimoto, L., C. Habita, J. P. Beressi, M. Delepine, C. Besse, A. Cambon Thomsen, I. Deschamps, J. I. Rotter, S. Djoulah, M. R. James and et al. Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q [see comments]. *Nature* 371: 161-4, 1994.

Hauser, E. R., M. Boehnke, S. W. Guo and N. Risch. Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 13: 117-37, 1996.

Hermann, R., M. Knip, R. Veijola, O. Simell, A. P. Laine, H. K. Åkerblom, P. H. Groop, C. Forsblom, K. Pettersson-Fernholm and J. Ilonen. Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes-indication of an increased environmental pressure? *Diabetologia* 46: 420-5., 2003.

Hermann, R., K. Lipponen, M. Kiviniemi, T. Kakko, R. Veijola, O. Simell, M. Knip and J. Ilonen. Lymphoid tyrosine phosphatase (LYP/PTPN22) Arg620Trp variant regulates insulin autoimmunity and progression to type 1 diabetes. *Diabetologia* 49: 1198-208, 2006.

Hirschhorn, J. N. Genetic epidemiology of type 1 diabetes. *Pediatr Diabetes* 4: 87-100, 2003.

Holmans, P. Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 52: 362-74, 1993.

Hyttinen, V., J. Kaprio, L. Kinnunen, M. Koskenvuo and J. Tuomilehto. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 52: 1052-5, 2003.

Jahromi, M., A. Millward and A. Demaine. A CA repeat polymorphism of the IFN-gamma gene is associated with susceptibility to type 1 diabetes. *J Interferon Cytokine Res* 20: 187-90., 2000.

Johansson, S., B. A. Lie, J. A. Todd, F. Pociot, J. Nerup, A. Cambon-Thomsen, I. Kockum, H. E. Akselsen, E. Thorsby and D. E. Undlien. Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1. *Genes Immun* 4: 46-53, 2003.

Kaprio, J., J. Tuomilehto, M. Koskenvuo, K. Romanov, A. Reunanen, J. Eriksson, J. Stengard and Y. A. Kesaniemi. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes

mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35: 1060-7, 1992.

Kennedy, G. C., M. S. German and W. J. Rutter. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription [see comments]. *Nat Genet* 9: 293-8, 1995.

Knip, M., R. Veijola, S. M. Virtanen, H. Hyöty, O. Vaarala and H. K. Åkerblom. Environmental triggers and determinants of type 1 diabetes. *Diabetes* 54 Suppl 2: S125-36, 2005.

Kruglyak, L., M. J. Daly, M. P. Reeve-Daly and E. S. Lander. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58: 1347-63, 1996.

Kruglyak, L. and E. S. Lander. Complete multipoint sib-pair analysis of qualitative and quantitative traits [see comments]. *Am J Hum Genet* 57: 439-54, 1995.

Kukko, M., T. Kimpimaki, S. Korhonen, A. Kupila, S. Simell, R. Veijola, T. Simell, J. Ilonen, O. Simell and M. Knip. Dynamics of diabetes-associated autoantibodies in young children with human leukocyte antigen-conferred risk of type 1 diabetes recruited from the general population. *J Clin Endocrinol Metab* 90: 2712-7, 2005.

Kurz, T., J. Altmueller, K. Strauch, F. Ruschendorf, A. Heinzmann, M. F. Moffatt, W. O. Cookson, F. Inacio, P. Nurnberg, H. H. Stassen and K. A. Deichmann. A genome-wide screen on the genetics of atopy in a multiethnic European population reveals a major atopy locus on chromosome 3q21.3. *Allergy* 60: 192-9, 2005.

Kyogoku, C., C. D. Langefeld, W. A. Ortmann, A. Lee, S. Selby, V. E. Carlton, M. Chang, P. Ramos, E. C. Baechler, F. M. Batliwalla, J. Novitzke, A. H. Williams, C. Gillett, P. Rodine, R. R. Graham, K. G. Ardlie, P. M. Gaffney, K. L. Moser, M. Petri, A. B. Begovich, P. K. Gregersen and T. W. Behrens. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 75: 504-7, 2004.

Kyvik, K. O., A. Green and H. Beck-Nielsen. Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. *Bmj* 311: 913-7, 1995.

Ladner, M. B., N. Bottini, A. M. Valdes and J. A. Noble. Association of the single nucleotide polymorphism C1858T of the PTPN22 gene with type 1 diabetes. *Hum Immunol* 66: 60-4, 2005.

Lan, M. S., C. Wasserfall, N. K. Maclaren and A. L. Notkins. IA-2, a transmembrane protein of the protein tyrosine phosphatase family, is a major autoantigen in insulin-dependent diabetes mellitus. *Proc Natl Acad Sci U S A* 93: 6367-70, 1996.

Lander, E. and L. Kruglyak. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results [see comments]. *Nat Genet* 11: 241-7, 1995.

Lee, Y. A., U. Wahn, R. Kehrt, L. Tarani, L. Businco, D. Gustafsson, F. Andersson, A. P. Oranje, A. Wolkertstorfer, A. v Berg, U. Hoffmann, W. Kuster, T. Wienker, F. Ruschendorf and A. Reis. A major susceptibility locus for atopic dermatitis maps to chromosome 3q21. *Nat Genet* 26: 470-3, 2000.

Lie, B. A., J. A. Todd, F. Pociot, J. Nerup, H. E. Akselsen, G. Joner, K. Dahl-Jorgensen, K. S. Ronningen, E. Thorsby and D. E. Undlien. The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet* 64: 793-800, 1999.

Liu, K. and S. V. Muse. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-9, 2005.

Lowe, R. M., J. Graham, G. Sund, I. Kockum, M. Landin-Olsson, J. B. Schaefer, C. Torn, A. Lernmark and G. Dahlquist. The length of the CTLA-4 microsatellite (AT)N-repeat affects the risk for type 1 diabetes. Diabetes Incidence in Sweden Study Group. *Autoimmunity* 32: 173-80., 2000.

Lucassen, A. M., C. Julier, J. P. Beressi, C. Boitard, P. Froguel, M. Lathrop and J. I. Bell. Susceptibility to insulin dependent diabetes mellitus maps

to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat Genet* 4: 305-10, 1993.

Luo, D. F., R. Buzzetti, J. I. Rotter, N. K. Maclaren, L. J. Raffel, L. Nistico, C. Giovannini, P. Pozzilli, G. Thomson and J. X. She. Confirmation of three susceptibility genes to insulin-dependent diabetes mellitus: IDDM4, IDDM5 and IDDM8. *Hum Mol Genet* 5: 693-8, 1996.

Magistrelli, G., P. Jeannin, N. Herbault, A. Benoit De Coignac, J. F. Gauchat, J. Y. Bonnefoy and Y. Delneste. A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. *Eur J Immunol* 29: 3596-602, 1999.

Marron, M. P., L. J. Raffel, H. J. Garchon, C. O. Jacob, M. Serrano-Rios, M. T. Martinez Larrad, W. P. Teng, Y. Park, Z. X. Zhang, D. R. Goldstein, Y. W. Tao, G. Beaurain, J. F. Bach, H. S. Huang, D. F. Luo, A. Zeidler, J. I. Rotter, M. C. Yang, T. Modilevsky, N. K. Maclaren and J. X. She. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet* 6: 1275-82., 1997.

Marron, M. P., A. Zeidler, L. J. Raffel, S. E. Eckenrode, J. J. Yang, D. I. Hopkins, H. J. Garchon, C. O. Jacob, M. Serrano-Rios, M. T. Martinez Larrad, Y. Park, J. F. Bach, J. I. Rotter, M. C. Yang and J. X. She. Genetic and physical mapping of a type 1 diabetes susceptibility gene (IDDM12) to a 100-kb phagemid artificial chromosome clone containing D2S72-CTLA4-D2S105 on chromosome 2q33. *Diabetes* 49: 492-9., 2000.

Mein, C. A., L. Esposito, M. G. Dunn, G. C. Johnson, A. E. Timms, J. V. Goy, A. N. Smith, L. Sebag Montefiore, M. E. Merriman, A. J. Wilson, L. E. Pritchard, F. Cucca, A. H. Barnett, S. C. Bain and J. A. Todd. A search for type 1 diabetes susceptibility genes in families from the United Kingdom [see comments]. *Nat Genet* 19: 297-300, 1998.

Merriman, T., R. Twells, M. Merriman, I. Eaves, R. Cox, F. Cucca, P. McKinney, J. Shield, D. Baum, E. Bosi, P. Pozzilli, L. Nistico, R. Buzzetti, G. Joner, K. Ronningen, E. Thorsby, D. Undlien, F. Pociot, J. Nerup, S. Bain, A. Barnett and J. Todd. Evidence by allelic association-

dependent methods for a type 1 diabetes polygene (IDDM6) on chromosome 18q21. *Hum Mol Genet* 6: 1003-10., 1997.

Merriman, T. R., I. A. Eaves, R. C. Twells, M. E. Merriman, P. A. Danoy, C. E. Muxworthy, K. M. Hunter, R. D. Cox, F. Cucca, P. A. McKinney, J. P. Shield, J. D. Baum, J. Tuomilehto, E. Tuomilehto-Wolf, C. Ionesco-Tirgoviste, G. Joner, E. Thorsby, D. E. Undlien, F. Pociot, J. Nerup, K. S. Ronningen, S. C. Bain and J. A. Todd. Transmission of haplotypes of microsatellite markers rather than single marker alleles in the mapping of a putative type 1 diabetes susceptibility gene (IDDM6). *Hum Mol Genet* 7: 517-24., 1998.

Miller, S. A., D. D. Dykes and H. F. Polesky. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16: 1215, 1988.

Morahan, G., D. Huang, B. D. Tait, P. G. Colman and L. C. Harrison. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science* 272: 1811-3, 1996.

Morahan, G., D. Huang, S. I. Ymer, M. R. Cancilla, K. Stephen, P. Dabadghao, G. Werther, B. D. Tait, L. C. Harrison and P. G. Colman. Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat Genet* 27: 218-21, 2001.

Mueller, J. C. Linkage disequilibrium for different scales and applications. *Brief Bioinform* 5: 355-64, 2004.

Nagamine, K., P. Peterson, H. S. Scott, J. Kudoh, S. Minoshima, M. Heino, K. J. Krohn, M. D. Lalioti, P. E. Mullis, S. E. Antonarakis, K. Kawasaki, S. Asakawa, F. Ito and N. Shimizu. Positional cloning of the APECED gene. *Nat Genet* 17: 393-8, 1997.

Nejentsev, S., Z. Gombos, A. P. Laine, R. Veijola, M. Knip, O. Simell, O. Vaarala, H. K. Åkerblom and J. Ilonen. Non-class II HLA gene associated with type 1 diabetes maps to the 240-kb region near HLA-B. *Diabetes* 49: 2217-21, 2000.

Nerup, J., P. Platz, O. O. Andersen, M. Christy, J. Lyngsoe, J. E. Poulsen, L. P. Ryder, L. S. Nielsen, M. Thomsen and A. Svejgaard. HL-A antigens and diabetes mellitus. *Lancet* 2: 864-6, 1974.

Nistico, L., R. Buzzetti, L. E. Pritchard, B. Van der Auwera, C. Giovannini, E. Bosi, M. T. Larrad, M. S. Rios, C. C. Chow, C. S. Cockram, K. Jacobs, C. Mijovic, S. C. Bain, A. H. Barnett, C. L. Vandewalle, F. Schuit, F. K. Gorus, R. Tosi, P. Pozzilli and J. A. Todd. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. *Hum Mol Genet* 5: 1075-80., 1996.

Noble, J. A., A. M. Valdes, M. Cook, W. Klitz, G. Thomson and H. A. Erlich. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* 59: 1134-48, 1996.

O'Connell, J. R. and D. E. Weeks. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63: 259-66, 1998.

Palmer, J. P., C. M. Asplin, P. Clemons, K. Lyen, O. Tatpati, P. K. Raghu and T. L. Paquette. Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science* 222: 1337-9, 1983.

Paterson, A. D. and A. Petronis. Sex of affected sibpairs and genetic linkage to type 1 diabetes. *Am J Med Genet* 84: 15-9., 1999.

Pociot, F., R. Veijola, J. Johannesen, P. M. Hansen, T. Lorenzen, A. E. Karlsen, H. Reijonen, M. Knip and J. Nerup. Analysis of an interferon-gamma gene (IFNG) polymorphism in Danish and Finnish insulin-dependent diabetes mellitus (IDDM) patients and control subjects. Danish Study Group of Diabetes in Childhood. *J Interferon Cytokine Res* 17: 87-93., 1997.

Pritchard, J. K. and M. Przeworski. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69: 1-14, 2001.

Pugliese, A., M. Zeller, A. Fernandez, Jr., L. J. Zalcberg, R. J. Bartlett, C. Ricordi, M. Pietropaolo, G. S. Eisenbarth, S. T. Bennett and D. D. Patel. 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: 293-7, 1997.

Qu, H., M. C. Tessier, T. J. Hudson and C. Polychronakos. Confirmation of the association of the R620W polymorphism in the protein tyrosine phosphatase PTPN22 with type 1 diabetes in a family based study. *J Med Genet* 42: 266-70, 2005.

Rabin, D. U., S. M. Pleasic, J. A. Shapiro, H. Yoo-Warren, J. Oles, J. M. Hicks, D. E. Goldstein and P. M. Rae. Islet cell antigen 512 is a diabetes-specific islet autoantigen related to protein tyrosine phosphatases. *J Immunol* 152: 3183-8, 1994.

Ramos, P. S., J. A. Kelly, C. Gray-McGuire, G. R. Bruner, A. N. Leiran, C. M. Meyer, B. Namjou, K. J. Espe, W. A. Ortmann, M. Reichlin, C. D. Langefeld, J. A. James, P. M. Gaffney, T. W. Behrens, J. B. Harley and K. L. Moser. Familial aggregation and linkage analysis of autoantibody traits in pedigrees multiplex for systemic lupus erythematosus. *Genes Immun* 7: 417-32, 2006.

Reddy, M. V., M. Johansson, G. Sturfelt, A. Jonsen, I. Gunnarsson, E. Svenungsson, S. Rantapaa-Dahlqvist and M. E. Alarcon-Riquelme. The R620W C/T polymorphism of the gene PTPN22 is associated with SLE independently of the association of PDCD1. *Genes Immun* 6: 658-62, 2005.

Redondo, M. J., M. Rewers, L. Yu, S. Garg, C. C. Pilcher, R. B. Elliott and G. S. Eisenbarth. Genetic determination of islet cell autoimmunity in monozygotic twin, dizygotic twin, and non-twin siblings of patients with type 1 diabetes: prospective twin study. *Bmj* 318: 698-702, 1999.

Reed, P., F. Cucca, S. Jenkins, M. Merriman, A. Wilson, P. McKinney, E. Bosi, G. Joner, K. Ronningen, E. Thorsby, D. Undlien, T. Merriman, A. Barnett, S. Bain and J. Todd. Evidence for a type 1 diabetes susceptibility locus (IDDM10) on human chromosome 10p11-q11. *Hum Mol Genet* 6: 1011-6, 1997.

Reed, P. W., J. L. Davies, J. B. Copeman, S. T. Bennett, S. M. Palmer, L. E. Pritchard, S. C. Gough, Y. Kawaguchi, H. J. Cordell, K. M. Balfour and

et al. Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping [see comments]. *Nat Genet* 7: 390-5, 1994.

Reijonen, H., J. Ilonen, M. Knip, B. Michelsen and H. K. Åkerblom. HLA-DQ beta-chain restriction fragment length polymorphism as a risk marker in type 1 (insulin-dependent) diabetes mellitus: a Finnish family study. *Diabetologia* 33: 357-62., 1990.

Risch, N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 40: 1-14, 1987.

Risch, N. Linkage strategies for genetically complex traits. I. Multilocus models [see comments]. *Am J Hum Genet* 46: 222-8, 1990.

Risch, N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs [see comments]. *Am J Hum Genet* 46: 229-41, 1990.

Risch, N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs [published erratum appears in Am J Hum Genet 1992 Sep;51(3):673-5] [see comments]. *Am J Hum Genet* 46: 242-53, 1990.

Risch, N. and K. Merikangas. The future of genetic studies of complex human diseases [see comments]. *Science* 273: 1516-7, 1996.

Rowe, R. E., B. Wapelhorst, G. I. Bell, N. Risch, R. S. Spielman and P. Concannon. Linkage and association between insulin-dependent diabetes mellitus (IDDM) susceptibility and markers near the glucokinase gene on chromosome 7. *Nat Genet* 10: 240-2, 1995.

Rytkonen, M., E. Moltchanova, J. Ranta, O. Taskinen, J. Tuomilehto and M. Karvonen. The incidence of type 1 diabetes among children in Finland--rural-urban difference. *Health Place* 9: 315-25, 2003.

Rytkonen, M., J. Ranta, J. Tuomilehto and M. Karvonen. Bayesian analysis of geographical variation in the incidence of Type I diabetes in Finland. *Diabetologia* 44 Suppl 3: B37-44, 2001.

Sham, P. C. and D. Curtis. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 59: 323-36, 1995.

Singal, D. P. and M. A. Blajchman. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 22: 429-32, 1973.

Sjöroos, M., A. Iitia, J. Ilonen, H. Reijonen and T. Lovgren. Triple-label hybridization assay for type-1 diabetes-related HLA alleles. *Biotechniques* 18: 870-7., 1995.

Smyth, D., J. D. Cooper, J. E. Collins, J. M. Heward, J. A. Franklyn, J. M. Howson, A. Vella, S. Nutland, H. E. Rance, L. Maier, B. J. Barratt, C. Guja, C. Ionescu-Tirgoviste, D. A. Savage, D. B. Dunger, B. Widmer, D. P. Strachan, S. M. Ring, N. Walker, D. G. Clayton, R. C. Twells, S. C. Gough and J. A. Todd. 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: 3020-3, 2004.

Smyth, D. J., J. D. Cooper, R. Bailey, S. Field, O. Burren, L. J. Smink, C. Guja, C. Ionescu-Tirgoviste, B. Widmer, D. B. Dunger, D. A. Savage, N. M. Walker, D. G. Clayton and J. A. Todd. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38: 617-9, 2006.

Spielman, R. S., R. E. McGinnis and W. J. Ewens. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52: 506-16, 1993.

Stead, J. D., J. Buard, J. A. Todd and A. J. Jeffreys. Influence of allele lineage on the role of the insulin minisatellite in susceptibility to type 1 diabetes. *Hum Mol Genet* 9: 2929-35., 2000.

Tegoshi, H., G. Hasegawa, H. Obayashi, K. Nakano, Y. Kitagawa, M. Fukui, S. Matsuo, M. Deguchi, M. Ohta, M. Nishimura, N. Nakamura and T. Yoshikawa. Polymorphisms of interferon-gamma gene CA-repeat and interleukin-10 promoter region (-592A/C) in Japanese type I diabetes. *Hum Immunol* 63: 121-8., 2002.

TheWHOMultinationalProjectForChildhoodDiabetesGroup. Familial insulin-dependent diabetes mellitus (IDDM) epidemiology: standardization of data for the DIAMOND Project. The WHO Multinational Project for Childhood Diabetes Group. *Bull World Health Organ* 69: 767-77, 1991.

Thomson, G. Mapping disease genes: family-based association studies. *Am J Hum Genet* 57: 487-98, 1995.

Todd, J. A., J. I. Bell and H. O. McDevitt. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 329: 599-604, 1987.

Todd, J. A., N. M. Walker, J. D. Cooper, D. J. Smyth, K. Downes, V. Plagnol, R. Bailey, S. Nejentsev, S. F. Field, F. Payne, C. E. Lowe, J. S. Szeszko, J. P. Hafler, L. Zeitels, J. H. Yang, A. Vella, S. Nutland, H. E. Stevens, H. Schuilenburg, G. Coleman, M. Maisuria, W. Meadows, L. J. Smink, B. Healy, O. S. Burren, A. A. Lam, N. R. Ovington, J. Allen, E. Adlem, H. T. Leung, C. Wallace, J. M. Howson, C. Guja, C. Ionescu-Tirgoviste, M. J. Simmonds, J. M. Heward, S. C. Gough, D. B. Dunger, L. S. Wicker and D. G. Clayton. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*, 2007.

Turpeinen, H., A. P. Laine, S. Nejentsev, M. Sjöroos, J. Ilonen, O. Simell, R. Veijola, M. Knip and H. K. Åkerblom. CD86 gene polymorphisms: no association with Type I diabetes among Finnish subjects. *Diabetologia* 45: 1041-2., 2002.

Ueda, H., J. M. Howson, L. Esposito, J. Heward, H. Snook, G. Chamberlain, D. B. Rainbow, K. M. Hunter, A. N. Smith, G. Di Genova, M. H. Herr, I. Dahlman, F. Payne, D. Smyth, C. Lowe, R. C. Twells, S. Howlett, B. Healy, S. Nutland, H. E. Rance, V. Everett, L. J. Smink, A. C. Lam, H. J. Cordell, N. M. Walker, C. Bordin, J. Hulme, C. Motzo, F. Cucca, J. F. Hess, M. L. Metzker, J. Rogers, S. Gregory, A. Allahabadia, R. Nithiyananthan, E. Tuomilehto-Wolf, J. Tuomilehto, P. Bingley, K. M. Gillespie, D. E. Undlien, K. S. Ronningen, C. Guja, C. Ionescu-Tirgoviste, D. A. Savage, A. P. Maxwell, D. J. Carson, C. C. Patterson, J. A. Franklyn, D. G. Clayton, L. B. Peterson, L. S. Wicker, J. A. Todd and S. C. Gough. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423: 506-11, 2003.

Undlien, D. E., T. Friede, H. G. Rammensee, G. Joner, K. Dahl-Jorgensen, O. Sovik, H. E. Akselsen, I. Knutsen, K. S. Ronningen and E. Thorsby. HLA-encoded genetic predisposition in IDDM: DR4 subtypes may be associated with different degrees of protection. *Diabetes* 46: 143-9, 1997.

Vafiadis, P., S. T. Bennett, E. Colle, R. Grabs, C. G. Goodyer and C. Polychronakos. Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes. *J Autoimmun* 9: 397-403, 1996.

Vafiadis, P., S. T. Bennett, J. A. Todd, J. Nadeau, R. Grabs, C. G. Goodyer, S. Wickramasinghe, E. Colle and C. Polychronakos. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15: 289-92, 1997.

Vafiadis, P., H. Ounissi-Benkalha, M. Palumbo, R. Grabs, M. Rousseau, C. G. Goodyer and C. Polychronakos. Class III alleles of the variable number of tandem repeat insulin polymorphism associated with silencing of thymic insulin predispose to type 1 diabetes. *J Clin Endocrinol Metab* 86: 3705-10, 2001.

Van Autreve, J. E., B. P. Koeleman, E. Quartier, F. Aminkeng, I. Weets, F. K. Gorus and B. J. Van der Auwera. MICA is associated with type 1 diabetes in the Belgian population, independent of HLA-DQ. *Hum Immunol* 67: 94-101, 2006.

Van der Auwera, B. J., C. L. Vandewalle, F. C. Schuit, F. Winnock, I. H. De Leeuw, S. Van Imschoot, G. Lamberigts and F. K. Gorus. CTLA-4 gene polymorphism confers susceptibility to insulin-dependent diabetes mellitus (IDDM) independently from age and from other genetic or immune disease markers. The Belgian Diabetes Registry. *Clin Exp Immunol* 110: 98-103, 1997.

Vang, T., M. Congia, M. D. Macis, L. Musumeci, V. Orru, P. Zavattari, K. Nika, L. Tautz, K. Tasken, F. Cucca, T. Mustelin and N. Bottini. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet* 37: 1317-9, 2005.

Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 6: 109-18, 2005.

Velaga, M. R., V. Wilson, C. E. Jennings, C. J. Owen, S. Herington, P. T. Donaldson, S. G. Ball, R. A. James, R. Quinton, P. Perros and S. H. Pearce. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. *J Clin Endocrinol Metab* 89: 5862-5, 2004.

Verge, C. F., N. J. Howard, L. Irwig, J. M. Simpson, D. Mackerras and M. Silink. Environmental factors in childhood IDDM. A population-based, case-control study. *Diabetes Care* 17: 1381-9, 1994.

Verge, C. F., P. Vardi, S. Babu, F. Bao, H. A. Erlich, T. Bugawan, D. Tiosano, L. Yu, G. S. Eisenbarth and P. R. Fain. Evidence for oligogenic inheritance of type 1 diabetes in a large Bedouin Arab family. *J Clin Invest* 102: 1569-75., 1998.

Virtanen, S. M., E. Hypponen, E. Laara, P. Vahasalo, P. Kulmala, K. Savola, L. Rasanen, A. Aro, M. Knip and H. K. Åkerblom. Cow's milk consumption, disease-associated autoantibodies and type 1 diabetes mellitus: a follow-up study in siblings of diabetic children. Childhood Diabetes in Finland Study Group. *Diabet Med* 15: 730-8, 1998.

Virtanen, S. M., E. Laara, E. Hypponen, H. Reijonen, L. Rasanen, A. Aro, M. Knip, J. Ilonen and H. K. Åkerblom. Cow's milk consumption, HLA-DQB1 genotype, and type 1 diabetes: a nested case-control study of siblings of children with diabetes. Childhood diabetes in Finland study group. *Diabetes* 49: 912-7, 2000.

(WTCCC), W. T. C. C. C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-78, 2007.

Zavattari, P., R. Lampis, C. Motzo, M. Loddo, A. Mulargia, M. Whalen, M. Maioli, E. Angius, J. A. Todd and F. Cucca. Conditional linkage disequilibrium analysis of a complex disease superlocus, IDDM1 in the HLA region, reveals the presence of independent modifying gene

effects influencing the type 1 diabetes risk encoded by the major HLA-DQB1, -DRB1 disease loci. *Hum Mol Genet* 10: 881-9, 2001.

Zaykin, D., L. Zhivotovsky and B. S. Weir. Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* 96: 169-78, 1995.

Zheng, W. and J. X. She. Genetic association between a lymphoid tyrosine phosphatase (PTPN22) and type 1 diabetes. *Diabetes* 54: 906-8, 2005.

Zhernakova, A., P. Eerligh, P. Barrera, J. Z. Wesoly, T. W. Huizinga, B. O. Roep, C. Wijmenga and B. P. Koeleman. CTLA4 is differentially associated with autoimmune diseases in the Dutch population. *Hum Genet* 118: 58-66, 2005.

ORIGINAL PUBLICATIONS