Understanding type 1 diabetes genetics – approaches for identification of susceptibility genes in multi-factorial diseases

Regine Bergholdt

This review has been accepted as a thesis together with eight previously published papers by the University of Copenhagen, October 7, 2008, and defended on January 14, 2009.

Steno diabetes Center and Hagedorn Research Institute, Denmark.

Correspondence: Elletoften 35, 2800 Kgs. Lyngby, Denmark.

E-mail: rber@hagedorn.dk

Official opponents: Dag E. Undlien, Norway, and Steen Madsbad.

Dan Med Bull 2009;56:1-39

1. PRE-INTRODUCTION

The aim of this thesis is to review and discuss different approaches for identifying susceptibility genes and evaluate their evidence in multi-factorial diseases using type 1 diabetes (T1D) as a model. The intention has not been to evaluate the individual methods in detail, but to discuss classical and novel methods in a conceptual perspective. The underlying scientific work is described in the papers (Bergholdt et al. 1995; Bergholdt et al. 2000; Bergholdt et al. 2004a; Bergholdt et al. 2005a; Bergholdt et al. 2005c; Bergholdt et al. 2006; Bergholdt et al. 2007a; Bergholdt et al. 2007b) and this thesis will evaluate our own data in the context of what has been, and has become, approaches for mapping disease genes. T1D can be seen as a model for diseases of multi-factorial origin and even though there are differences among such diseases, the concepts applied for the present studies are expected to be applicable to the genetics of other complex diseases. Multi-factorial diseases are not caused by variation in a single or a few genes but rather result from variants in several genes, each contributing only a weak effect.

The main purpose of this thesis is to review how numerous disease genes may be identified. More recent approaches integrate biological and genetic information in gene mapping and focus increasingly on genome-wide approaches. The developments in gene identification have been tremendous. However, despite new methods and large studies, the overall impression is that data inconsistencies exist and novel approaches integrating genetic and functional information are needed. These may include novel analytical methods, such as global/genome-wide approaches, use of information from other autoimmune diseases and rare syndromes, more functional approaches as well as studies of interactions at the genetic or functional level.

The thesis will focus on approaches for identification of susceptibility genes and cover approaches applied in the papers (Bergholdt et al. 1995; Bergholdt et al. 2000; Bergholdt et al. 2004a; Bergholdt et al. 2005a; Bergholdt et al. 2005c; Bergholdt et al. 2006; Bergholdt et al. 2007a; Bergholdt et al. 2007b). Candidate gene approaches, genome-wide linkage and association analyses, examples of the use of comparative genetics as well as novel approaches for identifying genes and gene-gene interactions, and approaches for functional prioritization of candidate genes will be covered and discussed in the context of our own work.

2. INTRODUCTION – TYPE 1 DIABETES

Type 1 diabetes (T1D) [MIM 222100] is a chronic disease, affecting up to 0.4% of individuals in some populations by the age of 30 years, with an overall lifetime risk of nearly 1% (Mølbak et al. 1994; Karvonen et al. 2000). T1D is caused by absolute insulin deficiency due to destruction of the pancreatic beta cells. The majority of T1D cases are believed to develop as a result of immune-mediated destruction of the beta cells, leaving a small proportion of idiopathic cases in which immune markers cannot be detected with existing methodology. These are probably caused by other pathogenetic mechanisms such as rare genetic syndromes, β-cell lytic virus infections, or environmental factors (Alberti and Zimmet 1998). T1D is associated with an increased risk of premature death due to acute factors as well as often accompagnied by chronic disabling complications, including retinopathy and blindness, nephropathy, neuropathy and cardiovascular disease (Borch-Johnsen 1989). The incidence of T1D is increasing in most parts of the world and age at onset has, in some populations, decreased over the last decades (Onkamo et al. 1999; Svensson et al. 2002; Hyttinen et al. 2003). T1D incidence is, however, highly variable among different ethnic populations, ranging from 0.1/100.000 per year in a region of China to more than 40/100.000 per year in Finland (Onkamo et al. 1999; Green et al. 2000), equivalent to a 400-fold variation in incidence. The explanation for these wide differences in risk among ethnic groups, are most probably due to differences in both genes and environment and their interactions.

The etiology of T1D is unknown, but it is recognized that both genetic and environmental determinants are important in defining disease risk. That T1D has a strong genetic component is well established (Hirschhorn 2003) and is supported by studies of familial aggregation, monozygotic (MZ) and dizygotic (DZ) twins, as well as genetic admixture studies demonstrating that the closer the genetic relationship between two individuals is, the more likely they are to be concordant for the disease. MZ twins share 100% of their genes while DZ twins only have 50% of their genes in common, both supposedly share environmental risk factors. In T1D the concordance for MZ twins is estimated to be 27-60%, whereas a range of 2-13% is seen for DZ twins (Hyttinen et al. 2003). A model with additive genetic and individual environmental effects estimated that 88% of phenotypic variance in a nation-wide Finnish twin study was due to genetic factors and the remaining variance due to unshared environmental factors (Hyttinen et al. 2003). The degree of familial clustering can be expressed as the sibling relative risk, λ_S, a measure for the ratio of disease prevalence in a sibling of a diabetic patient. λ_S is defined as the recurrence rate for a sibling divided by the risk in the general population (Risch 1987), Figure 1. The higher the λ_S, the easier it is to find or map the genetic cause of a disease. The λ_S value for T1D is 15, and for T2D only 3. That λ_S is higher for T1D than T2D, even though familial clustering is stronger for T2D than T1D (40% vs. 10%), is due to the much higher frequency of T2D in the general population. The absolute risk of disease for a sibling of a T1D patient in European populations is 6% (Thomson et al. 1988; Waldhör et al. 1999; Harjutsalo et al. 2005) compared to

Familial clustering:

$$\lambda_S = \frac{\text{risk to siblings}}{\text{population prevalence}} = \frac{6\%}{0.4\%} = 15$$

For a specific genetic locus:

$$\lambda_{HLA} = \frac{\text{exp. IBD 0}}{\text{obs. IBD 0}} = \frac{25\%}{8\%} = 3.1$$

HLA contribution: $15^x = 3.1 \Rightarrow \log 15^x = \log 3.1 \Rightarrow x = \log 3.1 / \log 15 = 0.42$

Figure 1. Calculation of familial clustering of T1D, as well as an example of the genetic contribution from a specific locus, the HLA locus, in a multiplicative model. The percentages are average numbers based on the current literature.

0.4% for the general population (Mølbak et al. 1994; Karvonen et al. 2000), leading to the λ_S-value of 15, Figure 1. Hence T1D is 15 times more common in siblings of T1D patients than in the general population.

When several genetic factors contribute to a disease, it can be estimated how much of the λ_S-value each accounts for, assuming a multiplicative model of inheritance. In T1D concordance at the HLA locus explains about half of the λ_S-value of 15, Figure 1.

No environmental factor has definitively been demonstrated responsible for triggering or modifying T1D pathogenesis, despite much research. Viral infections (e.g. coxackievirus, cytomegalovirus, rubella-virus), toxins, vaccine administration, early cow's milk exposure and climatic influences are some of the factors which have been studied. Prospective studies of infants developing anti-islet antibodies or diabetes have, however, not consistently been able to demonstrate effects of early cow's milk exposure, breast feeding, enteroviral infections or timing of vaccination, reviewed in (Atkinson and Eisenbarth 2001; Knip et al. 2005).

Studies of interactions between environmental factors and genes aim at describing how they jointly influence the risk of developing human disease. If genetic and environmental factors are only considered separately, estimates of the proportion of the disease explained by genes, environment and their interaction may be incorrect. Intuitively, most likely to be of importance are interactions among gene products and environmental exposures that have a plausible biological explanation, e.g. involved in the same biological pathway. While advance knowledge regarding biological impact of interactions is most often limited, the field is likely to be explored much more in the future.

In summary, like most common genetic disorders T1D is characterized by complex inheritance and cannot be explained by a simple model of dominant, recessive or intermediate inheritance of a specific set of genes. Genes provide susceptibility or protection and the disease may be a result of variations in several genes, with the majority only contributing weak effects. Due to the complex interplay between genetic and environmental factors, complex polygenetic diseases are also termed multi-factorial.

3. DISSECTION OF THE GENETICS OF MULTI-FACTORIAL DISEASES

Mapping a genetic disease requires identification of the genetic variability contributing to the disease. Heritability, linkage, association, and twin studies of many common disorders have suggested multifactorial contribution. In conditions involving single gene defects mutations are considered causative. In multi-factorial diseases the genetic contributions are generally considered to be susceptibility loci, each of which may contribute small effects to disease susceptibility in an additive or multiplicative manner and in the context of environmental and perhaps epigenetic factors. Human sequence variations can influence disease risk, e.g. by conferring changes in the encoded protein and thereby directly affecting the phenotype (premature stop codons, protein truncation mutants etc.), as well as by more subtle effects e.g. by altering mRNA expression levels or stability, in vivo regulation or protein interactions. Relating a genotype to a measurable phenotype therefore is a challenging task. It has been successful in monogenic disorders, in which mutation of a single gene is both necessary and sufficient to cause disease in any given individual. Most common disorders, however, are complex traits. No single gene segregates tightly with disease either individually or in families, and it has proven difficult to localize putative disease genes to chromosomal locations. Even though, most common diseases are strongly influenced by genetics, only few genes have been identified that are partly responsible for the familial clustering of these diseases. Whereas Mendelian diseases are often characterized by being rare but with high penetrance, complex diseases are relatively common and influenced by gene variants with low and incomplete penetrance at the individual level, some however may have substantial impact within specific populations (Lander and Schork 1994). The latter phenomenon is described by the population attributable risk, an epidemiological measure that expresses the fraction by which the disease would be reduced if the risk factor was not present in the population (Benichou 2002). The population attributable risk is important from a public health perspective but does not tell anything about the risk of the individual. It is determined by the strength of the association and the prevalence of the marker in the general population. The population attributable risk is high in monogenic rare disorders, e.g. 50% in cystic fibrosis, but low for rare alleles in complex diseases. If the disease associated allele is common, the population attributable risk increases, e.g. the insulin gene (*INS*) VNTR, for which the frequency in the population of the class I alleles (see chapter 4) is 70% (in T1D ~80%) and the population attributable risk 43% (Bell et al. 1984; Barratt et al. 2004).

The human genome comprise 3.2 billion base pairs and it has been estimated that it contains approximately 12 million single nucleotide polymorphisms (SNPs) (http://www.ncbi.nlm.nih.gov/SNP/index.html) corresponding to one SNP every 300 base pairs. Common variants with minor allele frequencies above 10% have been estimated to occur every ~600 bp (Kruglyak and Nickerson 2001), whereas one in every 1,000-2,000 nucleotides of coding sequence are supposed to be variable (Przeworski et al. 2000; Kruglyak and Nickerson 2001; Sachidanandam et al. 2001; Tsunoda et al. 2004). Polymorphisms refer to variants with an allele frequency of at least 1% in a given population (Kruglyak and Nickerson 2001). However, apart from single base changes, which are the main source of genetic and phenotypic human variation, also different structural variations are present in the genome. These include deletions, insertions, duplications, translocations and copy number variants which also contribute to disease susceptibility and human diversity (Feuk et al. 2006).

4. TYPE 1 DIABETES GENETICS

In T1D a few established candidate genes exist, i.e. the human leukocyte antigen (HLA) class II genes on chromosome 6p21 (Nerup et al. 1974; Cucca et al. 2001), which exerts a major effect, as well as the Insulin (*INS*) gene on chromosome 11p15 (Bell et al. 1984; Barratt et al. 2004), the *CTLA4* gene on chromosome 2q33 (Nistico et al. 1996; Ueda et al. 2003) and the *PTPN22* gene on chromosome 1p13 (Bottini 2004; Smyth et al. 2004), all conferring modest effects. Additionally, a few recently suggested genes seems promising with disease association replicated in at least two cohorts, i.e. the *IL2RA/CD25* on chromosome 10p15.1 (Vella et al. 2005), the *IFIH1* gene on chromosome 2q24.3 (Smyth et al. 2006a) and the very recently suggested *ITPR3* gene on chromosome 6pter-p21 (Roach et al. 2006). These genes were identified by different means, discussed later.

The most important genes are located within the major histocompatibility complex (MHC) HLA class II region on chromosome 6p21, a locus termed *IDDM1*, which alone accounts for approximately 40-45% of the genetic susceptibility to T1D (Rich 1990; Pociot and McDermott 2002). Association between autoimmune diseases and alleles of genes in the MHC region is among the most consistent findings in human genetics. Genetic, functional, structural and animal model studies all support that the HLA genes are the major genetic components in T1D susceptibility. The association between HLA and susceptibility to T1D was made in the early 1970's (Nerup et al. 1974) and has been consistently reproduced since then. Genes in the region are involved in the immune response e.g. by presentation of antigenic peptides to T lymphocytes. But even though the function of the proteins encoded by these genes is well known, their specific contribution to the pathogenesis and the exact mechanisms by which the locus confers susceptibility to immune-mediated destruction of the pancreatic islets is still not clear (Atkinson and Eisenbarth 2001; Pociot and McDermott 2002). Early studies indicated that HLA class II genes (*DRB1* and *-DQB1*) were the most important genes in the region (Thomsen et al. 1975), and

it has consistently been demonstrated that they are the primary determinants of *IDDM1* (Cucca and Todd 1996; She 1996; Thorsby 1997; Dorman and Bunker 2000; Ronningen et al. 2001; Undlien et al. 2001; Koeleman et al. 2004; Lambert et al. 2004). However, due to the strong linkage disequilibrium (LD) between these loci it has been very difficult to elucidate the effect of individual *HLA-DQ* or *-DR* genes. The contribution of the *IDDM1* region is easily detectable in genome-wide linkage analysis, as indicated by a LOD score of 116 in a recent combined T1D genome scan (Concannon et al. 2005). The influence of this region on genetic susceptibility to T1D is complex, with epistasis between *DQB1* and *DRB1*, as demonstrated by disease association of particular *DQB1-DRB1* haplotypes, as well as trans- or genotype effects involving *DQA1, DQB1* and *DRB1* as well as probably yet unrecognized genes that modify class II risk. Susceptibility effects for HLA class II haplotypes range across a 200-fold risk gradient (Pociot and McDermott 2002; Lambert et al. 2004). There is evidence that the degree of risk conferred by different combinations/haplotypes of class II alleles is determined by the predicted structure and function of peptide-binding pockets of the *DRB1* molecule (Cucca et al. 2001). The peptide-binding ability of the class II molecule is dependent on certain amino acids of the *HLA-DQB1* and *DRB1* chains (Chao et al. 1999; Latek et al. 2000; Stratmann et al. 2000; Cucca et al. 2001; Lee et al. 2001b; Siebold et al. 2004).

Although the classical HLA genes represent good candidates given their immunological roles, LD surrounding these genes has made it difficult to rule out effects from neighbouring genes, many with immune function, in influencing disease susceptibility. A role for MHC complex genes other than class II genes has previously been studied in *HLA-DR3/4* heterozygous individuals (Thomsen et al. 1988; Pociot et al. 1991), as well as by studying *HLA-DPB1* alleles (Erlich et al. 1996; Lie et al. 1997; Noble et al. 2000; Cucca et al. 2001; Valdes et al. 2001; Cruz et al. 2004), the antigen-processing genes: *TAP1, TAP2, LMP2*, and *LMP7* (van Endert et al. 1994; Caillat-Zucman et al. 1995; Undlien et al. 1997), as well as the tumor necrosis factor and lymphotoxin genes (*TNF* and *LT*), which showed some evidence for association independently of *DRB1* and *DQB1* (Pociot et al. 1991; Pociot et al. 1993; Bidwell et al. 1999; Bidwell et al. 2001). Furthermore, another class of MHC genes, MHC Class I chain-related (MIC) genes have been identified. The *MICA* genes is located between the *TNFA* and the *HLA-B* genes (MHC Sequencing Consortium 1999), and was demonstrated to be independently associated with T1D in several populations (Gambelunghe et al. 2000; Bilbao et al. 2002; Sanjeevi et al. 2002; Gupta et al. 2003; Tica et al. 2003). Despite intensive efforts in the analysis of these and the classical HLA genes no causal variants have yet been identified in T1D. Studies of classical HLA genes have demonstrated that probably an extended haplotype, rather than a single variant, is associated with the disease. This suggests that one should consider all genes of the MHC region rather than focusing only on the classical HLA genes.

As for the HLA region, the importance of the insulin gene, *INS*, was originally suggested by association studies (Bell et al. 1984; Barratt et al. 2004). *IDDM2* refers to a region on chromosome 11p15.5, which is polymorphic and maps to a variable number of tandem mini-satellite repeats (VNTR). The VNTR is flanking (5') the insulin gene, which is expressed specifically in the β-cell and thymus. The *INS* VNTR most probably represents the primary locus for *IDDM2*, and antibodies against insulin are early detectable auto antibodies in T1D (Bell et al. 1984; Barratt et al. 2004). The *IDDM2* locus has an odds ratio for T1D of approximately 3, whereas the locus specific estimated sibling genetic risk ratio λ_S is only 1.16. The VNTR ranges from 26 to more than 200 repeats of a 14-15 bp DNA sequence divided into three discrete classes according to size. The class I/I homozygous genotype (26-63 repeats) seems to increase the risk of developing T1D 2-4-fold, whereas class III alleles (141-209 repeats) are associated with a dominant protection from T1D. Class II alleles (~80 repeats) are virtually absent in Caucasoid populations (Bell et al. 1984; Barratt et al. 2004). The population attributable risk for the *INS* VNTR in T1D is 43%. The class I alleles of the *INS VNTR* which confers genetic risk to T1D lead to lower insulin mRNA expression in the thymus as well as higher insulin mRNA expression in the β-cell compared to the dominant protective class III alleles. This may attenuate the development of central tolerance to insulin, at the same time as providing high antigen expression in the β-cell (Pugliese et al. 1997). Furthermore there is evidence for interaction between the *INS* and *HLA* loci in conferring susceptibility to T1D (Nerup et al. 2001).

Another confirmed locus conferring T1D risk is the cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*) gene (*IDDM12*) on chromosome 2q33, which is involved in modulating immune responsiveness. Variation within the *CTLA4* gene has been shown to associate with T1D, as well as many other autoimmune diseases (Nistico et al. 1996; Kristiansen et al. 2000a; Pociot 2002; Ueda et al. 2003; Pociot 2004). The encoded molecule is a co-stimulatory receptor conferring an inhibitory effect on T-cell activation. Several different *CTLA4* gene variants have been identified and demonstrated associated with autoimmune diseases as T1D, systemic lupus erythematosus, celiac disease, Graves disease and autoimmune hypothyroid disease, and the gene may be a common susceptibility factor in autoimmunity in general (Pociot 2004). The most comprehensive SNP and LD mapping analysis of this locus (Ueda et al. 2003) identified a 3'untranslated region (UTR) SNP (*CT60* or *G6230A*, rs3087243) as the predominant marker for Graves disease and autoimmune hypothyroidism, as well as for the risk for T1D. The effect in T1D was, however, weaker than in the other diseases, the population attributable risk for the *CT60* SNP in T1D is 7%. However, the modest λ_S value (of 1.01) predicted for the associated SNPs at the *CTLA4* locus are unlikely to account fully for the magnitude of the observed evidence for linkage to T1D of the region (regional, locus-specific effect of $\lambda_S = 1.19$) (Concannon et al. 2005). Based on the functional data on *CTLA4* observed in (Ueda et al. 2003) and other studies (Pociot 2004) no clear molecular model to explain the increased risk for autoimmunity has yet emerged. Further studies are needed to clarify the functional role of *CTLA4* in T1D pathogenesis.

Finally a functional variant of the *PTPN22* gene (encoding the lymphoid-specific phosphatase, protein tyrosine phosphatase non-receptor-type 22, LYP) on chromosome 1p13 has been reported to be significantly associated with T1D as well as with other autoimmune diseases (Bottini 2004; Smyth et al. 2004). LYP, encoded by the *PTPN22* gene, is an inhibitor of T-cell activation, acting by dephosphorylating T-cell receptor-associated kinases. *PTPN22* is thereby a candidate gene for all autoimmune diseases mediated by T-cells. A SNP (rs2476601) in *PTPN22*, which results in an amino acid substitution, has been shown to have functional consequences for kinase binding involved in signalling via T-cell receptors. Knockout mice deficient for *PTPN22* show selective dysregulation in the effector/memory T-cell system (Hasegawa et al. 2004). Strong association of this coding variant of the *PTPN22* gene to T1D was identified originally in case-control designs in two ethnically different populations (Bottini 2004). Replication of the association to T1D was recently published in two reports; a large T1D family as well as a large case-control cohort (Smyth et al. 2004), and in a second report association was confirmed in a large cohort of US T1D multiplex families (Onengut-Gumuscu et al. 2004). The population attributable risk for the associated SNP is 8.4% but varies considerably between populations. Taken together these reports support the *PTPN22* gene to play a role in T1D, a finding recently further supported by our own studies (Chelala et al. 2007). In this collaborative study we additionally demonstrated evidence for association of the previously strongest associated SNP with the presence of GAD autoantibodies, which was restricted to patients with more than 10 years of disease duration. This may help define a subgroup of patients with long-term persistence of GAD autoantibodies (Chelala et al. 2007). Familial clustering of several autoimmune diseases is well

documented and *PTPN22* may be a link in a shared etiology or common factors contributing to autoimmune diseases. Recent reports have demonstrated association of the same SNP in *PTPN22* to rheumatoid arthritis in two independent samples (Begovich et al. 2004), to human systemic lupus erythematosus (Kyogoku et al. 2004), Graves' disease (Smyth et al. 2004; Velaga et al. 2004), as well as Addison's disease (Velaga et al. 2004). The *PTPN22* gene maps to chromosome 1p13, a region which has been weakly linked to rheumatoid arthritis (Jawaheer et al. 2003), as well as to systemic lupus erythematosus (Gaffney et al. 1998). Regarding T1D linkage in the recent largest genome scan the chromosome 1p11-p12 region was only weakly supported ($p = 1.4\times10^{-2}$); however, *PTPN22* was contained within the LOD-1 support interval (Concannon et al. 2005). Despite an allelic odds ratio of approximately 1.7 for *PTPN22*, the λ_S for this coding variant is only 1.05, probably explaining the difficulties in identifying this region in genome scans and the large materials needed to convincingly confirm the association. It has been estimated that to detect linkage with a p-value <0.001 with 50% power, approximately 8,000 affected sib-pairs would be needed for a fully informative genetic map (Concannon et al. 2005). It was furthermore estimated that under a multiplicative model, the contribution of *PTPN22* to T1D, based on the odds ratio, was approximately 2%, as compared to 40-50% for the HLA (Concannon et al. 2005).

The *IL2RA* (*CD25*) gene on chromosome 10p15.1 was recently evaluated due to its status as a biological candidate gene (Vella et al. 2005). The gene encodes the IL-2Rα subunit of the IL-2 receptor (*CD25*) complex. The expression of *CD25* on regulatory T cells is essential for their function in suppressing T cell immune responses and thereby autoimmune diseases (Salomon et al. 2000; Malek and Bayer 2004; Viglietta et al. 2004). Furthermore, a rare mutation in the gene has been shown to lead to severe human immunodeficiency (Sharfe et al. 1997). The gene was evaluated for association to T1D by use of a LD approach, in which haplotype tagged SNPs (Johnson et al. 2001), see chapter 5, in the gene were identified after re-sequencing exons and regulatory regions (Chapman et al. 2003; Clayton et al. 2004). In a multi-locus test evidence for association of the gene with T1D was demonstrated for a large case-control cohort, as well as an independent family collection (Vella et al. 2005). Since this haplotype tagging approach based on LD, and the multi-locus test applied, use the combined information from all the haplotype tagged SNPs (Vella et al. 2005), future fine mapping of the region is needed to identify the causal variant. A potential functional significance of such a variant is unknown.

In an interim analysis of a genome-wide association study with only non-synonymous SNPs, the interferon-induced helicase (*IFIH1*) gene on chromosome 2q24.3, was recently identified (Smyth et al. 2006a). The analysis included 6,500 non-synonymous SNPs, and approximately 2000 cases and 2000 controls. According to power calculations, it was estimated that the study would be able to detect only the strongest loci, with odds ratios >1.7, and that it may contain many false-positive findings (Smyth et al. 2006a). The method using non-synonymous SNPs is discussed in chapter 5.5. The third most associated SNP in the study, was a SNP in the *IFIH1* gene (odds ratio for the minor allele 0.82, $p=4.7\times10^{-5}$). Number one SNP was the now established coding SNP in *PTPN22*, mentioned above, number two was a SNP in the CAPSL gene at chromosome 5 (Smyth et al. 2006a). The *IFIH1* SNP was genotyped in additional cases and controls and the finding was statistically supported in a family collection of more than 2000 T1D families (Smyth et al. 2006a). Furthermore, additional SNPs in the region were genotyped, but no SNP has yet been identified as the causal (Smyth et al. 2006a). The *IFIH1* gene encodes a receptor for double stranded RNA. It is an early type 1 interferon-β responsive gene, which may contribute to apoptosis of virally infected cells in anti-viral immune responses (Smyth et al. 2006a), also of putative importance in T1D pathogenesis (Knip et al. 2005).

Very recently, a new T1D candidate gene was proposed, the Inositol 1,4,5-Triphosphate Receptor 3 (*ITPR3*) gene on chromosome 6qter-q21 (Roach et al. 2006). The gene was identified in an attempt to fine map the MHC region, for which the causal gene(s), as described in chapter 4, are still not known. The HLA DR/DQ locus has consistently been demonstrated associated and linked to T1D (Concannon et al. 2005). The extensive LD has, however, made it difficult to prove other associations in or close to the MHC which are independent and not explained by LD with HLA DR/DQ. The singular but broad linkage peak for the MHC region has been speculated to contain additional gene(s) and contain several local maxima (Herr et al. 2000). In the non-obese diabetic (NOD) mouse, a locus termed *Idd16*, which predisposes to T1D distinctly from MHC class II association, is in humans syntenic to a region immediately centromeric to the MHC locus (Deruytter et al. 2004). Roach et al. (Roach et al. 2006) used a high-density SNP panel, with markers spaced closer than in previous studies. A MHC SNP set from Illumina, containing 2,260 SNPs, covering the MHC region and the region centromeric to this, was used. The higher resolution, as compared to previous studies should be advantageous in distinguishing putative separate peaks of association. By testing for single-marker associations two peaks were detected; one confirming the HLA-DR/DQ locus and the other containing the *ITPR3* gene, centromeric to and outside the classical MHC region. The *ITPR3* gene is not localized in a region of extended LD and two-locus regression analysis supported a MHC class II independent influence on T1D (Roach et al. 2006). The most significant SNP was located in an intron of the *ITPR3* gene and the association was demonstrated in two Swedish case-control cohorts, however, of modest size, i.e. 181 cases and 182 controls and 462 cases and 299 controls, respectively. The combined p-value was 1.3×10^{-6} (recessive odds ratio = 2.5 (95% CI: 1.7-3.9) and the estimated population-attributable risk was 21.6% (95% CI: 10-31%). The *ITPR3* gene is expressed in beta cells, were it is upregulated in response to glucose stimulation (Lee et al. 2001a). *ITPR3* is furthermore known to mediate second messenger signaling by releasing Ca^{++} from intracellular stores in response to inositol triphosphate (Roach et al. 2006). This is a promising gene, but further fine mapping, in order to define the causal variant in *ITPR3* or nearby, as well as replication in other populations and in larger collections are needed to confirm the observation, and its apparent independence of HLA conferred effects. In our own studies of protein-protein interaction networks build from observed genetic interactions (Bergholdt et al. 2007b) we identified the *ITPR3* gene as central, and together with three other genes responsible for interactions between the HLA region and other genetic regions. Since these networks were of high statistical significance this further support the putative importance of the *ITPR3* gene in T1D.

Genome-wide scans have been intensively used in the search for genetic determinants for T1D. The first scans for linkage to T1D, using fewer than 100 affected sib-pair families, identified chromosome 6p21 (*IDDM1*) as the major T1D risk locus (Davies et al. 1994; Hashimoto et al. 1994), as well as other minor loci. Subsequent studies also identified putative T1D loci on several other chromosomes, as well as the *IDDM1* region (Concannon et al. 1998; Mein et al. 1998; Cox et al. 2001; Nerup et al. 2001). The originally identified putative loci, conferring susceptibility to T1D were termed *IDDM1-IDDM18*, approved by the Human Gene Nomenclature Committee (www.gene.ucl.ac.uk/nomenclature) (*IDDM 9* and *IDDM14* were reserved but never published). *IDDM1* and *IDDM2* (the HLA and *INS* components) were both originally identified by a candidate gene approach based on case-control studies. The remaining *IDDM* loci, except *IDDM17*, have all been discovered by earlier linkage studies using affected sib-pair families either in whole or partial genome scans. However, despite the fact that there was strong statistical evidence supporting linkage for some of these regions in the initial reports, most regions have not been clearly established in multiple populations and quite little overlap were observed between the

different scans, reviewed in (Pociot and McDermott 2002). Recently a joint analysis of data from previous T1D genome-wide scans (Cox et al. 2001; Nerup et al. 2001) was performed (Concannon et al. 2005). The Type 1 Diabetes Genetics Consortium (T1DGC) (http://www.t1dgc.org) merged data from the previous large genome scans and added new data from 254 families (Concannon et al. 2005). This family collection containing 1435 affected sib-pair families provided ~95% power to detect a locus with a locus-specific $\lambda_S > 1.3$ and $p=10^{-4}$, and represents one of the largest genome scans ever performed in a multi-factorial disease. Some original *IDDM* loci were confirmed while other previously suggested loci were excluded. Strong evidence for the HLA region (*IDDM1*) was observed (LOD score 116, and a locus specific λ_S of 3.34), whereas just moderate support (LOD scores: 1.66-3.34 and λ_S: 1.10-1.19) for T1D linkage was obtained for non-chromosome 6 loci. Nine regions were identified that supported non-HLA-linked susceptibility (Concannon et al. 2005), among these the insulin gene region, the *CTLA4* region, a region on chromosome 10 and two regions on chromosome 16, Table 1.

This study is an example of a concerted action (of the T1DGC) to increase sample size and thereby provide power to the levels necessary for convincing identification of T1D loci of also modest effects (Concannon et al. 2005). A major barrier to T1D gene identification, given the likely small locus-specific contribution (low λ_S) for non-HLA genes, probably has been the limited number of available affected sib-pair families with T1D. The increased sample size of this study allowed the exclusion of more than 80% of the human genome for locus-specific effects of $\lambda_S \geq 1.3$, excluding that a major T1D locus or gene exist in these regions.

However, loci of modest or minor effects on T1D susceptibility are not excluded, as evidenced by locus specific λ_S values below 1.19 for all loci apart from HLA in the genome scan (Concannon et al. 2005). To determine locus specific effects of this magnitude (λ_S values of 1.05-1.30), it has been estimated that more than 4,300 affected sib-pairs will be needed to demonstrate such effects, with a significance level corresponding to that determined by Lander and Kruglyak (Lander and Kruglyak 1995) for reporting significant linkage, $p < 2 \times 10^{-5}$. Collection of this amount of affected sib-pairs is one of the goals for the T1DGC.

The established candidate genes in T1D (HLA class II genes, *INS*, *CTLA4* and *PTPN22*) were all originally identified as biologically plausible candidate genes demonstrating association to T1D. Until now, no specific gene conferring risk to T1D has been identified based on linkage analyses. However in other complex diseases, e.g. Crohns disease (*NOD2/CARD15*), asthma (*ADAM33*), schizophrenia (*NRG1*) and myocardial infarction (*ALOX5AP*), this has only quite recently been successful (Hugot et al. 2001; Ogura et al. 2001; Stefansson et al. 2002; Van-Eerdewegh et al. 2002; Helgadottir et al. 2004).

Interestingly, identification of some recently suggested T1D candidate genes were by different means. The *IFIH1* gene was suggested based on a genome-wide association approach, although using a limited number of non-synonymous SNPs and resulting from an interim analysis (Smyth et al. 2006a). The *ITPR3* gene was identified based on a SNP based fine mapping of the MHC region (Roach et al. 2006), and may, if replicated, also in other populations thereby be shown to be the first T1D gene identified based on fine mapping of a linkage region. The *IL2RA/CD25* gene was suggested based on its role as a biological candidate gene (Vella et al. 2005). These latter findings needs further replication, but hold promise for future identification of T1D genes by different means, the importance of using large sample sizes and the increased use of SNPs in gene mapping are underlined.

5. IDENTIFICATION OF SUSCEPTIBILITY GENES

The sequence of the human genome has been completed, and millions of SNPs have been deposited in public databases. In the most commonly used database of SNPs, dbSNP (www.ncbi.nlm.nih.gov/SNP) at present more than 12 million SNPs are registered, of which approximately six million have been validated. For a complete understanding of the genetic variation underlying all common diseases a complete re-sequencing of the entire genome in many affected as well as un-affected individuals for each disease would be preferable, however not technically or economically feasible despite tremendous improvement in re-sequencing and genotyping technologies. Several other approaches exist for attempting to map susceptibility genes underlying common diseases. The two major classical approaches are candidate gene studies and genome-wide studies of randomly selected markers (microsatellites or SNPs) (see below). Common disease susceptibility alleles may not be considered true disease genes, because although necessary contributors, they are not sufficient to cause disease. These alleles may be more common in unaffected than in affected individuals in the general population. However, at the molecular level their products can be thought of as molecular variants of biochemical pathways and as components in complex multi-component networks that contribute in additive ways to the disease phenotype. Individually, they may have little or no disease effect but may lead to disease through combinatorial interaction and in different environmental settings. The so-called common variant/multiple disease hypothesis states that many disease genes may not be disease specific. Common deleterious alleles, found at a relatively high frequency in the population may play a role in related clinical phenotypes in the context of different genetic backgrounds and under different environmental conditions (Becker 2004). Another hypothesis often referred to is that susceptibility to T1D is conferred by common alleles of normal genes recurring in unfavorable combinations (Nerup et al. 1994; Pociot 1996). In line with this is the common variant/common disease hypothesis stating that the genetic risk for common diseases will often be due to disease-producing alleles found at relatively high frequency (>1%) (Becker 2004). An underlying assumption is that the deleterious effect of each disease producing allele is, on average, relatively low which might explain why these alleles are maintained in the population at high frequencies (common alleles of normal genes). However, it remains to be resolved if most common genetic factors are specific for diseases or may lead to multiple related diseases, or

Table 1. Chromosomal regions identified in Concannon et al. (Concannon et al. 2005). Modified from (Concannon et al. 2005). The chromosomal region, as well as the top scoring marker of each region is provided. Its position in cM, its LOD score, corresponding λs value and the LOD-1 interval surrounding each marker, are indicated. Additionally empiric and genome-wide P-values are provided.

Chromosomal region	Marker	Position (cM)	LOD score	λs	LOD-1 interval	Empiric p-value	Genome-wide emp. p-value < 0.05
2q31-q33	*D2S2167*	192	3.34	1.19	177-204	9.0×10^{-5}	Yes
3p13-p14	*D3S1261*	98	1.52	1.15	78-112	8.2×10^{-3}	
6p21	*TNFA*	47	116.3	3.35	46-48	ND (10^{-52})	Yes
9q33-q34	*D9S260*	150	2.20	2.20	138-161	1.5×10^{-3}	
10p14-q11	*D10S1426*	61	3.21	3.21	52-66	1.2×10^{-4}	Yes
11p15	*D11S922*	2	1.87	1.87	0-14	3.4×10^{-3}	
12q14-q12	*D12S375*	81	1.66	1.66	77-83	5.8×10^{-3}	
16p12-q11.1	*D16S3131*	56	1.88	1.88	26-71	3.3×10^{-3}	
16q22-q24	*D16S504*	108	2.64	2.64	100-121	4.9×10^{-4}	Yes
19p13.3-p13.2	*INSR*	25	1.92	1.92	0-43	3.0×10^{-4}	

whether the effects of disease alleles are different when they are in other genotypic backgrounds, in other genetic combinations or are influenced by other epigenetic or environmental factors.

5.1. LINKAGE DISEQUILIBRIUM

LD is the phenomenon describing that two or more loci are inherited together more often than would be expected by chance (Ardlie et al. 2002a). This inheritance of chromosomal regions from shared ancestors without recombination, known as haplotypes, maintains particular combinations of alleles in the population (Reich et al. 2002). LD which is established in a population at some point in time is, however, gradually diminished by the actions of recombination and novel random mutations (Kruglyak 1999; Reich et al. 2002). A region characterized by no (or little) recombination is collected in a haplotype block, whereas regions with high recombination usually are described as divided into separate haplotype blocks. LD thus describes the non-random correlation between alleles at a pair of SNPs, usually defined by D' or r^2 values. An international joint effort to create a genome-wide map of LD and haplotype blocks has been established through the HapMap project (www.hapmap.org). The hope is that by knowing the haplotype block structure of the entire genome, one could capture the genetic variability of the genome by genotyping a much smaller number of SNPs that describe each haplotype block (Johnson et al. 2001). These haplotype tagging SNPs (htSNPs) are defined by the fact that information of their allelic variation provides relatively unambiguous information about variation at another locus (Johnson et al. 2001). Haplotypes thereby allow mapping of disease-susceptibility alleles without the need to identify and test every SNP across each chromosomal region (Kruglyak 1999; Risch 2000; Johnson et al. 2001) and many studies of genetic variants and association with a disease are based on the concept of LD. It may be assumed that a variant allele itself is partly responsible for the phenotype or it may be assumed that it is a marker allele in tight LD with the disease allele. An association study in the latter case is known as an indirect association approach and is the one used in whole-genome association mapping (Kruglyak 1999), Figure 2. In other words, LD can be determined between individual SNPs, e.g. for constructing haplotype maps across a genomic region, as well as between an examined SNP and another SNP that is the one responsible for the phenotype causing mechanism, in which case the examined SNP is regarded as a phenotype marker (Kruglyak 1999).

Association studies make use of both categories. In the first case the primary goal is to reduce the amount of genotyping needed in order to cover a particular region so that almost complete information about variation in the region is obtained (Johnson et al. 2001). One study suggested that LD might exist only within distances of 3 kb (Kruglyak 1999), whereas several others have detected LD/significant association within 50 kb, 1 Mb or even much much further, of a susceptibility locus (Collins et al. 1999; Moffatt et al. 2000; Abecasis et al. 2001; Reich et al. 2001) An allele may show strong LD with the functional variant in one population but exhibit weak LD in other populations (de Bakker et al. 2006), or may show strong LD with one marker but exhibit weak LD with other nearby markers, even in the same population (Carlson et al. 2004). Such phenomena may make it difficult to replicate genetic associations among populations and to obtain consistent results within a genomic region in the same population. A way to improve the power of association studies and increase the probability of replications, may be to construct haplotype blocks by studying LD patterns across the genome and then optimally select a set of robust htSNPs such that all common variants are either directly genotyped or in strong LD with the genotyped tagSNPs (Goldstein 2001; Johnson et al. 2001; Stephens et al. 2001; Gabriel et al. 2002; Zhang et al. 2002; Ke and Cardon 2003; Xiong et al. 2003; Carlson et al. 2004). However, it is unclear whether haplotype block patterns and tagSNPs are consistent between populations and among repeated sampling from within a population. Recently, several large genomic regions (of ~500 kb) have been comprehensively examined as part of the ENCODE project (http://www.genome.gov/ENCODE/) (ENCODE Project Consortium 2004), which involved re-sequencing of 96 chromosomes in the selected regions to ascertain all common variants, as well as genotyping of all SNPs that are either in dbSNP or were identified by the re-sequencing. This strongly confirmed the patterns of long segments of strong LD that were seen in earlier studies (Collins et al. 1999; Moffatt et al. 2000; Abecasis et al. 2001; Reich et al. 2001).

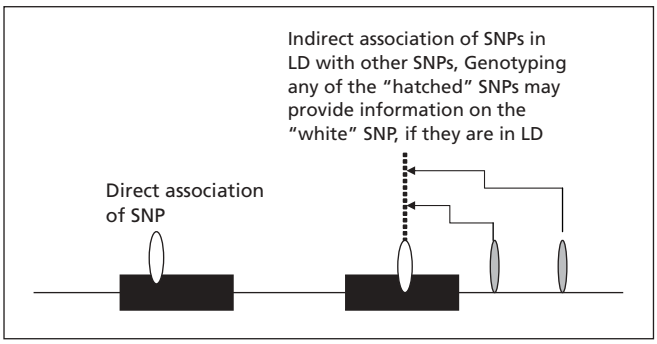

Figure 2. Direct and indirect association of a SNP. The two situations are illustrated. Left: Direct association of a SNP with a disease, e.g. a SNP located in the disease-causing gene. However other SNPs in LD with a true causal SNP may demonstrate association as well, termed indirect association (right).

It has been shown that also copy number polymorphisms are an important class of genetic variation in the human genome (Iafrate et al. 2004; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005). Special emphasis has recently been on deletion variants and methods to detect such polymorphisms (Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006). It was demonstrated that most common deletions (minor allele frequency >10%) could be assayed by neighboring SNPs, due to high degree of LD with nearby SNPs (Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006). Deletions causing severe diseases are rare, whereas more benign deletions are widespread (Iafrate et al. 2004; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005). Methods have recently been developed by which deletions can be evaluated in SNP-based whole genome association studies; they are not directly observed in ordinary gene mapping studies. The underlying principle is that segregating deletions will leave a footprint in SNP genotype data, seen by i.e. mendelian inconsistencies, Hardy-Weinberg disequilibrium or null genotypes (Bellanne-Chantelot et al. 2005; Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006). A recent study, which re-sequenced the MHC haplotypes, e.g. different DR3 haplotypes, demonstrated that copy number variants were quite common and accounted for 13% of the variants identified in the region (Traherne et al. 2006). Copy number variants have not yet been evaluated for association to diseases, but offer a potentially interesting and additional tool. Further studies are needed to clarify their role and usefulness as markers of disease.

For mapping genes in complex diseases different approaches can be used. The principal distinction lies between candidate gene approaches, in which biological information plays the primary role in selecting genes for detailed analysis, and positional cloning strategies, in which the major effort is in localizing susceptibility variants through examination of observed or inferred patterns of co-segregation between the disease phenotype and marker genotypes within families or populations.

5.2. CANDIDATE GENE APPROACHES

The traditional candidate gene approach has a focused strategy, it is a classical hypothesis-driven approach and usually focuses on one or a few genes, proteins or biological pathways of high information

content and pathophysiological importance. Candidate gene studies have until recently been the only practical alternative to classical linkage analysis. Genes can either be functionally/biologically selected candidates or positional candidates, selected by e.g. their location in a region of linkage. In both cases studies are hypothesis-based, relying on some kind of evidence or prior knowledge that the chosen genes might affect risk of disease. Additionally, the use of new tools of genomics (e.g. transcriptional profiling and proteomics) from human or animal studies may also identify candidate genes. Candidate gene approaches are based on testing for association, either of known allelic variants in selected genes or of newly identified variants after re-sequencing parts of, or the entire selected gene(s). In its simplest form allele or genotype frequencies are compared between cases and controls, alternatively family-based controls can by used, with the advantage of avoiding a potential problem of population stratification in the control samples. Family-based designs have the advantages that they are robust against population admixture and stratification, and allow both linkage and association to be tested for (Laird and Lange 2006), but do require more genotyping. For family-based studies, the TDT test (Spielman et al. 1993) can be used, which implies that under the null hypothesis, Mendel's laws determine which marker alleles are transmitted to the affected offspring from heterozygous parents. The test compares the observed number of transmitted alleles with what was expected for Mendelian transmission. An excess of alleles of one type among affected offspring indicates that a disease-susceptibility locus for a trait of interest is linked and associated with this marker locus. Originally TDT was used to test for linkage between a disease and a genetic marker in the presence of (already detected) disease association (LD) (Spielman et al. 1993; Spielman and Ewens 1996; Laird and Lange 2006). Since both linkage and association between the trait and the marker have to be present for the TDT to reject the null-hypothesis (Spielman et al. 1993; Spielman and Ewens 1996; Laird and Lange 2006), the TDT is now typically used as a test for association in family materials (Hirschhorn and Daly 2005). Association is due to lack of recombination between the marker locus and the disease-susceptibility locus. Even in the absence of prior evidence for association, TDT has been shown to be valid and can be used to test a marker, or a set of markers, for association to disease, using a data set comprising parents and one or more affected sibling(s) (Spielman and Ewens 1996). TDT is a non-parametric test, and does not require specification of disease model, or assumptions about the distribution of the disease in the population, to be valid. An extension to the classical TDT test is the Sib-ship-TDT, which also uses information from unaffected siblings (Spielman and Ewens 1998).

Case-control studies involve groups of affected and unaffected subjects all being unrelated and sampled from the same population. This approach is used for identifying statistically significant differences in allele frequency and/or genotype distributions between a genetic variant and the disease. Recruiting probands and their relatives may be more difficult than recruiting cases and controls, except probably for childhood diseases, where suitable controls are difficult to obtain. A disadvantage of the family based approach is that for late-onset diseases it may be difficult to obtain affected subjects with living parents. Furthermore, the requirement for living parents may introduce an age-of-onset bias towards younger patients for diseases that usually arise late in life. Subjects may tend to be characterized by a less severe phenotype, if they have living parents, or may be affected by subtypes of disease, e.g. a more severe phenotype related to early onset of a disease. Both situations could contribute to population stratification bias. However, family designs are generally more robust against population substructures, and use of family-based samples is a way to avoid severe stratification biases, since only within-family comparisons are made and data are thereby not affected by population structure (Spielman and Ewens 1998).

Furthermore family-based approaches have other theoretical advantages. Samples can be used to determine whether an allele has different effects on disease, dependent on whether inherited maternally or paternally. Parent-of-origin effects describes two phenomenas; genomic imprinting, which results from epigenetic modification of the genome leading to unequal transcription of parental alleles, i.e. expression of alleles depend upon the sex of the parent, from which they were inherited (Reik and Walter 2001; Glaser et al. 2006), and secondly, parent-of-origin effects on mutation rate, the preferential occurrence of some new spontaneous mutations in either the father's or mother's germ line, e.g. base substitutions or chromosomal abnormalities (Glaser et al. 2006). Discordant sib designs can control for the effects of shared environment. However, pure sibship-based association studies may be underpowered relative to case-control studies, although differences in statistical power between population and family-based designs are generally small. A study compared use of trios to case-control studies and found that the number of simplex families (trios) required for TDT was virtually equal to the number of cases needed for similar power in case-control studies with an equal number of cases and controls (McGinnis et al. 2002). This implies that less genotyping (one case and one control versus at least parents and one affected child) may be necessary in case-control studies to gain the same power and for many diseases they may even be easier to collect than families (McGinnis et al. 2002). Another design uses a strategy of collecting a much larger number of controls than cases (McGinnis et al. 2002), which might be advantageous if controls are easier to ascertain than cases for some diseases, in order to obtain sufficient power. However, cases and controls should always be genotyped the same way, in the same laboratory and derive from the same ethnically homogeneous source to minimize risk of false positive associations, due to population stratification. Laird et al. showed by simulations, that for rarer diseases (prevalence e.g. 0.1%) use of 200 trios was more powerful than 200 cases and 200 controls, whereas for more common diseases (prevalence e.g. 14%) the opposite was the case, although differences were small (Laird and Lange 2006).

Candidate gene studies often suffer from a significant lack of information concerning the underlying complex regulatory pathways, and therefore are limited by their need to focus on only few already recognized and selected proteins or genes. Studies are dependent on having predicted the correct gene or genes, usually on the basis of biological hypotheses or the location of the candidate within a previously determined region of linkage. Even if hypotheses are broad, e.g. involving testing all genes in a certain (known) functional pathway or cascade, the approach is limited by lack of complete information regarding pathways and cascades, and will probably not identify all genetic risk factors for the disease in question. The most important limiting factor is lack of complete knowledge of the pathogenesis for most complex diseases, as well as too limited sample sizes.

Association studies using known and common variants are cheaper and simpler than complete re-sequencing of selected candidate genes. To capture all variation in and around a gene, re-sequencing of the entire gene: exons, introns and regulatory regions (i.e. 5' and 3'UTR and further up- and downstream of the gene) is necessary in ideally a large number of cases and controls. Re-sequencing is, however, expensive and laborious, and in many studies so far, only coding regions of one or a few genes were sequenced. Questioning this approach, is the fact that there are several examples of disease-variants located in introns and regulatory regions (Bergholdt et al. 1995; Bergholdt et al. 2000; Horikawa et al. 2000; Prokunina et al. 2002; Helms et al. 2003; Tokuhiro et al. 2003; Ueda et al. 2003; Bergholdt et al. 2004a). SNPs in coding regions are often implicitly accepted as "the answer" when an association with a missense variant is detected. This, often without the same functional evidence, as required for a SNP in a non-coding region, and often despite the presence of many nearby variants that might be equally or more strongly associated with disease. As an example, the well known codon 17 polymorphism (*Thr17Ala*) in the *CTLA4* gene, which has been shown to be associated with several autoimmune

diseases, turned out to show association, only because it was in strong LD with a regulatory polymorphism in the non-coding region (*CT60, G6230A*, rs3087243 in 3'UTR), which is more strongly associated with disease and therefore more likely to be causal (Ueda et al. 2003). Proper interpretation of results from association studies might be challenging, particularly when considering rare non-coding variants. A recent study observed that re-sequencing high-priority candidate genes, in which severe loss of function variants were known to cause Mendelian disorders of lipid metabolism, demonstrated that some of these genes additionally harbored less severe, relatively rare missense variants, which were associated with high, but not extreme, levels of high-density lipoprotein (Cohen et al. 2004). This highlights the complexity and importance of correctly classifying phenotypes and sub-phenotypes.

Candidate gene association studies have been used for many years and have identified many of the genes known to contribute to susceptibility to common disease. They have been powerful in identifying common variants underlying complex traits like T1D, e.g. HLA (Nerup et al. 1974; Cucca et al. 2001), *INS* (Bell et al. 1984; Barratt et al. 2004), *CTLA4* (Nistico et al. 1996; Ueda et al. 2003), *PTPN22* (Bottini 2004; Smyth et al. 2004) and *IL2RA/CD25* (Vella et al. 2005).

Despite that many positive results have come from association studies, often initial positive reports of association have not been replicated. In a study of 166 putative associations (studied at least three times) only six appeared to be consistently replicated (Hirschhorn et al. 2002). Possible reasons for this lack of replication include differences in ethnicity (i.e. susceptibility allele frequencies, penetrance levels etc.), lack of statistical power, multiple-hypothesis testing, variability in study designs or in phenotype definition, population substructure, different statistical models and other biological or genetic heterogeneities between studies or populations (Hirschhorn et al. 2002; Laird and Lange 2006). Moreover the contribution of publication bias may be considerable (Lohmueller 2003), since initial positive associations are more likely to be published than negative findings. Furthermore, reasons for inconsistencies could also be false positive findings due to type 1 errors or population stratification. Prior odds for any biological candidate to contain variants capable of influencing disease susceptibility is low, given the modest effect sizes anticipated and combined with often small sample sizes assayed. Together with a probable use of too liberal thresholds for declaring significance, it is perhaps not surprising that many false positive associations have been reported. False negative findings could be due to lack of power in replication studies, and of course true differences between study populations or between different environmental or genetic modifiers can exist (Hirschhorn et al. 2002). With genome-wide association studies, involving hundreds of thousands of markers, such issues become even more important (Laird and Lange 2006).

Recently, guidelines have been proposed for more robust association studies in multi-factorial diseases (Editorial 1999; Cardon and Bell 2001; McCarthy et al. 2005). These include replication of significant findings in additional study groups of similar ethnic origin and exploration of data sets from dissimilar ethnicities, incorporation of diverse study designs, functional assessment of the presumed aetiological variants in order to provide biological substantiation of statistical findings, as well as sufficiently large sample sizes. Furthermore, genotyping a haplotype tagged set of variants across a gene may provide a systematically capture of variation in the gene. A strategy of using meta-analyses may also be used to overcome some of the problems.

With the improvements in genotyping technology, focus in gene mapping has shifted towards association studies, using large numbers of SNPs or other markers, genotyped in known linkage regions, candidate gene regions or on a genome-wide basis.

In our own studies we used the classical candidate gene approach in evaluation of the human interleukin-1 type I receptor gene, *IL1RI* (Bergholdt et al. 1995; Bergholdt et al. 2000), the Casitas-B-lineage lymphoma b gene, *CBLB* (Bergholdt et al. 2005c) and the interleukin-12-subunit B gene, *IL12B* gene (Bergholdt et al. 2004a). Methods used were considered appropriate at the time these studies were performed. In studies of the *IL1RI* and *CBLB* gene we performed mutational scanning by Southern blotting and/or sequencing by different methods (Bergholdt et al. 1995; Bergholdt et al. 2000; Bergholdt et al. 2005c). After identification of variants, these were genotyped in relevant cohorts and data were evaluated for association to T1D.

IL1RI was selected due to its obvious nature as a T1D candidate gene, being the membrane-bound receptor for IL-1, and thereby essential in IL-1 signaling, known to be crucial in beta cell death (Nerup et al. 1994; Mandrup-Poulsen 1996). In our studies of the *IL1RI* gene (Bergholdt et al. 1995; Bergholdt et al. 2000), we de novo sequenced the complex promoter region, comprising different alternative promoters, and searched for mutations herein. In the first report one variant was detected, initially by Southern Blotting, after which, the SNP and the region surrounding it was characterized by sequencing ~400 bp (Bergholdt et al. 1995). These methods were state-of-the-art methods for detection of mutations at that time and automated sequencing was a newly developed tool. We demonstrated T1D association of the identified SNP in a case-control material, genotyped by PCR-RFLP, comprising 262 cases and 189 controls (Bergholdt et al. 1995), which is a small material according to standards of today, but considered appropriate at that time. T1D association of this SNP in the *IL1RI* gene was subsequently confirmed in the UK population and in non-high risk HLA individuals in a Finnish cohort (Metcalfe et al. 1996), but could not be confirmed in a larger Danish T1D family material of 245 families (Kristiansen et al. 2000b). In the second report (Bergholdt et al. 2000), we de novo sequenced the remaining and larger part of the *IL1RI* promoter region. Four novel polymorphisms were identified. Three were genotyped (the fourth were in complete LD with one of the others and were not analyzed further), in a collection of 253 Danish T1D families (1097 individuals). Of these, 103 families were simplex families (one affected offspring), 150 were multiplex families, i.e. families with two T1D affected offspring. Genotyping was performed by PCR-RFLP assays designed for each variant and analyses of genotyping data were performed by Sib-TDT analysis (Spielman et al. 1993; Spielman and Ewens 1996; Spielman and Ewens 1998); methods still widely used for evaluation of candidate genes. One SNP in the promoter region demonstrated T1D association (Bergholdt et al. 2000). We evaluated IL-1 receptor type 1 plasma levels and stratified according to genotypes of this SNP, and demonstrated statistically significant difference among genotypes, suggesting an allele dosage effect, which, however, was not disease dependent (Bergholdt et al. 2000). The major concern regarding these studies are the limited sample sizes used, e.g. for mutation scanning by sequencing only five to eight individuals, implying that we could have missed even common variants. Variants in the *IL1RI* gene have also been demonstrated associated to other diseases, e.g. the HinfI variant we detected, was associated with increased risk of helicobacter pylori infection, but not gastric cancer (Hartland et al. 2004), to our knowledge it has not been tested in other autoimmune diseases (Hollegaard and Bidwell 2006). Another SNP in the gene was associated with the metabolic syndrome (McCarthy et al. 2003).

Studies of the Komeda diabetes-prone rat, a spontaneous animal model of T1D, characterized by autoimmune destruction of pancreatic beta cells, as well as signs of autoimmunity, identified the *Cblb* gene (Casitas-B-lineage lymphoma b), as a major susceptibility locus (Yokoi et al. 2002). Cbl-b is an ubiquitin-protein ligase and functional studies have indicated an important inhibitory role in T-cell co-stimulation. Furthermore, in human T1D, an increased frequency of other autoimmune diseases is often seen (Payami et al. 1989; Chikuba et al. 1992; Tait et al. 2004), resembling the phenotype of the KDP rat, suggesting the human *CBLB* gene to be a candi-

date for conferring susceptibility to human T1D and autoimmunity in general. We therefore screened the coding regions of the human *CBLB* gene for mutations, using automated sequencing, in a panel of 24 individuals affected with several autoimmune diseases, as well as five controls (Bergholdt et al. 2005c). One variant in exon 12 was in a large family material of 480 families significantly associated to T1D (Bergholdt et al. 2005c). Human *CBLB* gene variants have also been tested for association to T1D by two other research groups (Kosoy et al. 2004; Payne et al. 2004). The exon 12 SNP was in both these studies identified by sequencing, but not genotyped further due to a low minor allele frequency (MAF of 0.031 and 0.043), based on frequencies in sequencing panels of 16 and 32 subjects, respectively (Kosoy et al. 2004; Payne et al. 2004). The minor allele frequencies in these studies were, however, comparable with the one observed by us (0.024) (Bergholdt et al. 2005c), but their strategy may have missed important information regarding this SNP.

Furthermore, these studies were performed in heterogeneous populations, which may bias results due to population stratification. A recent minor case-control report could not demonstrate association of a SNP in exon 10 of *CBLB* to Graves disease in children (Chen et al. 2005). We performed pair-wise LD analyses, by use of Lewontin's standardized disequilibrium coefficient D'(Daly 2001; Gabriel et al. 2002) of the SNPs detected in *CBLB*. With a threshold of D' >0.8 for reporting LD, which is generally accepted (Carlson et al. 2004), we detected one large LD block and two smaller LD blocks, covering most of the gene. This indicates that genotyping just few SNPs covering these blocks, should capture most of the variation in the gene, Figure 3.

LD was, however, not absolute (D' <1), which might explain lack of association of neighboring SNPs belonging to the same block. The LD-map constructed, using GOLD (Abecasis and Cookson 2000) for visualization, for the *CBLB* SNPs we detected, is shown in Figure 3.

Our studies of the *CBLB* gene were performed before the HapMap project (www.hapmap.org) was completed. Today, information from HapMap could have provided information about haplotype blocks and which SNPs to genotype directly. The HapMap project is, by no doubt, a valuable resource for candidate gene studies, reducing or ideally removing the need for re-sequencing a chosen candidate gene.

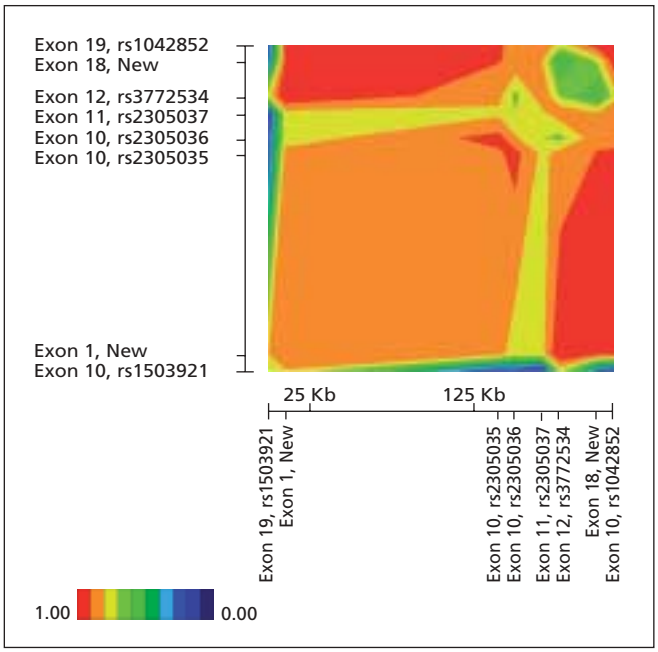

Figure 3. LD plot of identified and genotyped SNPs in the CBLB gene. SNPs are named by their exonic position as well as their dbSNP name if possible. Pairwise LD-coefficients (Lewontin's standardized disequilibrium coefficient D') are plotted by use of GOLD software. The color-coded scale is provided to the right. D'values between 0.9 and 1 are shown in red,D'between 0.8 and 0.9 in orange etc.

Functional studies supporting a candidate gene is of obvious importance, and may be viewed as a supplement or even an alternative to replication studies, especially when only small sample sizes are available. Also for the *CBLB* gene, we performed functional evaluation, and implication for functionality was demonstrated (Bergholdt et al. 2005c). An isoform of *CBLB* was demonstrated to be increasingly expressed in lymphocytes after cytokine exposure, suggesting a possible role of this isoform in T1D (Bergholdt et al. 2005c). We also evaluated potential gene-gene interaction between the *CBLB* and *CTLA4* genes involved in the same functional pathway of T-cell activation. This was done by stratifying *CBLB* data according to high risk versus lower risk genotypes of the established *CTLA4* marker, *CT60*, which is associated to T1D (Ueda et al. 2003). Since *CBLB* association to T1D was strengthened in the high risk group, observed by an increased distorted transmission in the subgroup of families with the high risk *CTLA4* genotype, interaction among the two genes were suggested (Bergholdt et al. 2005c).

We selected the *IL1RI* and *CBLB* genes as candidate genes based on their functional implication in pathways of high relevance in T1D pathogenesis. The cytokine gene, Interleukin-12 subunit B, *IL12B*, was selected for further examination based on its implication as a candidate gene for *IDDM18*, a region suggested by linkage analysis to be potentially important for T1D (Warrington and Bengtsson 1994; Huang et al. 2000; Morahan et al. 2001). IL-12 influences T-cell responses (Adorini 2001) and drives the differentiation of T-lymphocytes towards the Th1 subset, characterized by production of cytokines leading to cell mediated immunity (Manetti et al. 1993; Liblau et al. 1995; Tian et al. 1998). Furthermore in the non-obese diabetic (NOD) mouse, IL-12 was shown to play a primary role in T1D induction (Trembleau et al. 1995; Lamont and Adorini 1996; Adorini 2001). We sought to replicate a strong association of a 3'UTR SNP in the *IL12B* gene with T1D, as well as an observation of allele-dosage effects on the mRNA level of the expressed gene, reported by others (Morahan et al. 2001). We could not, in two large family collections (337 + 795 T1D families) replicate T1D association. We also typed adjacent microsatellites and performed haplotype analyses, but could still not confirm association. Neither functional significance nor a significant allele dosage effect on either mRNA or protein levels could be demonstrated in our studies (Bergholdt et al. 2004a). In conclusion, association of *IL12B* has been demonstrated in some T1D populations (Morahan et al. 2001; Yang et al. 2006), but not in several others (Johansson et al. 2001; Dahlman et al. 2002; Davoodi-Semiromi et al. 2002; McCormack et al. 2002; Nistico et al. 2002; Holm et al. 2003; Bergholdt et al. 2004a), suggesting genetic heterogeneity. One study only observed association in late onset T1D (Windsor et al. 2004). Functional studies on *IL12B* variants have been conflicting as well (Morahan et al. 2001; Dahlman et al. 2002; Seegers et al. 2002; Bergholdt et al. 2004a; Stanilova and Miteva 2005). Further, our data demonstrated inter-individual differences in stimulated IL-12 expression, however, not related to genotype of the 3'UTR SNP, underlining the complexity of the IL-12 system. Large inter-individual differences have been observed by others as well (Peng et al. 2006). The significance of the *IL12B* gene in T1D is therefore not yet clear. Association of the *IL12B* gene has been shown to other diseases, i.e. asthma (Morahan et al. 2002b; Randolph et al. 2004), chronic hepatitis C (Houldsworth et al. 2005) and cerebral malaria (Morahan et al. 2002a), however, neither in rheumatoid arthritis (Orozco et al. 2005a), nor with the presence or severity of coronary artery disease (Momiyama et al. 2005).

5.3. CLASSICAL GENOME-WIDE LINKAGE ANALYSES

Linkage analysis is the study of co-segregation of DNA variants with a phenotype or disease. No assumptions of biological mechanisms

are made, and linkage studies are thereby not limited by specific hypotheses of disease pathogenesis. Since the function of the majority of genes characterized to date is unknown, and many unidentified genes exist, linkage analysis provides an objective tool for identifying yet unknown genes or chromosomal regions contributing to multi-factorial diseases. Genome-wide scans use affected sib-pairs or multiplex families, studied with a large set of polymorphic markers covering the entire human genome. Allele sharing is compared with disease-state and a statistical estimate of excess allele sharing can be calculated, leading preferentially to identification of the genomic position of loci, that influences the disease. Classical linkage analyses are usually perfomed using microsatellites, i.e. DNA sequences made up of repeats 2-8 nucleotides in length, randomly distributed throughout the genome. These are more polymorphic and thereby more informative, than a similar number of bi-allelic SNPs. Genome wide linkage analyses normally include at least 300-400 microsatellite markers, resulting in an average marker density of approximately 10 cM or less. The number of microsatellites may subsequently be extended in order to fine-map a specific area. The quite low resolution implies that signals from linkage scans can be broad. Data from families are required to detect linkage, as co-segregation of chromosomal loci can be observed only in relatives. Statistical support for linkage is presented in the form of a LOD score (logarithm (log10)) of the likelihood ratio for linkage). A LOD score of 3, taken as strong evidence of linkage, corresponds to odds of 1000:1 in favor of linkage. In studies of affected sib-pairs usually a non-parametric linkage score is used. Thresholds developed for linkage mapping of monogenic disorders with complete information of genotype and phenotype, may, however, not be optimally suited for mapping more complex polygenic disorders. Lander and Kruglyak (Lander and Kruglyak 1995) therefore proposed that the LOD threshold for "significant" genome-wide linkage in sib-pair studies should be raised to 3.6 ($p<\times10^{-5}$), while that for "suggestive" genome-wide linkage (random single occurrence in a genome-wide scan) should be set at 2.2 ($p<7\times10^{-4}$). In addition they suggested reporting all nominal p-values<0.05 without any claim for linkage.

A typical positive first-stage genome scan usually provides linkage to one or more genetic intervals of approximately 10-30 cM in genetic width, which may include up to hundreds or thousands of genes. This is a huge distance at the molecular level, quite often with few biological clues as to the exact molecular basis of the linkage signal.

Linkage studies have been most successful in mapping genes underlying monogenic diseases, based on co-segregation of markers with the disease in affected large pedigrees. Markers are typically microsatellite markers, flanking the disease gene. Genome scans have also been used in the analysis of many complex diseases (where linkage peaks are generally wider), i.e. autoimmune and infectious diseases, phsychiatric diseases, hypertension, as well as animal models of complex human diseases. However, linkage studies have had only limited success in identifying genes for multi-factorial diseases, and the genes can in most cases only explain a small fraction of the overall heritability of the diseases. Several factors may be implicated in this lack of success, i.e. low heritability, inability of a standard set of microsatellite markers, spaced 10 cM apart, to extract complete information about inheritance and inadequately powered study designs (Hirschhorn and Daly 2005). Limitations of genome scans when applied to multi-factorial diseases may furthermore include heterogeneity in disease phenotypes, population and ethnic differences, imperfect statistical and analytical methods, and the use of different polymorphic marker panels and genetic maps in different studies. This has in several cases resulted in lack of replication and identification of different loci by different research groups studying the same disease. Possible solutions may imply using denser marker sets, larger sample sizes or larger pedigrees (Hirschhorn and Daly 2005).

Even when convincing evidence of linkage is obtained, extensive studies of putative candidate genes are still required. Systematic fine mapping of identified regions is needed to characterize variation(s) of the gene or genes responsible for the observed linkage signals, and to move from a broad region of linkage, usually exceeding 10 cM (~10 million basepairs), to the causal gene or genes within a region. This should be followed by functional characterization (in vitro and in vivo) of the biological effect of the identified genetic variants. Systematic fine mapping is laborious and complex. Linkage peaks may cover many genes (up to 1000) and may compromise several susceptibility loci, as suggested from animal models (Todd and Wicker 2001; Nguyen et al. 2002). Furthermore, a linkage signal may be observed due to chance clustering of several disease loci, each with relatively weak locus-specific effects (Concannon et al. 2005). Such complexity probably also, at least partly, explains some of the difficulties in obtaining reproducible association results. Fine mapping of identified linkage regions, in the same populations, is crucial in order to characterize the disease associated variants. Recently, the first identifications of disease-associated genes in complex diseases (Mb. Crohn, Asthma, Schizophrenia and myocardial infarction), using this strategy, were published (Hugot et al. 2001; Ogura et al. 2001; Stefansson et al. 2002; Van-Eerdewegh et al. 2002; Helgadottir et al. 2004). Another approach for increasing power in fine mapping linkage regions, may be the use of stratification/enrichment for cases from i.e. multiplex families, or from families carrying the strongest evidence for linkage.

Genome-scale linkage analyses in T1D have resulted in significant linkage for the MHC region and a number of non-MHC loci, however, with varying levels of statistical significance. In no case, so far, has a specific gene variant been proven to be the cause of genetic linkage to T1D.

In our own studies we fine mapped a region on chromosome 21, which showed linkage to T1D in the Scandinavian T1D genome scan (Nerup et al. 2001). The observed linkage was novel, and mapped to the distal part of chromosome 21 (21q21.3-qTel). The initial evidence for linkage was indicated by a maximum LOD score of 1.23 (p=0.009) (Nerup et al. 2001), and linkage was most evident in the Danish population (maximum LOD score of 2.33 (p=0.009)), Figure 4. The linked region on chromosome 21 comprised approxi-

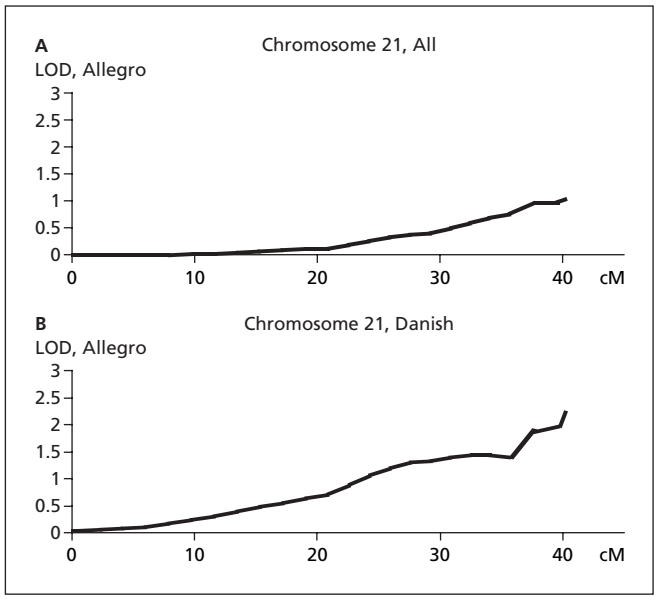

Figure 4. Linkage curve for chromosome 21. Obtained from the Scandinavian T1D genome scan. The curve represents **A.** data from the combined Scandinavian material (Danish, Swedish and Norwegian) as well as **B.** exclusively from the Danish population, modified from (Nerup et al. 2001). The X-axis represents the length of chromosome 21 (~57 cM), demonstrating that the peak of the curve is towards the distal part of the chromosome. Y-axis represents LOD-score as calculated by the Allegro software.

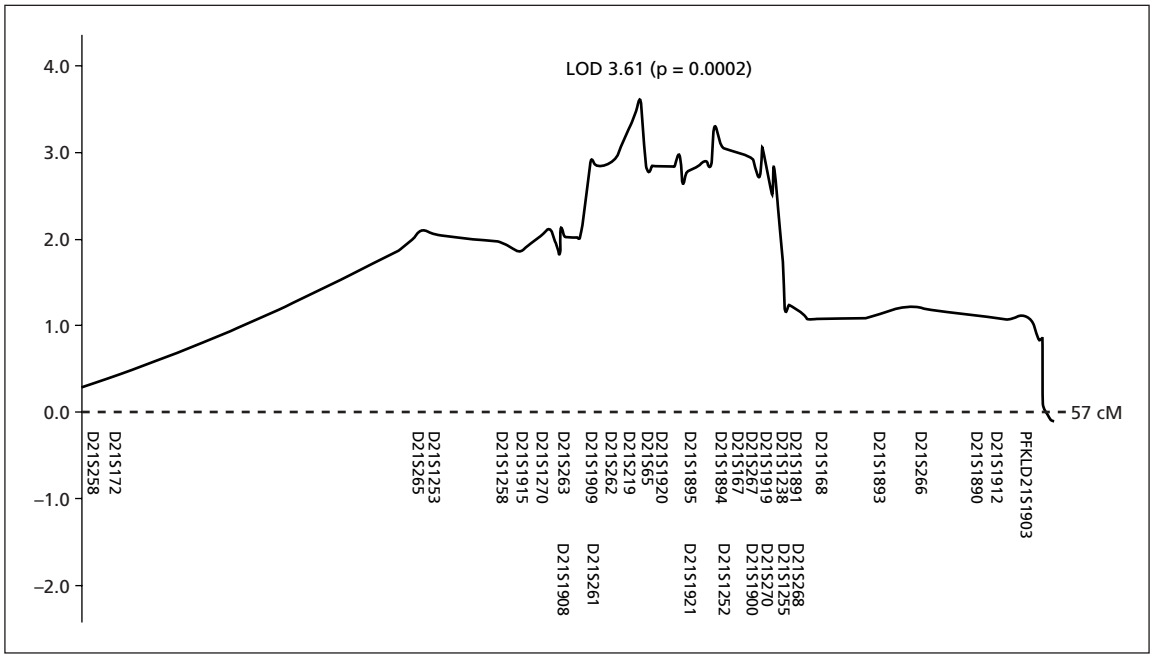

Figure 5. Multipoint LOD score analysis of a T1D linked region on chromosome 21. Modified from (Bergholdt et al. 2005a). X-axis: position on chromosome 21 in cM, Y-axis: NPL LOD score. Below the X-axis the position of the corresponding microsatellite markers are plotted.

mately 20 Mb, which equals ~35 cM (www.ncbi.nlm.nih.gov/genemap).

Furthermore interaction analyses had recently given additional support to the chromosome 21 region, as being important in T1D, since in neural network and decision tree analyses of Danish and Swedish T1D genome scan data, the same region was identified (Pociot et al. 2004).

Almost the complete sequence of chromosome 21 was known when these studies were performed and 225 genes were identified on the chromosome (Hattori et al. 2000). That the chromosome was almost fully sequenced meant that new strategies for identification of genes could be applied instead of classical positional cloning, and that the position of genes, microsatellites and sequence variations were identifiable, though some uncertainty still existed about the exact map order of some of the markers.

The strategy used was to narrow the linked region, initially by linkage analysis including an increased number of microsatellites covering the region of interest in a multipoint NPL analysis. For the current study we initially used the Marshfield map (http://research.marshfieldclinic.org) for selection of microsatellite markers, but after advances in the completion of the human genome sequence, we used marker order and inter-marker distances based on the physical map (NCBI, National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov) of chromosome 21 for further analyses. A recent study has demonstrated the importance and impact of correct marker order for linkage analyses (Gretarsdottir et al. 2002).

A study of asthma susceptibility demonstrated that a one-LOD decrease in the support interval of the linkage peak, is likely to contain the susceptibility locus if the region has been saturated with markers (Van-Eerdewegh et al. 2002). In the present study the information content of the region, with the additional markers, was 95-100%, and we obtained a maximum NPL score of 3.61, Figure 5.

We identified the corresponding support interval using NPL statistics (referred to as the "one LOD drop" interval) and found that it covered approximately 6.3 Mb, which was further studied by LD mapping with gene-based SNPs. We identified 32 candidate genes in the narrowed region, Figure 6.

Several strategies could be chosen from here. The ideal one includes re-sequencing of all candidate genes in order to identify all mutations. However, due to the still quite significant number of 32 genes, we chose to examine the genes by identifying SNPs in coding regions (cSNPs) of the genes and initially screened the identified 74 cSNPs in a panel of 96 diabetic subjects. SNPs with a minor allele frequency above 3%, were further examined. These SNPs were genotyped and examined for T1D association in a larger T1D family material (253 families). None of the SNPs showed association to T1D, whereas one SNP in the *TTC3* gene did show distorted transmission to unaffected offspring, which might support a possible protective effect.

The *TTC3* gene is expressed in most human tissues, including pancreas. It has been speculated that overexpression of this gene may be involved in some of the morphological anomalies observed in Down's Syndrome (Tsukahara et al. 1996). We also performed stratification of SNP data, by dividing genotype data into subgroups, according to simplex or sib-pair family status, and according to families carrying the linkage signal in the region and families not consistent with linkage, respectively. This approach is based on a theory predicting that genetically loaded individuals of families presumably provide a stronger signal, thereby facilitating identification of disease susceptibility variants (Fingerlin et al. 2004). The strategy

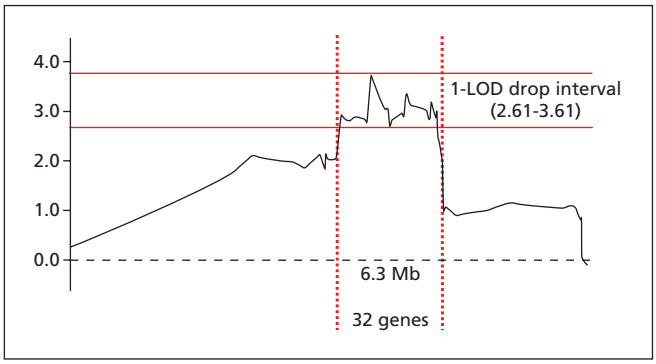

Figure 6. Identification of the 1-LOD drop interval of a T1D linked region on chromosome 21. The figure shows schematically how the 1-LOD drop interval is calculated (3.61-1=2.61) and correlated to microsatellite marker positions on the chromosome. These positions are subsequently used to identify candidate genes in the region, i.e. the genes positioned between the two microsatellite markers flanking the 1-LOD drop interval.

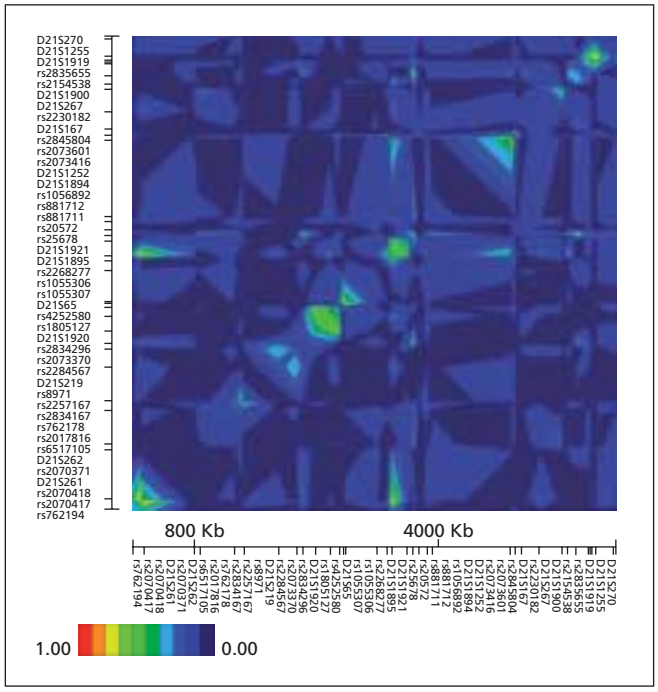

Figure 7. LD analysis, by HaploXT and GOLD, of chromosome 21 SNPs and microsatellites. Pairwise LD coefficients (D') between all genotyped SNPs and microsatellites in the region were calculated and visualized by GOLD. The color-coded scale is shown. Markers and SNPs are ordered according to their physical order along chromosome 21.

of selecting for further studies families that show evidence for linkage to a region, has been used previously (Horikawa et al. 2000; Van-Eerdewegh et al. 2002). Interestingly, we demonstrated significant T1D association of a *TTC3* SNP in the sib-pair family subgroup, and in addition three other SNPs in *OLIG2*, *CBR1* and *KCNE1* showed association in the subgroup of families responsible for the observed T1D linkage of the region.

Additionally, the degree of linkage disequilibrium in the linked region of 6.3 Mb was evaluated by pair-wise LD measurements. We were, however, only able to identify two blocks, extending more than two markers with D'-values >0.5. Thus, there does not appear to be large blocks with strong LD in this region in the Danish population, in agreement with previous studies on chromosome 21 (Olivier et al. 2001; Patil et al. 2001), although one block containing five markers was demonstrated, spanning approximately 100 kb, Figure 7. These data support that one or more T1D linked genes exists in the region, even though no evidence for association was found for the tested SNPs in the complete family material. Despite the evidence for linkage in the Danish population, linkage to T1D on chromosome 21 has not been reported in other genome scans. However, it cannot be ruled out that it was missed in earlier studies due to inclusion of relatively few markers from this region on chromosome 21 (Davies et al. 1994; Hashimoto et al. 1994; Concannon et al. 1998; Mein et al. 1998; Cox et al. 2001). It can be estimated that the information content of this particular region in previous genome scans did not exceed ~0.66. In the most recent combined genome scan of Scandinavian and UK/US data, chromosome 21 was not identified either (Concannon et al. 2005). For this genome scan 22 markers for the UK/US data set and 8 markers for the Scandinavian data set were used on chromosome 21 (7 markers overlapped) (Concannon et al. 2005). If population heterogeneity, regarding this chromosome 21 region, do exist, the "Scandinavian" contribution may not be strong enough compared to the large number of UK/US families, to show up in the combined results.

Certainly, further genotyping and refinement of the genetic map corresponding to this region on chromosome 21 is important. Attempts to replicate our observation have so far not been published. LD mapping with SNPs coupled with genome scans are increasingly being used and are considered valuable for fine mapping and genetic association studies (Ardlie et al. 2002a; Tabor et al. 2002). Genotyping a large amount of SNPs is still time-consuming and costly, but important in terms of fully characterizing and fine mapping a linkage region. Our study reduced the region linked to T1D from more than 20 Mb to 6.3 Mb and increased the NPL score to 3.61. Furthermore, the number of possible susceptibility genes was reduced to 32. Although we have not yet been able to identify a single gene variant responsible for the observed linkage, we demonstrated significant support for several combinations of markers within this region. Furthermore, some evidence for association to T1D of variants of the *TTC3*, *OLIG2*, *KCNE1* and *CBR1* genes in conditioned analyses was observed (Bergholdt et al. 2005a). As our data suggest the existence of T1D linked gene(s) in the region, further characterization of the candidate genes on chromosome 21 is important, including analyses of regulatory and promoter regions, in which variants of potential importance for regulation of gene expression and thereby functional significance, might be located. Possible significance of non-coding SNPs in disease predisposition of e.g. T1D and T2D, have been demonstrated several times, (e.g. *CALP10*, *CTLA4*, *RUNX1*) (Horikawa et al. 2000; Prokunina et al. 2002; Helms et al. 2003; Tokuhiro et al. 2003; Ueda et al. 2003). Our subsequent analyses of chromosome 21 are discussed later in this thesis.

5.4. COMPARISON OF LINKAGE VS. ASSOCIATION STUDIES

The capability of association and linkage studies to detect allelic association is influenced by several factors, e.g. power of the study, sample size, effect size and the frequencies of the alleles (Zondervan 2004). The typical frequencies of variants, that underlie common diseases are largely unknown, but several examples of common variants (e.g. frequencies >1%) have been shown to influence disease susceptibility, most of which, however, only increase the risk of disease by two-fold or less, when examined in large populations (Lohmueller 2003). For most common diseases the phenotype presumably is determined by a combination of susceptibility and/or protective genes and interactions between multiple genetic and environmental factors.

Hence, any individual genetic variant will probably have a small effect on disease risk. It is reasonable to suppose that common alleles, as well as rare alleles, will contribute to common disease. It has been demonstrated that linkage analysis often is more powerful than association analysis for identifying rare high-risk disease alleles,

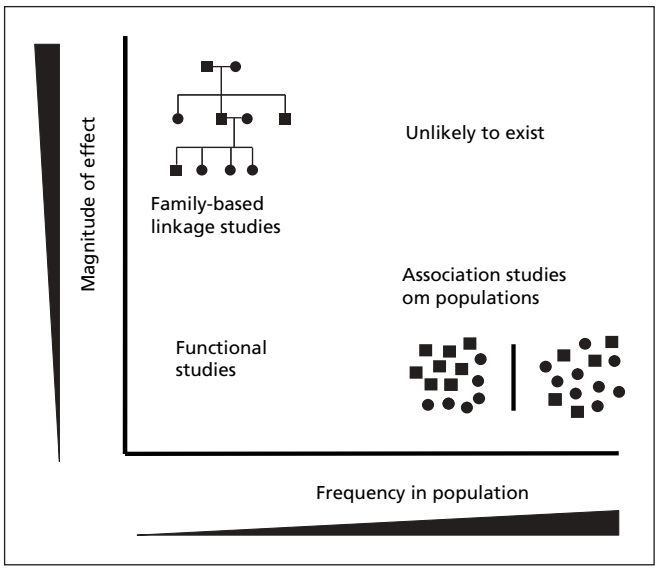

Figure 8. Method of choice for obtaining maximum power for evaluation of different variants depends on allele frequency and severity/function of the variant.

Table 2. Power calculations for genome-wide linkage analyses, and detection of minor susceptibility loci (i.e. $\lambda s = 1.12$) for T1D.

	Number of affected sib-pairs needed $\lambda s = 1.12$. Power	
Evidence for linkage	90%	99%
Suggestive ($p=7.4 \times 10^{-4}$)	4,400	6,700
Significant ($p=2.2 \times 10^{-5}$)	6,375	9,092

whereas association studies are expected to be more powerful for the detection of common disease alleles that confer small disease risk (Risch and Merikangas 1996).

Furthermore it has been pointed out that genetic association studies can provide not only greater power, but also better resolution of location than linkage studies (Risch and Merikangas 1996). Method of choice for obtaining maximum power for evaluation of different variants depends on allele frequency and severity/function of the variant in question, as indicated in Figure 8.

Based upon the existing sample sizes, used in the recent combined T1D genome scan (Concannon et al. 2005), it is unlikely that non-HLA loci exist with odds ratios >3 ($\lambda_S > 1.3$).

Recent power calculations of number of affected sib-pairs needed in genome-wide linkage analyses to detect linkage for minor risk loci, according to the criteria, set forward by Lander and Kruglyak (Lander and Kruglyak 1995) for suggestive and significant linkage are shown in Table 2. Collecting more than 4,400 T1D affected sib-pairs are one of the goals of the Type 1 Diabetes Genetics Consortium (www.t1dgc.org). Association studies of genes with a plausible biological mechanism in T1D as well as genome-wide linkage studies both continue to be valuable methods. Increasingly, however, the focus is moving to genome-wide association studies, using SNPs, which may have the power to permit detection of the additional genes with effects in the appropriate range, i.e. odds ratios <2, if robustly powered and performed. This approach is now technically possible and will probably be broadly applied in the future.

5.5. GENOME-WIDE ASSOCIATION STUDIES

Genome-wide association studies might be advantageous for identifying variants associated with common diseases (Hirschhorn and Daly 2005; Wang et al. 2005). The definition of a genome-wide association study is an association study that surveys most of the genome for causal genetic variants. Since no assumptions are made about the genomic location of the causal variants, this approach could use the strengths of association studies without having to guess the identity of the causal genes (Hirschhorn and Daly 2005). Genome-wide association studies may include up to hundreds of thousands of SNPs (Hirschhorn and Daly 2005; Wang et al. 2005), much larger numbers than in classical linkage studies with less than thousand microsatellites. Using this amount of markers and corresponding statistical tests may, however, result in numerous hypothesis-testing issues (Hirschhorn and Daly 2005; Wang et al. 2005). Genome-wide association studies are unbiased, quite comprehensive and can be used in the absence of convincing evidence regarding the function or location of the causal genes leading to a specific disease, and can obtain greater power than classical linkage analysis. Selecting an appropriate sample size is important and because variants that contribute to complex traits are likely to have modest effects, large sample sizes are necessary. Required sample sizes are further increased by the large number of hypotheses that are tested in a genome-wide association study, p-values obtained should be corrected for multiple testing. It has been proposed that $p=5\times10^{-8}$ (equivalent to a p=0.05 after a Bonferroni correction for 1 million independent tests) would be a conservative threshold for declaring a significant association in a genome-wide association study (Hirschhorn and Daly 2005). Alternatively, more liberal p-value thresholds could be used, but too liberal thresholds will require extensive follow-up and replication studies to distinguish false-positives from true associations. It has, however, been argued that a classical Bonferroni correction is too conservative, requiring too low p-values and too large sample sizes. With the high density of markers, significant LD between many markers and redundancy between single markers and multimarker haplotypes, which might also be tested for association, means that independency among tests is not the case. Alternative strategies for defining thresholds have been proposed, i.e. permutation testing, to assess the probability of having observed a particular result by chance (Hirschhorn and Daly 2005). An approach using two or three stages has also been proposed and may be advantageous in selecting a minimum amount of markers, while still obtaining reasonable power (Hirschhorn and Daly 2005). Thus, an initial scan is performed with a modest threshold for passing markers as positive, however, still with a threshold ensuring power for detecting loci explaining small fractions of the phenotypic variance (Hirschhorn and Daly 2005). Using this approach a limited number of false-positives will also pass, but in a second, independent population sample, similar in size or larger, the passed markers are tested again in stage 2. SNPs typed in stage 2 could even be genotyped by a different genotyping method, minimizing the risk of reporting false associations. A third step may then be used for even greater stringency (Hirschhorn and Daly 2005). To deal with the problem of multiple hypothesis testing such multi-stage designs may also be advantageous (Hirschhorn and Daly 2005), since the number of genotyped SNPs can be reduced in each stage, allowing genome-wide significance to be achieved step by step. (Laird and Lange 2006).

As it has been estimated that the genome contains approximately 12 million common SNPs (Kruglyak and Nickerson 2001), and that groups of neighbors correlated with each other exist, from which the genotype of one SNP perfectly predicts those of correlated neighboring SNPs (htSNPs) (Johnson et al. 2001), one SNP can serve as a proxy for many others, which are then indirectly assayed in an association screen. Once the patterns of LD are known for a given region, a few tag SNPs can be chosen such that they individually or in multimarker combinations (haplotypes) capture most of the common variation within the region (Johnson et al. 2001; Gabriel et al. 2002). The precise number of tagSNPs needed is currently not known, and depends on methods used to select the SNPs, the degree of long-range LD between blocks and the efficiency with which SNPs in regions of low LD can be tagged. Various algorithms for selecting tagSNPs have been proposed (Wang et al. 2005) It is obvious that a proportionally higher density of variants must be typed from regions showing low LD as compared to regions showing a high degree of LD to capture all information.

The amount of data produced will raise issues of analysis and interpretation. To be useful, markers tested for association must either be the causal allele or highly correlated (in LD) with the causal allele (Kruglyak 1999). Knowledge regarding LD patterns on a genome-wide scale is therefore essential. Attempts to determine LD patterns on a genome-wide scale are ongoing through the HapMap project [Consortium, 2003 #649] (www.hapmap.org), which provides a haplotype-based map of informative SNPs covering the entire genome, particularly useful for methods that use markers selected on the basis of LD. Phase I of the HapMap project has been completed (2005) and includes information on approximately 300 million genotypes (Altshuler et al. 2005). It has been demonstrated that large parts of the genome falls into segments of strong LD, within which variants are strongly correlated with each other (Daly 2001; Patil et al. 2001; Gabriel et al. 2002; Altshuler et al. 2005).

Another approach proposed for selecting SNPs for genome-wide association studies is the so called missense approach, in which only missense SNPs (SNPs changing the amino acid sequence encoded) are used (Botstein and Risch 2003). A high proportion of missense mutations are among the alleles underlying diseases of Mendelian inheritance, since they often cause severe changes in protein function. However, even for such diseases, an appreciable fraction of muta-

tions are outside the coding regions, questioning the validity of the approach. Since alleles underlying complex traits often confer more subtle effects on disease risk, they are more likely to include non-coding regulatory variants with a modest impact on expression (Hirschhorn and Daly 2005). Missense variants have been associated with multi-factorial diseases, and obviously they are more likely to have functional consequences. However, ascertainment bias may be substantial, since until recently, missense variants have been preferentially discovered and tested for disease association. In summary, such an approach is likely to be productive, whereas it remains unclear how well common risk alleles will be captured. It has been estimated that a typical gene contains one or two missense SNPs, hence the strategy would require genotyping of only 30,000 – 60,000 SNPs (Hirschhorn and Daly 2005). On the other hand identification of all common missense SNPs would require a substantial effort in advance. It has been suggested that inclusion of SNPs from conserved non-coding regions might be advantageous. By comparative genetics, i.e. between human and mouse and rat genomes, it has been demonstrated that large parts of non-coding DNA is highly conserved among species, suggesting functional significance of such regions (Loots et al. 2000; Pennacchio and Rubin 2001; Thomas et al. 2003; Boffelli et al. 2004). This, however, will require a substantially increased number of SNPs to be included, diminishing the advantage of the missense approach as compared to selecting markers based on LD. Alternative approaches, such as typing only few SNPs in or near coding regions of each gene, are limited in coverage. Such studies only survey the variants chosen and variants in LD with them, and for sure will miss regulatory variants (Buetow et al. 2001; De La Vega et al. 2002). So, unless the LD patterns of each gene are empirically determined, such an approach might have little success and can not be called a genome-wide association study. Also attempts to convert linkage studies into association studies by looking for association between disease and the microsatellites typed in a linkage study, will survey only a small fraction of the human genome, and cannot truly be considered genome-wide association studies.

To perform genome-wide association studies, the ability to genotype large amounts of SNPs in large samples, is crucial. Genotyping technology has considerably improved and become cheaper in recent years, and prices are expected to drop further. Low costs and high quality of genotyping are essential and required to make genome-wide association studies feasible. In recent years, chips (mapping arrays) have been designed (by Affymetrix, Illumina and Perlegen companies) containing panels of genome wide SNPs, offering the most cost-effective way of genotyping the large amount of SNPs necessary. SNPs were selected on different bases, e.g. chips comprising 10,000, 100,000 and 500,000 SNPs from Affymetrix (www.affymetrix.com), on which SNPs were not selected based on LD. Another chip comprising 317,000 SNPs, termed Hap300K from Illumina (www.illumina.com) were selected mainly based on HapMap phase 1 data, however, with increased amounts of SNPs close to genes and in the MHC region, as well as enriched in non-synonymous SNPs. Also Perlegen (www.perlegen.com) perform genome-wide SNP genotypings in collaboration with researchers and companies. Chips containing 1 million SNPs are underway. Using these platforms and chips, the SNPs can be genotyped at significant lower cost per SNP. The high degree of coverage for these SNP sets implies that they are likely to cover a significant fraction of the genome, and even if they are less efficient per marker than a purely LD based set, the amount of SNPs should be high enough to ensure acceptable coverage of the genome. It is important to note that none of these marker sets will be optimal for detecting the effects of rare variants with frequencies of 1% or less.

The impact of the problem of population stratification is unknown and controversial (Ardlie et al. 2002b; Freedman et al. 2004). If a population is of mixed ethnicity, population stratification may imply different subgroups within a population with different disease prevalences. Techniques have been developed to detect and correct for the phenomenon by typing dozens of unlinked markers (Devlin and Roeder 1999; Pritchard and Rosenberg 1999; Pritchard et al. 2000; Reich and Goldstein 2001). Recently it was demonstrated that in well-matched populations, population stratification was not a major problem, whereas mild stratification was difficult to rule out (Ardlie et al. 2002b; Freedman et al. 2004). Even mild stratification may, however, confer bias when searching for alleles with modest effects. The problem of false-positive associations is essential in studies using hundreds of thousands of markers, and should be considered when designing a study. Strategies to increase the efficiency of genome-wide association studies may imply use of more homogeneous populations, e.g. like the Scandinavian or use of enriched cases, e.g. by family history or low age at onset, which is expected to increase power.

Until very recently no truly genome-wide association study had been published, although a few previous studies have used part of the approach. A strong potential association between acute myocardial infarction and the lymphotoxin-alpha gene (*LTA*) was suggested in a study with relatively few individuals (initially less than 100) genotyped for approximately 65,000 SNPs (Ozaki et al. 2002). They also followed up their initial results in a replication panel, in which association of an intronic SNP in LTA was replicated. Although this was a large project involving millions of genotypes, the low power of the initial screening sample means that the rates of both false-negative and false-positives in this study were probably high. Additionally, it is not clear what fraction of the genome was surveyed by their set of SNPs, selected from a Japanese SNP database (Ozaki et al. 2002). The lymphotoxin-alpha gene was identified against the odds of what would have been expected, taking the small initial sample size into account. Only a few other reports of genome-wide screens using SNPs had been published until very recently: Macula degeneration (Klein et al. 2005), Parkinson's disease (Maraganore et al. 2005) and obesity (Herbert et al. 2006). Relatively small sample sizes were included and the Affymetrix 100K SNP chip or Perlegen Sciences SNP chip, which only covers a fraction of the human genome were used (Klein et al. 2005; Maraganore et al. 2005; DeWan et al. 2006; Herbert et al. 2006). In addition there are several reports from different diseases where only gene-based SNPs were used, only non-synonymous SNPs, or just very few SNPs in smaller cohorts, exemplified in (Namkung et al. 2005; Yamazaki et al. 2005; Abel et al. 2006; Rudd et al. 2006; Smyth et al. 2006a; Spector et al. 2006) – these are considered far from truly genome-wide, and are not discussed further, even though some have identified candidate genes.

One genome wide association study of inflammatory bowel disease was conducted and published in 2006 using a LD based set of approximately 300,000 markers (Illumina HumanHap 300 chip) in 500 cases and 500 controls (Duerr et al. 2006). The *CARD15* gene was confirmed and a variant in *IL23R* was identified (Duerr et al. 2006).

Since then, in 2007, the field has moved extremely fast forward, significantly demonstrating the feasibility and success of genome wide association studies in complex diseases.

The first genome wide association study of T2D has been published (Sladek et al. 2007). Five loci with variants conferring T2D risk were reported in a stage 1 and preliminary stage 2 analysis (Sladek et al. 2007). The follow-up case-control cohort used for stage 2 was selected based on less stringent inclusion criteria than the original cohort used for the genome wide association study, the significance of which is unknown. The complete two-stage analysis is underway and all loci detected do need replication in other samples. Sladek et al. (Sladek et al. 2007) used two marker sets, one based on LD data from the HapMap project supplemented by a gene-centered SNP set, in total approximately 400,000 markers. However, despite this high resolution, the often replicated T2D association of the *TCF7L2* gene was only detected by positive association of one single SNP, indicating that coverage of studies like this may still be insufficient to detect all important loci.

In T2D the genome side association study of Sladek et al. has rapidly been followed by several other T2D genome scans (Saxena et al. 2007; Scott et al. 2007; Steinthorsdottir et al. 2007; Wellcome Trust Case Control Consortium 2007) and a follow-up study (Zeggini et al. 2007). Together, these reports have provided convincing evidence for several new T2D regions, the *HHEX-IDE* region on chromosome 10, the *CDKAL1* region on chromosome 6, the *CDKN2A-2B* region on chromosome 9, the *IGF2BP2*-region on chromosome 3 and the *SLC30A8*-region on chromosome 8 (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007; Steinthorsdottir et al. 2007; Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). All studies were performed with large sample sizes, using individuals of Northern European ancestry, and there were only slight differences in phenotypic disease characterizations of cases among the studies. The new regions identified were all found in three or more of the scans, however for some, different SNPs were used to identify the same locus (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007; Steinthorsdottir et al. 2007; Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). In addition in a recent genome wide association scan for obesity the *FTO* gene were convincingly identified, also of presumed importance in T2D (Frayling et al. 2007). These new regions for which neither the causal gene or genes, nor the causal variant is known, adds to the established T2D genes, i.e. *KJNJ11, PPARG, TCF7* and *WFS1* (identified through candidate-gene studies) and the *TCF7L2* gene identified through fine-mapping, reviewed in (Frayling 2007). The association of *TCF7L2* is the top finding in all the genome wide association scans for T2D, it has an odds ratio of 1.37, whereas all other ten T2D loci or genes only has odds ratios between 1.10 and 1.20 (Frayling 2007). However, the studies published does not rule out other yet unknown signals or even larger signals since the chips used does not cover all common, and very little of rare variation, in the genome.

Regarding T1D, in 2007 the Wellcome Trust Case Control Consortium published their large genome wide association study of 2,000 cases from each of seven common diseases, and 3,000 shared controls from two separate cohorts. The seven diseases were bipolar disorder, coronary heart disease, Chrohn's disease, hypertension, rheumatoid arthritis, T1D and T2D (Wellcome Trust Case Control Consortium 2007). For genotyping the 500 K GeneChip Affymetrix set was used.

For T1D, this report demonstrated seven regions that showed strong evidence of association at $p < 5 \times 10^{-7}$, i.e. 12q13, 12q24 and 16p13 demonstrating the highest signals in themselves, as well as 4q27 and 12p13 identified by multi-locus analyses and 18p11 and 10p15 identified by combined analysis of autoimmune cases (T1D, Rheumatoid arthritis and Crohn's disease) (Wellcome Trust Case Control Consortium 2007). Except the chromosome 10p15 region, containing the *CD25* gene, which encodes an IL-2 receptor (Vella et al. 2005), these strongest associated regions have not previously been demonstrated associated to T1D. The two chromosome 12 regions, 12q13 and 12q24 both map to regions of high LD containing many positional candidate genes, including several genes involved in immune signaling. In the 16p13 region only two genes of unknown function are located. The 4q27 region contains the IL-2 and IL-21 genes, which together with the 10p15 association signal may highlight IL-2 signaling as highly interesting in T1D. Also observations in the NOD mouse support IL-2 as important (Yamanouchi et al. 2007). The associated region on 18p11 only contains a single candidate gene, the *PTPN2* (protein tyrosine phosphatase, non-receptor type 2), member of the same family as *PTPN22*. Regions corresponding to the already established T1D candidate genes *MHC, CTLA4, PTPN22, IL2RA/CD25* and *IFIH1* were also identified in this genome wide association scan with p-values <0.001 (Wellcome Trust Case Control Consortium 2007). Significance levels corresponding to these loci were much lower than the strongest associated loci, and the *INS* gene region was not detected at all (Wellcome Trust Case Control Consortium 2007). No markers close to the *INS* gene were present on the chip used. Replication studies are required to confirm associations from genome wide association studies. John Todd et al. therefore followed up this primary genome wide association scan in another paper, aiming at validating the initial findings regarding T1D. Eleven regions from the initial study were sought replicated, as well as a few regions identified in their other previous scan, comprising only non-synonymous SNPs, in which they originally identified the *IFIH1* gene (Smyth et al. 2006a). The *MHC* region was excluded from the follow up analysis (Todd et al. 2007). In this second report 4,000 new individuals with T1D, 5,000 controls and approximately 3,000 family trios were used (Todd et al. 2007). Robust association to T1D of only four of the newly identified chromosomal regions, 12q24, 12q13, 16p13 and 18p11, could be replicated (Todd et al. 2007). Loci that could only be identified in the primary report, but not reproduced in the second, are probably either of small effect or false-positives, underlining the risk of reporting false-positives even in such large material. Evidence for the four replicated regions is very strong, and should be followed by extensive fine mapping and functional studies in order to identify the causal genes in each region.

Another independent two-stage T1D genome wide association scan (comprising 563 cases, 1146 controls and 483 T1D family trios of European ancestry) has just been published (Hakonarson et al. 2007). In stage 1 550,000 SNPs were genotyped on the Illumina platform. This report confirmed known T1D loci as the *HLA* region, the *INS* gene region and the *PTPN22* region and identified a new locus on chromosome 16p13, corresponding to a LD block only comprising the *KIAA0350* gene (a sugar-binding C-type lectin) (Hakonarson et al. 2007). This locus was also identified in the Welcome Trust Case Control Consortium genome wide association scan (Todd et al. 2007; Wellcome Trust Case Control Consortium 2007).

Failure to detect known association signals does not provide conclusive exclusion of corresponding genes, since coverage in such scans of neither common nor rare variants are complete. Despite large sample sizes, power is apparently still limited to detect significance levels appropriate for genome wide studies. The recent genome wide studies do, however, represent a validation of the genome wide approach. An interesting feature as the use of shared controls for several disease phenotypes is furthermore elucidated (Wellcome Trust Case Control Consortium 2007). Additionally, it was demonstrated that the extent of population stratification in the British population was modest, when all individuals with non-European ancestry was excluded (Wellcome Trust Case Control Consortium 2007).

Several other true genome wide association studies are expected to appear in the near future. There will be a need for standardized criteria for establishing significance (perhaps based on permutation testing and replication in other populations) and for quality control. If this latter threshold is low, a high amount of incorrect genotypes, but a higher call rate will be the case, whereas if the threshold is more stringent the opposite is the case. It should be kept in mind, that common SNPs only capture common variation; rare variations and copy number variants are not captured. Many false positives will be generated, and replication of interesting findings will be essential. Extensive re-sequencing, genotyping and further functional studies are therefore mandatory in order to identify the causal gene/genes. But the field of genome wide association scans has, with these new reports, proven to be feasible, applicable and succesfull in complex diseases like T1D and T2D.

6. COMPARATIVE GENETICS

Comparative genetics implies a gain of genetic knowledge of e.g. T1D by comparison with other similar diseases, other kinds of diabetes, syndromes involving T1D, as well as animal models for the disease in question.

6.1. OTHER AUTOIMMUNE DISEASES

Comparative analysis of the position of candidate loci for T1D with candidate loci from other autoimmune and/or inflammatory diseases shows considerable overlap (Becker et al. 1998; Becker 1999; Becker 2004). This supports a hypothesis that at least part of the underlying genetic susceptibility to T1D may be shared with other clinically distinct autoimmune diseases, as rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, Crohn's disease etc. Such diseases may share susceptibility loci and thereby common biological pathogenetic pathways. Several of the diseases are like T1D characterized by antibody and/or T-cell reactivity, often preceding clinical onset (Roep 1996; Gottlieb and Eisenbarth 1998; Rabinovitch and Suarez-Pinzon 1998; Schranz and Lernmark 1998).

The MHC region has been demonstrated associated or linked to disease susceptibility in almost all autoimmune diseases. The protective effect of some HLA haplotypes in T1D, on the other hand, seem specific to T1D. The relationship of the MHC with immune/inflammatory diseases is thought to be due to the essential requirement of HLA molecules in normal antigen recognition, processing and presentation, ultimately leading to activation and progression of an immune system response.

This process, central to normal immune function, is disturbed in some individuals affected by disease, however, probably only in combination with other genetic and environmental modifiers causing autoimmune disease. In T1D as well as other diseases none of the genetic variants of the HLA region are disease specific but found in affected as well as un-affected individuals. A fraction of autoimmune T1D does not posses the otherwise characteristic HLA genotypes, which are thereby neither necessary nor sufficient for T1D. Additionally, the genetic location of many proposed T1D-loci overlap with loci from other autoimmune or inflammatory diseases. The genome scan approach has apart from in T1D (Davies et al. 1994; Hashimoto et al. 1994; Concannon et al. 1998; Mein et al. 1998; Cox et al. 2001; Nerup et al. 2001; Concannon et al. 2005), been used in other complex autoimmune/inflammatory diseases, e.g. multiple sclerosis (Ebers et al. 1996; Haines et al. 1996; Kuokkanen et al. 1996; Sawcer et al. 1996), Crohns (Hugot et al. 1996; Satsangi et al. 1996), rheumatoid arthritis (Remmers et al. 1996; Cornelis et al. 1998), psoriasis (Tomfohrde et al. 1994; Matthews et al. 1996), systemic lupus erythematosus (Kono et al. 1994; Morel et al. 1994; Gaffney et al. 1998; Moser et al. 1998) and asthma (Daniels et al. 1996). Familial association of different autoimmune diseases in the same pedigree (Bias et al. 1986; Grennan et al. 1986; Ginn et al. 1998), co-incidence of different autoimmune diseases in the same individual (Lorini et al. 1996; Yamato et al. 1997), and some shared clinical parameters of different autoimmune diseases (Martin et al. 1995), suggest a common biological basis for several autoimmune/inflammatory diseases. This does not, however, explain the tissue specificity for destruction and different clinical symptoms, observed for several of the autoimmune diseases. Additionally, a recent study of variants in 16 different candidate genes, previously associated with at least one autoimmune disease, did also not support the hypothesis, since a pronounced lack of association with other autoimmune diseases, of the variants tested, was observed (Smyth et al. 2006b). The genes tested were *CRP, FCER1B, FCRL3, CFH, SLC9A3R1, PAD14, RUNX1, SPINK5, IL1RN, IL1RA CARD15, IBD5*-locus (including *SLC22A4*), *LAG3, ADAM33* and *NFKB1* (Smyth et al. 2006b). Variation within the *CTLA4* gene has consistently and convincingly, however, been demonstrated associated to several autoimmune diseases, i.e. T1D, Graves disease and autoimmune hypothyroidism (Kristiansen et al. 2000a; Ueda et al. 2003; Pociot 2004). Also variation in *PTPN22* has recently been demonstrated associated to several autoimmune diseases apart from T1D, e.g. rheumatoid arthritis, Grave's disease, systemic lupus erythematosus, systemic sclerosis and Wegener's granulomatosis (Bottini 2004; Onengut-Gumuscu et al. 2004; Smyth et al. 2004; Carlton et al. 2005; Jagiello et al. 2005; Lee et al. 2005; Orozco et al. 2005b; Skorka et al. 2005; Gourh et al. 2006; Onengut-Gumuscu et al. 2006; Wellcome Trust Case Control Consortium 2007).

Pathways affecting pro-inflammatory and anti-inflammatory cytokines, T-cells, antibody regulation, apoptosis and hormonal control of the immune system are the most likely to be involved in related autoimmune diseases (Rose 1997). The eventual genetic basis underlying co-localization of loci between multiple diseases could for each locus be due to either the same allele of the same gene, to different alleles of the same gene or different genes in a complex of functionally related genes or it may be coincidental. A general pattern of locus co-localization is not found in human non-autoimmune disease (Becker et al. 1998).

Even though the end-stage phenotype of a specific autoimmune disease may be clinically distinct and/or organ specific, the pathogenesis of many autoimmune diseases may still, however, involve shared processes of immune regulation. Overlapping loci between T1D and other autoimmune/inflammatory diseases suggests that in some cases, common pathways may be involved in the etiology of clinically, distinct autoimmune diseases. Environmental factors, other genetic loci, as well as epigenetic mechanisms may contribute to disease or target-tissue specification. There has not yet been a specific effort to study different autoimmune/inflammatory diseases using a standard panel of polymorphic markers to allow direct comparisons.

6.2. TYPE 2 DIABETES

T1D and T2D [MIM #125853] has been considered genetically and pathophysiologically distinct entities, although clinically, they may sometimes be difficult to distinguish. Increasing evidence links T1D and T2D (Pietropaolo et al. 2000; Mathis et al. 2001; Wilkin 2001; Maedler et al. 2002), and there is accumulating evidence that different inflammatory mediators are of importance in the pathogenesis of both diseases. T2D is the result of insufficient insulin production to maintain normoglycemia in the face of insulin resistance and it may develop gradually over years. T2D comprises 85-90% of all diabetes, and an increasing incidence of T2D is seen in children and adolescents. Pancreatic beta cells are sensitive to a number of pro-apoptotic stimuli believed not only to be involved in the development of T1D but also T2D (Mathis et al. 2001; Donath et al. 2003). In T2D, beta cell deficit and increased beta cell apoptosis have also been reported (Butler et al. 2003) and may contribute to the relative insulin deficiency which characterizes T2D. Furthermore, cytokines induce insulin resistance in peripheral tissues (Starr et al. 1997; Kim et al. 2001a; Grimble 2002), and elevated circulating levels of pro-inflammatory cytokines in recent onset T2D subjects have been reported (Esposito et al. 2002; Kubaszek et al. 2003; Spranger et al. 2003). Hyperglycemia in T2D patients is accompanied by increased circulatory levels of pro-inflammatory cytokines (Esposito et al. 2002). Thus, cytokines produced locally as part of the autoimmune infiltrate of the pancreatic islets in T1D or expressed in beta cells and/or at high levels in circulation in T2D may be a common pathogenetic denominator in beta cell failure of the two diseases. However, in T1D cytokines produce beta cell destruction and hyperglycemia, i.e. they are causative, whereas in T2D hyperglycemia induces islet production of cytokines possibly leading to beta cell destruction, i.e. they are a consequence.

Approximately 10-15% of T2D patients have measurable circulating GAD antibodies indicating that they have a slowly progressing beta cell failure, so-called LADA (latent autoimmune diabetes of the adult) (Dahlquist et al. 1989; Li et al. 2001; Mathis et al. 2001; Naik and Palmer 2003). Clustering of T1D and T2D in families has been reported in several studies (Dahlquist et al. 1989; Carel et al. 1993; Li et al. 2001). Such observations could suggest overlap in the pathogenesis of T1D and T2D in some cases.

Use of comparative genetics between T1D and T2D is therefore obvious. Studies of "classical" T2D markers have been attempted in T1D materials. A consistent, but of modest effect, association of the *PPARG* (a nuclear receptor involved in glucose homeostasis, lipid

storage and adipocyte differentiation) *pro12ala* polymorphism to T1D has been demonstrated in a large T1D family material of 2355 European families, RR 1.15 (1.04-1.28) (Eftychi et al. 2004), as well as in our own Danish family material (490 families), a report in which also a meta-analysis (containing 1691 transmissions) demonstrated association (Johansen et al. 2006). T1D association of the *KCNJ11* (encoding Kir6.2, an inwardly rectifying potassium channel, forming the ATP-sensitive K^+-channel in complex with the sulfonylurea receptor) glu23lys variant, and the *TCF* (encoding the transcription factor, HNF-1α) ala98val variant could not be demonstrated (Eftychi et al. 2004; Johansen et al. 2006). The IRS1 (insulin receptor substrate 1, crucial in insulin signaling) gly972arg variant demonstrated association in some populations (Federici et al. 2003; Morrison et al. 2004), but could not be confirmed in other (Eftychi et al. 2004; Johansen et al. 2006).

Another approach for gaining knowledge about T1D genetics through T2D genetics, may be by studying the genetic basis of pathways and key molecules that may contribute to both T1D and T2D risk, e.g. it is well documented that cytokine production and action are under genetic influence.

Since it was first proposed that inflammation and activated innate immunity might play a role in the pathogenesis of T2D (Pickup et al. 1997), several studies have addressed the role of markers of inflammation, e.g. cytokines, in the development of T2D (Pickup 2004). In several studies, IL-6 and TNF-α are positively correlated with measures of insulin resistance and/or plasma insulin concentrations, body mass index (BMI), and circulating triglycerides (Muller et al. 2002; Temelkova-Kurktschiev et al. 2002a; Temelkova-Kurktschiev et al. 2002b). Cytokines may play a role also in the pathogenesis of T2D by direct effects on the beta cell. Combinations of TNF-α, IL-1β, IL-6 and IFN-γ can induce beta cell apoptosis through induction of signaling pathways that activate e.g. NFκB (Donath et al. 2003). Variation in genes encoding all these molecules may provide insight into the pathogenesis of T2D and T1D. Specific cytokine gene variants and their role in T1D and T2D are reviewed in (Bergholdt and Pociot 2006), demonstrating still limited overlap in genetic variants evaluated in both diseases.

A recently identified and consistently replicated T2D candidate gene is *TCF7L2* (Cauchi et al. 2006; Damcott et al. 2006; Florez et al. 2006; Grant et al. 2006; Groves et al. 2006; Scott et al. 2006; Zhang et al. 2006), a transcription factor involved in glucose homeostasis, capable of regulating pro-glucagon gene expression in entero-endocrine cells via Wnt signaling (Grant et al. 2006), and possibly early beta cell development (Cauchi et al. 2006). Hence, this gene may also be an interesting gene in T1D. The causal variant is not yet known. In T2D a population attributable risk of 21% was calculated (Grant et al. 2006), whereas no studies of association of *TCF7L2* with T1D has yet been published. Very recently the first genome-wide association study of T2D was published (Sladek et al. 2007). The *TCF7L2* gene on chromosome 10 was confirmed, and four novel loci identified, a variant in a zink transporter gene, *SLC30A8* on chromosome 8, expressed exclusively in beta cells, as well as three variants corresponding to two LD blocks containing genes potentially involved in beta cell development and function (*IDE-KIF11-HHEX* (chromosome 10) and *EXT2-ALX4* (chromosome 11)). The PAR calculated for these loci in common is approximately 70%. The subjects used in stage 1 of this genome-wide association study were chosen based on a positive family history of T2D and lack of obesity (Sladek et al. 2007). This may enrich for individuals carrying a smaller number of risk alleles with stronger effects and for variants in genes implicated in pancreatic development and beta cell function, thereby also of potential importance in T1D, which needs to be explored.

6.3. KNOWLEDGE FROM SYNDROMES

Another feature of comparative genetics is the use of knowledge from syndromes, in which T1D is one of the characteristics, and from rare forms of diabetes. Such studies may provide possibilities of examining disruption of individual components of glucose homeostasis. It should be taken into account that diabetes, resembling T1D, in some syndromes, is not associated with autoimmunity and may not have the same characteristics and underlying pathophysiological mechanisms as classical T1D. Some syndromes and rare forms of diabetes are characterized by a reduced or complete lack of beta cell mass, some manifest at birth, others are characterized by insulin resistance, reviewed in (Barrett 2001; Polak and Shield 2004). Diabetes may also result from mitochondrial disorders, of which the underlying mechanism of diabetes is unknown, but thought to be due to reduced insulin secretion (Barrett 2001). Also rare monogenic forms of diabetes can provide information (Porter and Barrett 2005). When a causative gene is identified, its contribution to genetic heterogeneity of more common types of diabetes, as well as functional effects of the relevant proteins, can be evaluated.

A few syndromes of relevance in T1D are discussed further in the following section.

6.3.1. Down's syndrome

In our own studies we have attempted to use comparative genetics by several means. Down's syndrome may serve as an example of how knowledge and studies regarding a quite common syndrome, only occasionally including T1D, may be used to extract information about T1D in general.

Down's syndrome (trisomy 21) [MIM 190685] is characterized by mental retardation and characteristic dysmorphic features, believed to be attributable to genes on chromosome 21, likely in a complex and interactive manner (Tsukahara et al. 1996; Gitton et al. 2002; Reymond et al. 2002; Kahlem et al. 2004; Patterson and Costa 2005). Down syndrome is characterized by an increased prevalence of a number of autoimmune diseases (Shield et al. 1999), and an increased prevalence of diabetes in Down's syndrome was a dogma, described in textbooks (Pickup and Williams 1991), based on a few old, biased and not population-based studies (Milunsky and Neurath 1968; Jeremiah et al. 1973; Van-Goor et al.; Anwar et al. 1998). Additionally, an earlier peak age of onset of diabetes had been suggested previously (Burch and Milunsky 1969). In none of the previous studies T1D was, however, distinguished from T2D. Pathogenesis and mechanisms behind diabetes in Down's syndrome are unclear, but signs of premature aging as well as life-style factors may bias the use of a simple age criteria for diagnosis, as was used in some previous studies. Increased autoimmunity in general in Down's syndrome has been proposed, and especially thyroid and coeliac disease have increased prevalences, including demonstration of auto-antibodies (Ivarsson et al. 1997; Karlsson et al. 1998; Agardh et al. 2002). In a study of nine Down's syndrome patients (age of patients unknown) with slightly impaired glucose tolerance, it was, however, not possible to detect insulin auto-antibodies, as an indication of autoimmune damage of pancreatic beta cells (Serrano-Rios et al. 1973). In a more recent study co-occurrence of more than two islet auto-antibodies was, however, demonstrated to be significantly more common in Down's syndrome individuals as compared to controls (Gillespie et al. 2006).

In Denmark, due to complete registries, unique opportunities exist for population based studies, and we therefore undertook an epidemiological study aiming at defining the prevalence of Down's syndrome in T1D. In this study, we demonstrated a prevalence of Down's syndrome in T1D of 3.8‰, which corresponds to a 4.2 times increased prevalence of Down's syndrome in T1D in Denmark, compared to the background population (Bergholdt et al. 2006). In line with some of the previous studies addressing diabetes in Down's syndrome, we also demonstrated a more than four times increased prevalence of T1D in Down's syndrome. Our study was the first proper demonstration of an increased prevalence of Down's syndrome in verified T1D, it was nation-wide population based,

used validated and complete registers, as well as confirmed diagnoses of T1D. Median age at diagnosis of T1D in Down's syndrome was 6 years in our study, lower than the general median age at onset for childhood T1D of 8 years, which might suggest a greater "genetic load", perhaps attributable to genes on chromosome 21.

To determine whether T1D in Down's syndrome is immune-mediated, also immuno-genetic factors should be evaluated, and first of all, estimation of the significance of the HLA region, in Down's syndrome individuals with T1D, is necessary. This has recently been addressed in two reports, both, however, limited by the small number of individuals with Down's syndrome and T1D (Bergholdt et al. 2006; Gillespie et al. 2006). We attempted to evaluate the significance of the HLA region, in Down's syndrome individuals with T1D. HLA-*DQB1* genotypes were evaluated, where this was possible. Although numbers were small (eight cases of T1D and Down's syndrome, of which HLA genotyping was possible in six), we demonstrated that one carried a dominantly protective allele, one only neutral alleles, three carried just one classical risk allele, and just one carried two classical T1D risk alleles (Bergholdt et al. 2006). In line with our findings, a newer report, comprising 40 individuals with Down's syndrome and T1D, also suggest that individuals with Down's syndrome generally do carry more high risk HLA class 2 genotypes than controls, but also that individuals with Down's syndrome and T1D carry less of the highest risk HLA class 2 genotypes when compared to T1D individuals in general (Gillespie et al. 2006). This might suggest that other genetic risk loci than the HLA region, as genes on chromosome 21, could be important in this group. It is tempting to speculate that the explanation for the increased prevalence, or at least part of the explanation, might be due to the trisomy 21, supporting the possible existence of T1D associated genes on this chromosome. The simplest mechanism would be that the extra chromosome 21 results in a linearly proportional increase of all gene products controlled by genes on chromosome 21. This might then produce an imbalance between gene products determined by chromosome 21 and those determined by genes on other chromosomes. Unfortunately, often no direct correlation between gene dosage and the amount or activity of gene products exists (Olson et al. 2004). Mechanisms by which increased dosage causes any specific Down's syndrome feature have not been established, despite extensive studies of Down's syndrome mouse models (Olson et al. 2004). However, a recent report indicated a ~1.5 fold over-expression of most of the genes in the trisomic region of the Ts1Cje Down's syndrome mouse model (Amano et al. 2004). A new promising aneuploid mouse model, Tc1, which carries an almost complete copy of human chromosome 21, have just been reported (O'Doherty et al. 2005). This Tc1 mouse has several characteristics of Down's syndrome, and is very promising for studies of Down's syndrome. A major concern, however, is that it is a mosaic, implying that not all cells are trisomic (O'Doherty et al. 2005). It is believed that some abnormalities of Down's syndrome are due to over-expression at specific loci, resulting from the presence of two identical copies of a susceptibility allele inherited from the parent of origin of trisomy (disomic homozygosity) (Feingold et al. 1995). So by inheriting two identical copies of a rare susceptibility allele, the suggestion is that an individual with Down's syndrome exceeds a liability threshold and manifests the disorder (Feingold et al. 1995). In Down's syndrome there is also an increased prevalence of a number of other autoimmune diseases, but to date no gene or chromosomal region have been able to explain this predilection for autoimmune disease (Shield et al. 1999). Our observation of a more than four times increased prevalence of Down's syndrome in T1D, and a lower median age at onset of T1D in Down's syndrome compared to other T1D individuals, substantiate the possible significance of genes on chromosome 21. Such genes may be important for the quite rare combination of T1D and Down's syndrome, but may also be involved in T1D pathogenesis in general. This represented the first population based study of Down syndrome in T1D. Further studies of auto-antibodies and the contribution of HLA and other genetic regions in T1D in Down's syndrome are clearly needed, including a larger number of well-defined T1D cases with Down's syndrome, to clarify diabetes in Down's syndrome. Also other molecular approaches, as global assessment of the transcriptome and proteome in such individuals, might be helpful new approaches. In addition to gene and protein expression patterns also temporal and spatial expression patterns of such may be important. When using a global approach also alterations in gene expression of genes on other chromosomes than chromosome 21, as well as of biological systems and pathways could be elucidated.

The distal part of chromosome 21 is known to harbor a few established candidate genes for autoimmunity, including T1D. One is the *AIRE* gene at 21q22.3, encoding a regulator of transcription, in which mutations have been shown to give rise to the disease APECED (Autoimmune Poly-Endocrinopathy-Candidiasis-Ectodermal Dystrophy syndrome) [MIM #240300] (Perheentupa 1996; Scott et al. 1998; Bjorses et al. 2000). APECED is a rare systemic autoimmune disorder of monogenic and autosomal-recessive inheritance (Meyer and Badenhoop 2002), where T1D is part of the syndrome in 20% of cases (Meyer and Badenhoop 2002). Almost 30 APECED causing mutations have been identified, of which two, prevails in Europe (Pearce et al. 1998; Meyer et al. 2001). The *AIRE* gene was originally considered a candidate gene of the linked region, and we tested these two prevailing mutations causing the APECED syndrome in Caucasoid populations, but no mutation carriers were identified. In addition, following our fine mapping, the *AIRE* gene was no longer localized in the linked region but distal to this, and we did therefore not continue with re-sequencing and further examination of the gene. Others have equally not been able to identify any of these two mutations in T1D populations (Nithiyananthan et al. 2000; Meyer et al. 2001; Meyer and Badenhoop 2002). Furthermore, disomic homozygocity at the APECED locus has been proven not to be able to explain autoimmunity in Down's Syndrome (Shield et al. 1999).

Another example of our use of comparative genetics in relation to chromosome 21 genes concerns the *CBR1* gene. By proteome-analysis of interleukin-1β (IL-1β) exposed diabetes-prone Bio Breeding (BB-DP) rat islets of Langerhans in vitro, the protein carbonyl reductase 1 (CBR1), encoded by the *CBR1* gene, was demonstrated by our group to be down-regulated five-fold by IL-1β exposure (Sparre et al. 2002). Since IL-1β is believed to be an important cytokine involved in destruction of the beta cells (Nerup et al. 1994), this finding suggested a role for *CBR1* in cytokine-mediated beta cell destruction and thereby in T1D pathogenesis. Analysis of seven cSNPs and two SNPs in the 3'UTR of the CBR1 gene did not demonstrate any significant association (Bergholdt et al. 2005a). However, in the "linked" family subgroup analysis, a significant association to T1D was demonstrated.

Another chromosome 21 gene, the Cu/Zn superoxide dismutase 1 (*SOD1*) is a potent antioxidant, and has been implicated in genetic predisposition to alloxan (AL) –induced diabetes in mice (Mathews and Leiter 1999; Mathews et al. 2002). Alloxan (AL) and streptozotocin (STZ) are pancreatic beta cell selective toxins that have been used to probe the mechanisms underlying oxygen-mediated damage to rodent beta cells. Pretreatment with or concomittant administration of *SOD1* confer protection against AL or STZ induced diabetes (Robbins et al. 1980; Oberley 1988), and cultured rodent islets can be protected from AL or STZ induced impairment (Fischer and Hamburger 1980; Oberley 1988). Specific activity of pancreatic *SOD1* have been shown to be significantly increased in AL-resistant (ALR/Lt) compared with AL-susceptible (ALS/Lt) mice (Mathews and Leiter 1999), and *SOD1* transgenic mice are protected against oxidative stress (Eizirik and Sandler 1989; Kubisch et al. 1997). We therefore not only genotyped cSNPs, but also known SNPs in the 3'UTR, of which none demonstrated association to T1D (Bergholdt et al. 2005a).

The Runt-related transcription factor 1 (*RUNX1*), also called AML1, is another candidate gene for autoimmunity, including T1D, located in the linked region on chromosome 21. Indications that *RUNX1* might be a common factor in autoimmunity have recently come from reports in three other autoimmune diseases, systemic lupus erythematosus, psoriasis and rheumatoid arthritis, in which disease-associated mutations in *RUNX1* binding sites of three different genes was demonstrated (Prokunina et al. 2002; Helms et al. 2003; Tokuhiro et al. 2003). Further, the association of the regulatory *PDCD1* variation (in a *RUNX1* binding site), originally found with SLE (Prokunina et al. 2002), has been replicated in T1D (Nielsen et al. 2003). In rheumatoid arthritis an intron 6 SNP in the *RUNX1* gene itself furthermore showed disease association in a case-control study (Tokuhiro et al. 2003). These reports make the *RUNX1* gene interesting in T1D, supported by the location of the linkage peak we reported, in exactly this region. RUNX1 protein is a DNA-binding transcription factor with the context-specific capability of activating or repressing gene expression (Hug et al. 2004). We were, however, not able to demonstrate association to T1D neither of the known coding SNPs in this gene, nor of the intronic SNP in intron 6 (Bergholdt et al. 2005a), which was found associated to rheumatoid arthritis (Tokuhiro et al. 2003). However, T1D association of borderline significance was observed in the "linked" family subgroup analysis. In addition, data from our group have demonstrated that the mRNA expression of *AML1* (*RUNX1*) was significantly downregulated in a rat insulin producing cell line after exposure to IL-1 using Affymetrix chip analysis (Nielsen et al. 2004), also suggesting that *RUNX1* regulated gene expression may be relevant in T1D pathogenesis.

The observation of a more than four times increased prevalence of Down's syndrome in T1D, and a lower median age at onset of T1D in Down's syndrome compared to other T1D individuals, substantiate the possible significance of genes on chromosome 21. Such genes may not only be important for the quite rare combination of T1D and Down's syndrome, but may also be involved in T1D pathogenesis in general. Further studies of auto-antibodies and the contribution of HLA and other genetic regions in T1D in Down's syndrome are clearly needed. Larger numbers of well-defined T1D cases with Down's syndrome is necessary to clarify T1D in Down's syndrome, and these should preferentially be collected in a collaborative effort by several countries in order to obtain sufficient numbers. Also other new molecular approaches, as global assessment of the transcriptome and proteome in such individuals, might be helpful in elucidating underlying biological mechanisms. In addition to gene and protein expression patterns also temporal and spatial expression patterns of such may be important. The increased frequency of T1D among Down's syndrome individuals should be kept in mind when caring for people with Down's syndrome. As the life expectancy of a 1 year old person with Down's syndrome is now more than 60 years and likely to improve (Bittles and Glasson 2004), T2D is expected, just for this reason, to become an increased "problem". Additionally, the prevalence of verified T2D in Down's syndrome should therefore be evaluated, as should environmental and genetic factors underlying T2D in Down's syndrome individuals. Focus should be on life-style intervention, diet and the importance of physical activity, also as potential methods for prevention of T2D in this group. Clinical care of Down's syndrome individuals with diabetes is a specialist job, with obvious difficulties. Prolonged and maybe even continued intensive education in self-management is necessary. In only a subgroup of people with Down's syndrome diabetes self-management are obtainable, in other cases collaboration with parents or people caring for these individuals are essential. Education and level of self-management, furthermore has to be adjusted in an individualized way. Little is known about the demanding specialized treatment and care of diabetes in Down's syndrome, and further research into this field is warranted.

In conclusion, T1D has an increased frequency in Down's syndrome individuals, which should be kept in mind. The frequency of T2D is unknown, but presumed to be a problem as well. Research dealing with underlying biological mechanisms, genes on chromosome 21 as well as other genes, prevention, clinical care and management/potential self-management of diabetes in people with Down's syndrome is clearly needed.

6.3.2. Wolfram syndrome

Wolfram syndrome (MIM 222300) is a rare autosomal recessive disorder (Wolfram and Wagener 1938; Barrett et al. 1995). The disorder is also referred to as DIDMOAD (diabetes insipidus, diabetes mellitus, optic atrophy and deafness). Insulin-dependent diabetes and bilateral progressive optic atrophy are necessary for diagnosing the syndrome (Fuqua 2000). The clinical picture is, however, variable, and may additionally include neurological symptoms (Rando et al. 1992) and psychiatric illness (Strom et al. 1998). Linkage with Wolfram syndrome has been reported for a chromosome 4p16.1 region (Polymeropoulos et al. 1994; Collier et al. 1996), containing the *WFS1* gene, in which mutations were identified in Wolfram syndrome families (Inoue et al. 1998; Strom et al. 1998; Hardy et al. 1999; Smith et al. 2004; Hansen et al. 2005). Diabetes seen in Wolfram syndrome is distinctive from T1D. It results from a selective islet beta cell loss, but with no evidence of autoimmunity (Karasik et al. 1989; Ishihara et al. 2004), suggesting that the *WFS1* gene could be generally important in formation or survival of beta cells. It has been suggested that mutations in *WFS1* may play a minor role in more common forms of diabetes (Inoue et al. 1998). *WFS1* mutations have been tested for association with different diseases and specific genotype/phenotype correlations. No association of mutations in coding regions of the *WFS1* gene could be demonstrated with T1D in our Danish family material (Larsen et al. 2004), whereas a Japanese study demonstrated association of one coding SNP in a small T1D case-control design (Awata et al. 2000). Also in T2D association of a two SNP-haplotype was demonstrated with *WFS1* in a case-control study (Minton et al. 2002), but further studies in T1D and T2D are needed to clarify the eventual contribution of the *WFS1* gene. Wolfram syndrome-associated diabetes is generally described as non-immune-mediated, however, one recent Japanese case report, demonstrated a Wolfram patient with a novel mutation in *WFS1*, including diabetes-related (GAD (glutamic acid decarboxylase) and IA-2 (insulinoma-associated antigen-2)) autoantibody positivity (Nakamura et al. 2006). The *WFS1* region on chromosome 4p16 corresponds to a region previously linked to T1D (Nerup et al. 2001; Larsen et al. 2004), however not confirmed in the recent combined genome scan (Concannon et al. 2005). Identification of subgroups of diabetic patients with specific genetic causes may be important in shedding light on genetics and mechanisms, also of putative importance for the pathogenesis of T1D in general. The exact pathogenesis of Wolfram syndrome-associated diabetes, however, remains unknown (Fuqua 2000).

6.3.3. Wollcot-Rallison syndrome

Wollcot-Rallison syndrome (OMIM 226980) is a rare autosomal recessive condition characterized by neonatal or very early-onset diabetes, epiphyseal dysplasia, renal impairment, acute hepatic failure and developmental delay (Wolcott and Rallison 1972). Diabetes associated with this syndrome is characterized by beta cell loss, leading to insulin deficiency, without autoimmune pathology (Wolcott and Rallison 1972; Delepine et al. 2000). The condition was in 2000 mapped to the chromosome 2p12 locus (Delepine et al. 2000), containing the eukaryotic initiation factor 2α kinase 3 gene, *EIF2AK3* (also known as pancreatic endoplasmic reticulum kinase, *PERK*). Mutations in *EIF2AK3* have been demonstrated associated with the syndrome (Delepine et al. 2000; Iyer et al. 2004; Senee et al. 2004; Porter and Barrett 2005; Durocher et al. 2006). *EIF2AK3* regulates protein translation and is highly expressed in pancreatic islets and plays a role in maintaining integrity of the pancreatic islets. The syn-

drome is extremely rare with only approximately 30 cases described world-wide (Delepine et al. 2000; Senee et al. 2004; Durocher et al. 2006). Findings have not indicated *EIF2AK3* as a common diabetes susceptibility gene in diabetes in general, it was not associated to T2D (Vaxillaire et al. 2001), whereas in T1D suggestive evidence of linkage to the region was observed in Scandinavian families (Nerup et al. 2001). Furthermore a South Indian study demonstrated association to T1D of two microsatellites in the gene region, but did not demonstrate any common polymorphisms in the *EIF2AK3* gene (Allotey et al. 2004). The role of this gene in T1D is therefore not clear either, and the potential use of genes underlying rare syndromes involving diabetes remains to be further elucidated.

6.4. ANIMAL MODELS

In animal models of autoimmune disease, the general disease phenotype can be broken down into individual loci having discrete subphenotypes, which is not possible in human studies. Such sub-phenotyping at specific loci can potentially provide clues to the functional significance of candidate genes at each locus, and may provide useful biological information in the search for candidate genes at specific human T1D loci. Transgene and knock-out studies in animal models allow studies of specific effects of genes either missing or over-expressed.

In T1D especially two animal models of human T1D have been successfully used, the non-obese diabetic mouse (NOD) and the bio-breeding rat (BB). Much insight into T1D pathogenesis and genetics, from syntenic regions, has come from studying these two animal models. A thorough review of the models and results obtained by use of them, is, however, considered outside the scope of this thesis. Some recent papers reviewing their use in T1D are (Todd and Wicker 2001; Lang and Bellgrau 2004; Leiter and von-Herrath 2004; Mordes et al. 2004; Mathews 2005; Melanitou 2005; Wicker et al. 2005; Yang and Santamaria 2006).

Another spontaneous rat model of T1D, the Komeda diabetes-prone (KDP) rat, is characterized by autoimmune destruction of beta cells, rapid onset of diabetes with no sex difference and no significant T-cell lymphopenia (Yokoi et al. 2002). Lymphocyte infiltration is also seen in other tissues as the thyroid gland, kidney, adrenal gland and pituitary, indicating general autoimmunity (Yokoi et al. 2002). Knowledge on genetics in the KDP rat was used in our own studies of the *CBLB* gene (Bergholdt et al. 2005c). In this rat model the major part of its genetic susceptibility for autoimmunity was shown to be accounted for by two loci, the MHC region on rat chromosome 20 and another locus on rat chromosome 11. This latter locus was demonstrated to contain the Casitas-B-lineage lymphoma b (*Cblb*) gene (Yokoi et al. 1997; Yokoi et al. 2002). A nonsense mutation in this gene in the KDP rat truncates the protein and transgenic rescue studies confirmed that that this mutation was pathogenetic and specific to the rat (Yokoi et al. 2002). Furthermore, functional studies of the protein, Cbl-b (an ubiquitin-protein ligase) demonstrated by use of knock-out mice, an important role in T cell co-stimulation, by acting as a negative regulator of autoimmunity (Bachmaier et al. 2000; Chiang et al. 2000). In Cblb-deficient mice, infiltration of lymphocytes in different endocrine tissue, was shown to be caused by enhanced T cell activation (Bachmaier et al. 2000; Yokoi et al. 2002). Based on this role of Cbl-b in regulation of autoimmunity, it was hypothesized that Cbl-b also could be dysregulated in humans and may contribute to autoimmune diseases in humans. In T1D an increased frequency of other autoimmune diseases is also often seen (Payami et al. 1989; Chikuba et al. 1992; Tait et al. 2004), and the *CBLB* gene was therefore chosen as a candidate gene for T1D and autoimmunity in general. We re-sequenced coding regions of the gene and demonstrated association to T1D of an exon 12 polymorphism, as described in chapter 5. The variant was rare (MAF = 0.024) and transmissions therefore few, underlining the importance of using large materials and the need for replication in other samples/populations.

In our study of the *CBLB* gene we also performed functional evaluation of the two known isoforms of the molecule. Expression levels of the two isoforms (one lacking exon 16) on the mRNA level, was evaluated constitutively and after cytokine stimulation. We demonstrated significant up-regulation, although not disease-specific, of the isoform lacking exon 16 upon cytokine exposure (Bergholdt et al. 2005c), which may suggest a possible role for this isoform in T1D pathogenesis (and/or other autoimmune diseases). No SNPs were detected in exon 16 or splice sites flanking this exon and no known functional elements of the molecule are encoded by exon 16. A mechanism for the suggested effect of cytokine stimulation is not known, and the involvement of several isoforms of *CBLB* also remains unclear. Due to the KDP phenotype, including several autoimmune diseases, we hypothesized that individuals with more than one autoimmune disease, may be more likely to harbor polymorphisms in *CBLB*. We therefore used individuals with more than one autoimmune disease for re-sequencing. Additionally, another comparative approach was used, since gene-gene interaction of *CBLB* with the *CTLA4* gene was evaluated. The *CTLA4* gene, involved in the same biological pathway of T cell regulation as *CBLB*, is an established candidate gene for many autoimmune diseases, e.g. Graves' disease, autoimmune hypothyroidism and T1D. Interaction between the two genes was evaluated by stratifying data from the *CBLB* SNP showing T1D association according to high risk genotype vs. lower risk genotypes of the established *CTLA4* marker, *CT60* (Ueda et al. 2003). An increase in transmission distortion was observed (Bergholdt et al. 2005c), and although based on few transmission, this may suggest interaction and supports genetic control of T cell regulation in autoimmunity.

7. NOVEL ANALYSIS APPROACHES FOR IDENTIFYING GENES

Identification of susceptibility genes and interaction among genes in complex genetic diseases poses many challenging problems. Traditionally, genetic linkage analysis is performed by searching for the marginal effect of a single putative trait locus. Methods for searching for multiple trait loci simultaneously are, however, limited. Most methods used for searching for interaction, focuses on two-locus traits by stratifying on evidence for linkage at one selected locus while searching for another (Buhler et al. 1997; Farrall 1997; Cox et al. 1999; Cordell et al. 2000; van Heel et al. 2003). Recently, the use of artificial neural networks has been applied to address the complexity stemming from interacting gene loci (Lucek et al. 1998; Curtis et al. 2001; Marinov and Weeks 2001; Ritchie et al. 2003a; Pociot et al. 2004).

7.1. EPISTASIS

Individual effects of minor susceptibility genes is, even in very large samples, difficult to demonstrate and may depend also on genetic variance at other loci (epistasis). Epistasis or gene-gene interaction is the concept of two or more loci or allele pairs, which combine interdependently to increase the susceptibility to a disease. Interactions between genes may contribute to the variation in complex traits, e.g. disease susceptibility (Cordell 2002; Carlborg and Haley 2004). Studies in model organisms suggest that epistasis may play an important role in the etiology of multi-factorial diseases and complex traits in humans. There is no consensus as to the best strategy for detecting epistatic interactions in humans (Cordell 2002; Carlborg and Haley 2004). A number of recent studies in humans and animals have identified loci that interact significantly but contribute little or no effect individually (Kuida and Beier 2000; Kim et al. 2001b; Carlborg et al. 2003). In T1D, attempts to show epistasis between classical T1D loci in humans, as well as animal models, have only provided very few examples (Cordell et al. 1995; Cordell et al. 2000; Cordell et al. 2001). Hence, new methods are in demand for detecting and characterizing epistasis, as well as for elucidating underlying biological interactions (Cordell 2002; Carlborg and Haley 2004).

Most gene products mediate their function within complex networks of interconnected proteins (Cusick et al. 2005). Therefore understanding effects of genetic interaction either by classical linkage approaches (Buhler et al. 1997; Farrall 1997; Cox et al. 1999; Cordell et al. 2000; van Heel et al. 2003) and/or novel analytical approaches is not simple. The presence of epistasis is important when investigating the genetic mechanisms underlying a complex disease because it implies that the effect of one locus on the development of a disease can be increased or decreased by effects at another locus, which is likely to result in a reduction in the power to detect the first locus (Cordell 2002). Interaction between different genetic loci, and probably also gene-environment interactions, may contribute to the complexity of the genetics of T1D. Interactions between susceptibility loci for T1D have been demonstrated in several studies (Davies et al. 1994; Cordell et al. 1995; Mein et al. 1998; Nerup et al. 2001; Concannon et al. 2005). In a large study of disease susceptibility and gene-gene interactions in a multiple sclerosis case-control data set, variation in 34 selected genes were evaluated (Brassat et al. 2006). A multifactor dimensionality reduction test (Ritchie et al. 2003b) was used to demonstrate one single variant, as well as a three locus combination showing association to MS, emphasizing the importance of looking at main effects as well as interactive effects in complex diseases (Brassat et al. 2006). Establishing the biological importance of interactions identified statistically, is, however, difficult, but important.

A conditional approach may be applied to genome scan data. A scan for epistasis could involve searching for modest individual effects, and then querying for interactions among such set of positive markers. Scans conditional on known positive results may be more powerful, as the variance explained by the major loci has been controlled for, thereby enhancing the signal from minor contributors. This indicates that initial screens for main effects also in genome-wide association studies are likely to be successful in many cases, without considering epistasis in the initial analysis, but only in subsequent analyses.

7.2. DATA MINING

Data mining technologies (based on pattern recognition) applied to genome wide linkage data might be a valuable tool, since these methods are able to detect complex interactions, not easily detectable by other methods. Data mining technologies can be used to identify marginal markers and combinations of marker-marker interactions predictive for disease susceptibility or protection. Importantly, looking at several interacting markers at a time may improve the ability to differentiate between patients and controls. Data mining and neural network-based methods have recently proven valuable in T1D (Pociot et al. 2004).

Genetic factors influence both susceptibility and resistance to disease, and data mining methods do allow evaluation of potentially protective effects (Pociot et al. 2004). This is not possible in classical linkage analyses, implying that for elucidation of such effects, novel analysis tools are necessary.

In our studies we have applied a data mining approach, using decision trees, for the analysis of genome scan data (Bergholdt et al. 2007b). Originally, the developed methods were based on decision tree construction and artificial neural networks (Pociot et al. 2004) applied to T1D genome scan data (genotyping data) from the Scandinavian genome scan (Nerup et al. 2001). We wanted to test whether the analytical methods previously developed, could use IBD sharing data, implying a reduced amount of information, as compared to real genotyping data used in (Pociot et al. 2004). We were given access to genome scan data from the Type 1 Diabetes Genetics Consortium (www.t1dgc.org) on 392 Scandinavian T1D multiplex families (411 affected sib-pairs) and 763 US/UK multiplex families (910 affected sib-pairs), which were part of the recent combined and largest T1D genome scan (Concannon et al. 2005). Data were recoded to IBD status and all microsatellite markers genotyped in both materials were analyzed.

We were able to detect and confirm the major loci also identified by the genome scan (Concannon et al. 2005): 6p21 (the *HLA* region), 2q31-33 (the *CTLA4* region), 16p12-q11.1, 11p15.5 (the Insulin gene region), 16q22-q24 and 10p14-q11 (Bergholdt et al. 2007b), confirming the validity of the approach. Markers corresponding to the HLA region showed predictive signals for T1D with significance levels several orders of magnitude higher than the other markers. Interestingly, also two markers corresponding to the previously identified *IDDM15* region (Delepine et al. 1997) on chromosome 6q21, were identified (Bergholdt et al. 2007b). In linkage studies identification of markers for this region generally requires special analyses, in order to separate their effect from the strong effect of the MHC region (Delepine et al. 1997; Concannon et al. 2005). A few new regions were also suggested, *D17S798*, *D2S125*, *D9S175*, and *D8S261* (ref 8). In the complete data set, also the *D4S403* was identified as of predictive value. This marker and the region surrounding it was previously demonstrated linked to T1D (Nerup et al. 2001; Larsen et al. 2004) and contains the *WFS1* gene associated with Wolfram syndrome, chapter 6.3.2. In the Scandinavian data sub-set a marker corresponding to the linkage region on chromosome 21 (Bergholdt et al. 2005a), was furthermore identified providing significant predictive value for T1D (Bergholdt et al. 2007b).

Further we evaluated potential interactions between markers by use of the Enterprise Miner tree model in SAS. We used the highest ranked marginal markers to search for interaction with other markers (on other chromosomes) and the resulting combinations were tested for association to T1D. Combinations of markers, with significant predictive or protective value for T1D, were identified for the *TNFA* marker (a marker for the HLA region) and the *D17S798* marker, interacting with markers on other chromosomes (ref. 8).

Interestingly, we demonstrated several rules of statistically significant interaction. One marker combination conferred protective value towards T1D (statistically more frequent in non-affected sib-pairs than affected sib-pairs), i.e. *TNFA – D11S910*. All other combinations (*TNFA-D4S403*, *TNFA-D2S177*, *TNFA-D1S229*, *TNFA-D13S170*, *TNFA-D16S287*, *D17S798-D2P25*, *D17S798-D5S429* and *D17S798-D1S197*) conferred increased susceptibility towards T1D as compared to each of the markers alone.

To evaluate a potential biological significance of the predicted rules of genetic interaction, we constructed protein-protein interaction networks, using proteins encoded by genetic regions 5 Mb on each side of the interacting micro-satellite markers. We used a high confidence, quality controlled system integrating and validating interactions from multiple databases and species and were able to demonstrate support for functionality for all predicted rules of genetic interaction (Bergholdt et al. 2007b). These networks predict, with high stringency and confidence criteria, the most likely candidate genes responsible for the interactions predicted, as well as the biological context in which they are involved in (Bergholdt et al. 2007b). The protein-protein networks and their value in selecting functional candidate genes are further discussed in chapter 8.2.

In conclusion, novel analytical methods, based on data mining and decision trees can confirm observations known from “classical” analyses and provide new information about genetic interactions. Genetic interaction as the basis for pathogenesis is supported for T1D, and our findings certainly suggest these novel methods as valuable in searching for yet unidentified genetic and functional interactions involved in pathogenetic processes. Strong functional support to the value of the predicted interactions has come from our demonstration of protein interactions corresponding to the genetic interactions, as described in chapter 8.2.

8. APPROACHES FOR FUNCTIONAL IDENTIFICATION OF CANDIDATE GENES

Classical functional evaluation of identified genetic variation is essential to verify the significance of a genetic variation and to under-

stand its eventual biological significance, and is usually performed in the end of a study. Such studies include evaluation of expressed levels of mRNA or protein in different genotype groups, reporter gene assays to evaluate promoter activity, as well as gene overexpression/knock-out studies. We have evaluated identified variants functionally in our studies of the *IL1RI* gene (Bergholdt et al. 2000), the *IL12B* gene (Bergholdt et al. 2004a) and the *CBLB* gene. Identified (Bergholdt et al. 2000; Bergholdt et al. 2005c) or selected (Bergholdt et al. 2004a) variants demonstrating T1D association were tested for functional significance by comparing mRNA and/or protein levels among different genotype groups, in search of allele dosage effects (Bergholdt et al. 2000; Bergholdt et al. 2004a; Bergholdt et al. 2005c). Although numbers were small and biological background variation/noise between individuals may be significant, we were able to demonstrate significant differences among genotype groups, indicating allele dosage effects. Functional studies have not yet been extended further in search for the biological mechanisms, explaining these associations. Functional studies may act as supplementary studies, as well as an alternative to e.g. large replication studies.

A thorough description of classical methods for evaluating functional significance of identified genetic variants is outside the scope of this thesis, which focuses on approaches for selection of candidate genes/gene mapping. This is also the case for other kinds of regulation, i.e. epigenetic mechanisms, which may influence functionality, but is not considered approaches for gene mapping. Epigenetic phenomena as maternal and paternal imprinting, expression regulation by micro-RNA, methylation and histone modification may complicate evaluation of gene expression, and will only be discussed here as factors complicating the analysis/interpretation of data.

Epigenetics is defined as biochemical mechanisms altering inherited information without changing the primary nucleotide sequence, i.e. changes to DNA other than mutations. These include covalent modification of nucleotides, e.g. methylation, which is the most common DNA modification in eukaryotes, and histone modification, which alter the binding affinity of proteins that mediate transcriptional acitivity (Richards 2006). Epigenetic modifications at the DNA, nucleosomal or chromosomal level, may affect gene expression and ultimately phenotypes (Richards 2006). The molecular pathways initiating the different epigenetic states on DNA sequences, have been a recent area of focus. Studies have demonstrated that small RNA molecules, i.e. RNAi (RNA interference molecules) can modify methylation or histone marks associated with transcriptional repression in particular genomic regions (Richards 2006).

MicroRNAs (miRNAs) have recently emerged as important regulators of gene expression controlling central biological processes. These small (approximately 22 nucleotide long RNA molecules) induce translational suppression, via interaction with target sites in the 3'UTR of genes (Carthew 2006; Ying et al. 2006). Direct roles in developmental processes have been described in a variety of species, and cellular processes as proliferation, morphogenesis, apoptosis and differentiation seem to be possibly regulated by miRNAs (Carthew 2006). Prediction of miRNA target sites is still controversial and not simple (John et al. 2004; Griffiths-Jones 2006; John et al. 2006). miRNA may play an important role in the development of T1D (Poy et al. 2004). We have attempted to evaluate the existence of micro-RNA binding sites in the chromosome 21 genes from the linked region, which could point to micro-RNA as a regulatory mechanism for the protein synthesis of the corresponding genes. Candidate genes in the linked region on chromosome 21 containing predicted micro-RNA binding sites in their 3'UTR region were: *SON, RUNX1, SOD1, DSCR1, TIAM1, HLCS* and *HUNK* (http://www.genome.ucsc.edu/).

Another example of an epigenetic phenomenon is parental imprinting, which is a parent-of- origin specific gene expression, whereby a single allele is differentially expressed depending on the sex of the parent transmitting the allele. Futhermore, X-chromosome inactivation, a sex-chromosome dosage-compensation mechanism in mammals that leads to transcriptional silencing of a large percentage of genes on one X-chromosome in females, is an epigenetic phenomenon. The significance of such molecular epigenetic mechanisms is, however, still unknown (Richards 2006), as is the relevance in T1D.

8.1. GENE EXPRESSION – TRANSCRIPTIONAL PROFILING

Gene expression involves the presence of one or more transcription factor that, by binding to a promoter, leads to production/increased production of a protein or its suppression, and gene expression is thoroughly regulated. Knowledge regarding differences in gene expression, under different conditions or in different tissues, may be useful in mapping genes. Expression levels of genes may be seen as intermediate phenotypes between genomic DNA sequence variations and more complex disease phenotypes (Cheung et al. 2003); and variation in gene expression has been shown in large part to be due to polymorphisms in DNA sequence (Cheung et al. 2003; Morley et al. 2004). Genes responsible for variation in gene expression have been suggested to have an impact on phenotype (Morley et al. 2004). High heritability of gene expression has previously been demonstrated (Cheung et al. 2003) and suggests that identification of genetic determinants of expression may provide insights into the molecular basis of complex traits, like T1D. Common expression patterns of genes may furthermore reflect shared regulatory mechanisms.

Efforts have been made to identify and prioritize positional candidate genes based on their expression (Franke et al. 2004; Franke et al. 2006). It was expected that differences in expression e.g. among cases and controls, would be due to cis-acting variants in genes, but surprisingly and importantly, it was demonstrated that for most genes, differences in gene expression were determined by variants in genes located elsewhere (trans-acting) (Schadt et al. 2003; Morley et al. 2004).

Recently, a number of studies have used approaches in which linkage mapping was combined with the identification of co-regulated genes using microarrays. These studies have enabled the discovery of trans-acting expression QTLs (Morley et al. 2004; Deutsch et al. 2005; Li and Burmeister 2005; Stranger et al. 2005). This may be a promising approach also for identifying interacting susceptibility genes in multi-factorial diseases. In expression QTL analyses in general, large numbers of individuals are genotyped for a panel of polymorphic markers and simultaneously phenotyped with microarray expression profiling, followed by statistical evaluation of linkage between markers and gene expression levels (Beyer et al. 2007). Data for such genetical genomics studies in T1D are, however, still limited. Use of microarray technology is instrumental for these studies and the use of microarrays has changed molecular biology from studies of individual biological functions of a few related genes or proteins towards a more global investigation of cellular processes, shown to be complex and interrelated in biological systems (Allison et al. 2006; Hoheisel 2006). Microarrays are mainly used for transcriptional profiling and analysis of transcripts of genes, but can also be used for genotyping, which are now also done routinely. Other approaches are also possible, e.g. genome-wide epigenetic analysis (Hoheisel 2006) and approaches evaluating dynamic traits, developmental issues, including a time perspective (time-dependent genetic effects), and more functional mapping of QTLs (Wu and Lin 2006).

In our own studies we suggested, that by looking for genes that do change expression level after cytokine exposure in human islets and in a rat beta cell line, it would be possible to select or prioritize genes with functional significance in T1D pathogenesis. Gene expression may reflect an intermediate phenotype. This can be used in detection of disease gene variants, if the presence of DNA variation leads to some functional consequence, e.g. represented by altered gene expression. Ultimately, DNA variation in one of the candidate genes may be shown to be associated with a disease associated phenotype,

such as the intermediate RNA expression level. Differences in expression levels may have other causes than genetic variation, which should also be explored. However, the current sample collection size of 8 human pancreatic islet preparations makes such analysis impossible. It will not be possible to study enough human islet preparations to establish significant correlations between gene expression and variations. We focused on an approach to pin-point potential functional candidate genes for further studies and have previously genotyped all coding SNPs in all the examined chromosome 21 genes (Bergholdt et al. 2005a). In the future, it would probably be more relevant and important to search for potential regulatory SNPs in these genes.

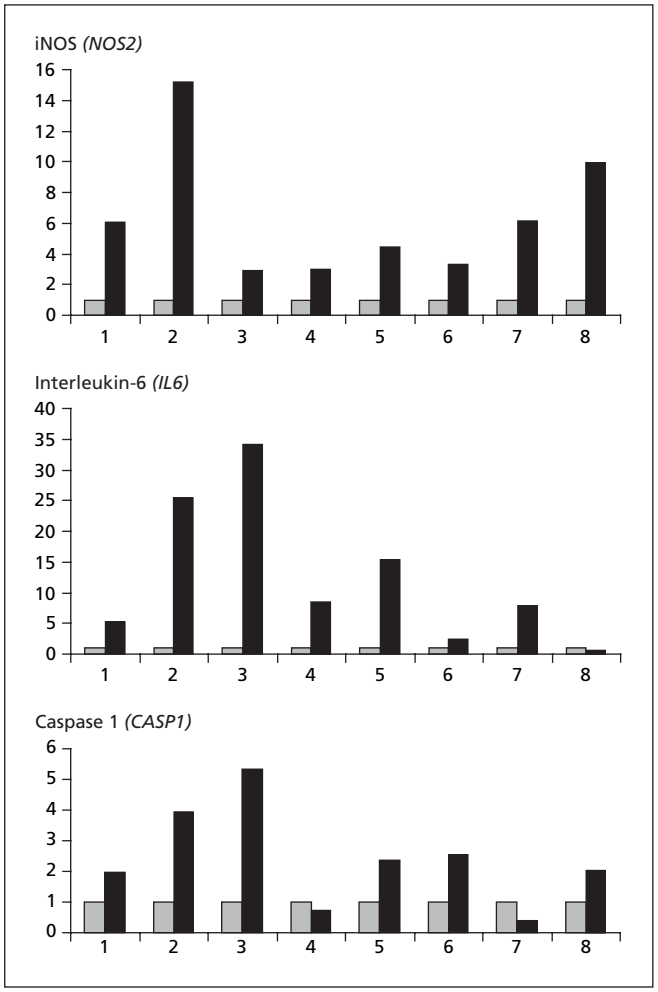

Figure 9. Expression of selected control genes for evaluation of cytokine stimulation.Gene expression was measured in human pancreatic islets (+/- cytokine sitmulation). Expression of these genes was used to ensure satisfactory cytokine stimulation in all preparations. Dark = un-stimulated, light = stimulated.

We used array-based approaches (Bergholdt et al. 2007a) for evaluation of the 32 positional T1D candidate genes identified on chromosome 21 (Bergholdt et al. 2005a). Eight human pancreatic islet preparations were initially used to compare expression levels of the human chromosome 21 genes in un-stimulated versus cytokine stimulated conditions. For this we used a so called Low Density Array (TaqMan) platform, containing expression assays for the selected genes as well as control genes (Bergholdt et al. 2007a). We searched for genes showing differential expression between cytokine stimulated and unstimulated conditions. Human genes down regulated after cytokine exposure of the islets were *ITSN1, TTC3, CHAF1B, HUNK, OLIG1* and *CBR1* (Bergholdt et al. 2007a), Figure 9.

Evaluation of T1D related gene expression in human pancreatic islets is of extreme importance. Testing the target tissue of T1D itself, in which the pathogenetic process takes place, is essential. Human pancreatic islets are, however, very limited as a resource for research, and we believe our observations in the eight pancreatic islet preparations are unique and valuable.

By use of this Low Density Array platform, we further compared expression levels of the same 32 human chromosome 21 genes in mRNA from human lymphocytes (the effector cells in T1D) obtained from 10 newly diagnosed T1D patients (sampled less than 20 weeks after diagnosis) and 10 HLA-matched controls. When comparing the means of expression levels in the control group with the T1D group, none of the comparisons, however, reached statistical significance. Genes most differentially expressed in the two groups were *ITSN1, CBR1, CLDN8, OLIG1* and *CHAF1B,* Figure 10, (Unpublished data by Regine Bergholdt).

Furthermore, a rat beta cell line *INS*-1αβ over expressing pdx-1 was used to evaluate gene expression of genes from this region. In this system we compared expression levels of rat homologues (if existing) of the human chromosome 21 genes in un-stimulated conditions with IL-1β stimulated conditions (Bergholdt et al.

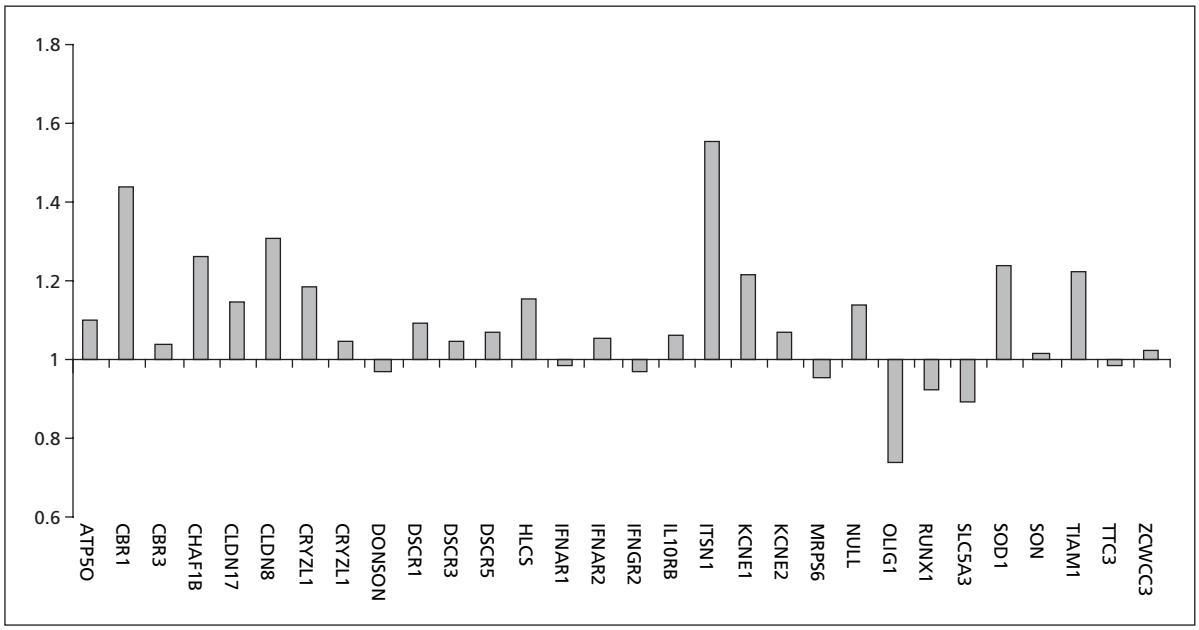

Figure 10. Relative expression levels of human chromosome 21 genes in lymphocytes in controls (n=10) versus T1D (n=10).

2007a). This analysis was part of a larger project evaluating differences in rat gene transcript expression between ± pdx-1 expression and ± IL-1β stimulation (Karlsen et al. Unpublished data). Extensive microarray analyses were performed using an Affymetrix rat oligonucleotide array, containing 30,000 rat gene transcripts. We focused exclusively on gene transcripts homologous to human chromosome 21 genes and looked exclusively at differences in gene expression, with and without IL-1β stimulation for 2 and 24h (Bergholdt et al. 2007a).

Use of other methods than methods controlling for mass significance are increasingly being accepted for microarray studies, despite the many hypotheses tested. We have chosen to use the FDR method to search for the most likely candidate genes, since our aim was to generate hypotheses regarding the potentially most likely genes, which then has to be examined further and validated.

By use of this approach, we identified several "functional" candidate genes. Five rat genes (corresponding to human chromosome 21 genes) significantly changed expression levels by IL-1β stimulation, *CBR3*, *DSCR1* and *CRYZL1* were down regulated, whereas *ATP5O* and *IFNGR2* were up regulated by IL-1β (Bergholdt et al. 2007a). Findings observed in this model system needs confirmation in additional functional studies using other systems, but may be valuable in generating testable hypotheses about genes of presumed functional significance. In the present study we have used model systems primarily addressing beta cell function and response.

Although it is not always straightforward to extrapolate from one species to the other, high concordance was observed between expression changes in the rat beta cell line and human pancreatic islets (Bergholdt et al. 2007a). Genes showing differential expression in the approaches used were quite similar, supporting the identified specific gene products as functionally interesting. Selected genes were involved in e.g. the NFκB-pathway, generally in cytokine signaling, oxidative phosphorylation, defense responses and apoptosis. Transcriptional profiling was used to prioritize among genes and select the ones with putative functional significance in T1D, Figure 11.

Similar findings in human islets and the rat beta cell line strength-

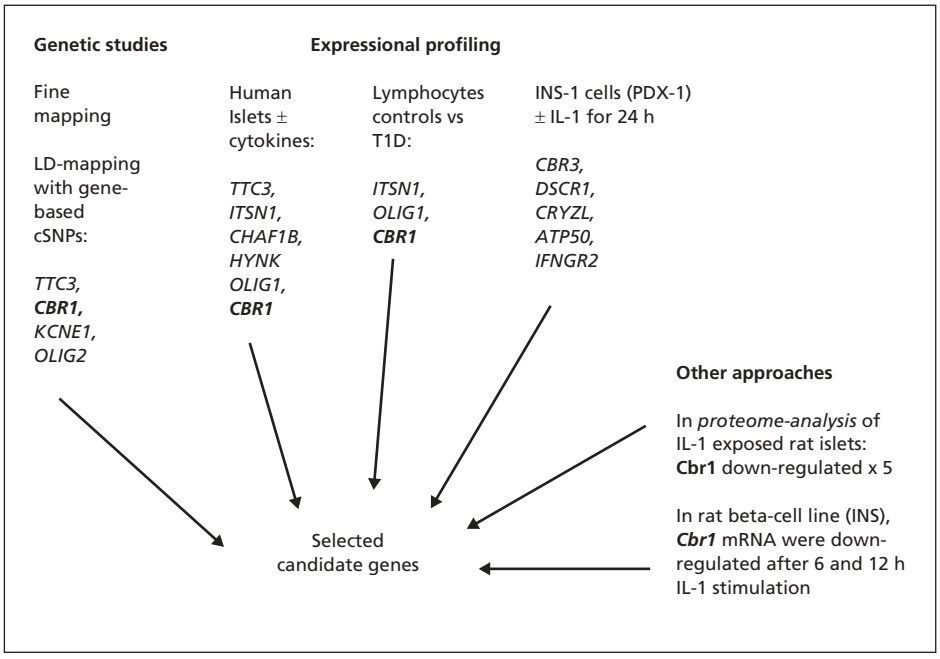

Figure 11. An example of an integrated approach used for selecting candidate genes on chromosome 21; most important are genes identified by several different approaches.

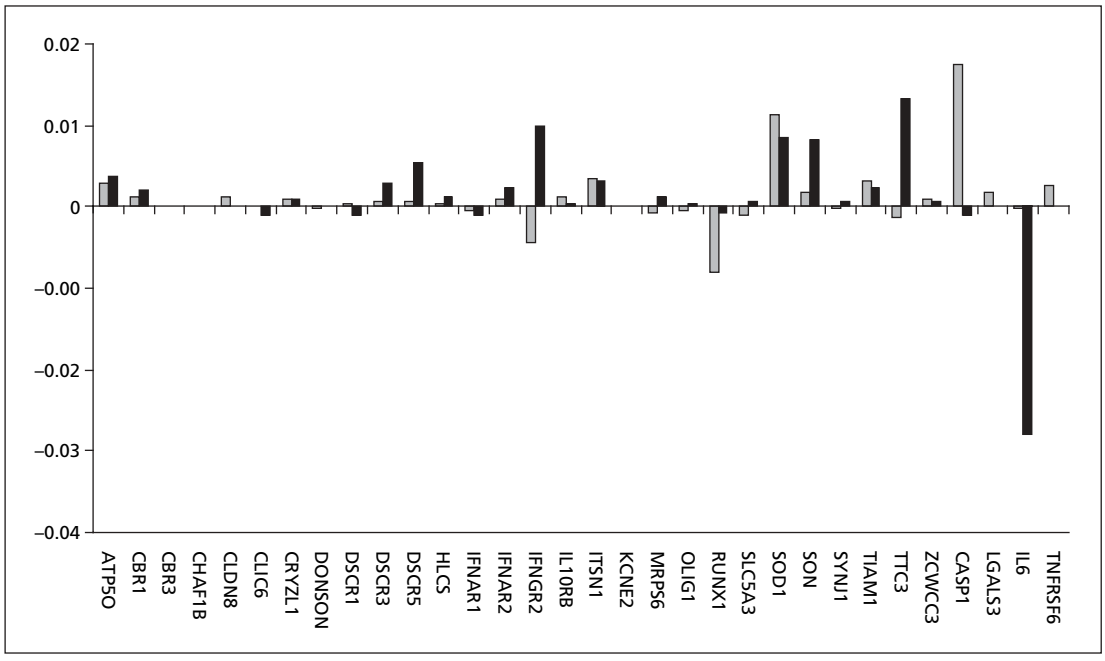

Figure 12. For each gene, average increase or decrease in expression level in control lymphocytes (n=10) versus T1D lymphocytes (n=10) (Grey bars), as well as in unstimulated pancreatic islets (n=8) versus cytokine stimulated islets (n=8) (Black bars) are shown. Only genes clearly expressed in both tissues were compared. Comparisons among this limited number of genes did not provide a clear picture.

ens the validity of the methods used, and supports this as a useful approach for gene mapping. We used this as follow-up on our fine mapping of the linkage region on chromosome 21, in which 32 positional candidate genes had been identified. Most important genes we hypothesize are genes identified by several different approaches, genetic as well as functional, Figure 11.

The analyses furthermore revealed several transcription factor binding-sites shared by several of the differently expressed genes. However, none of the binding sites demonstrated statistically significant over-representation (Bergholdt et al. 2007a). This suggests common regulation and functional interaction of this chromosome 21 region with regions on other chromosomes (e.g. via transcription factors), rather than among the chromosome 21 genes themselves. We think this is a valuable novel approach in selecting yet unidentified genes with a functional significance presumed to underlie multi-factorial genetic diseases. We have used the approach to prioritize chromosome 21 genes for further studies, and have especially focused on the *CBR1* (carbonyl reductase 1) gene, which we are in the process of evaluating further in genetic as well as functional studies, Figure 12 (Bergholdt et al. 2005b; Bang-Berthelsen et al. 2006).

For many diseases access to the relevant target tissue, for evaluating expression of genes, is a problem. This is obviously true for T1D, since access to human pancreatic islets is limited. It could be speculated that in many diseases, human lymphocytes could act as a kind of easily accessible RNA source, potentially useful as a surrogate measure for gene expression in other (target) tissues. To evaluate this hypothesis, we used the expression data on chromosome 21 genes in human lymphocytes (Unpublished data by Regine Bergholdt) to calculate whether these could be identified as being representative for the expression levels we have obtained in human pancreatic islets. Such a comparison is not a simple procedure, the biological variation is extensive and the islet preparations with and without cytokine stimulation is not independent. We therefore only performed a simple visual analysis in which we compared the average increase or decrease in expression level for each gene, among control lymphocytes versus T1D lymphocytes as opposed to unstimulated islets versus cytokine stimulated islets.

As seen in Figure 12, these analyses do not support a simple correlation between gene expression in lymphocytes and pancreatic beta cells under the experimental procedures applied here. Thus, when studying human beta cell gene expression, human lymphocytes might not be an obvious surrogate marker system. However, this clearly needs further exploration. We have therefore focused on the beta cell data, and the eight human islet preparations used, do represent one of the largest studies of human beta cell gene expression ever.

Whether the gene responsible for the chromosome 21 linkage peak is expressed in islets is unknown. T1D is an immune-mediated disease and the few T1D genes validated to date, are immune system genes. However, the immune system may destroy beta cells because of features in beta cells themselves, e.g. the way they react to the immunological challenge. This is the reason we chose to use an established model for T1D pathogenesis, the cytokine stimulation model, mimicking what is happening in the beta cell in T1D pathogenesis for these studies. We can of course not rule out that important genes are expressed in immunological tissue, but to elucidate this was not the scope of the present study.

8.2. FUNCTIONAL INTERACTION NETWORKS

Both genetic and physical data are essential to understand biological systems. For a complete picture they must be integrated and since genetic interactions define functional relationships between the genes, such integration is expected to provide insight into the translation of interactions to phenotype. Individual interactions can be integrated into larger network structures representing protein complexes or pathways. Whereas the genome is rather static, interaction networks are believed to be more dynamic and dependent on the context. They might be active only under certain conditions, in certain cell types or stages of development. Ideally, all conditions and cell types should be tested to capture this presumed variability.

For prioritization of positional candidate genes in broad linkage intervals the use of functional interaction networks (interactomes) may be a valuable method. If linkage intervals obtained for a disease are queried for functional interactions with each other and related to phenotype information for the disease, this holds promise for selection of putative disease genes for further investigation (Gandhi et al. 2006; Lage et al. 2007a). This is supposed to also be applicable to disease associated intervals identified from genome wide association studies, in which often little is known a priori about the functions of the significant loci. Studies like this might generate novel hypotheses regarding pathogenesis of human multi-factorial diseases. They have the potential of identifying new, previously unrecognized components of disease mechanisms, as well as of pin-pointing the most important protein complexes involved.

Furthermore, many diseases have overlapping clinical manifestations/sub-phenotypes and it could be speculated that this may be represented by genetic variation in the same functional pathways. Recently, the existence of so called disease sub-networks was suggested. It was demonstrated that proteins encoded by genes mutated in one inherited genetic disorder, were likely to interact with proteins known to cause similar disorders, presumably by sharing common underlying biochemical mechanisms (Gandhi et al. 2006).

A recent report demonstrated the feasibility of constructing such functional human gene networks and applying these to positional candidate gene identification (Franke et al. 2006). It was shown that obvious candidate genes are not always involved, and that taking an unbiased approach in finding candidate genes, e.g. by using functional networks may result in new testable hypotheses (Franke et al. 2006). In this report a functional human gene network was constructed by integrating information of genes, proteins, protein-protein interactions, microarray co-expressions etc. from a number of databases. Genetic diseases with at least three known disease genes were tested, 17 of the diseases were of complex inheritance. Artificial susceptibility loci with 100 genes around the known disease genes were constructed, and the functional networks then ranked the genes based on known functional interactions. In 54% of cases the known disease gene was ranked among the top five genes for each locus, corresponding to 2.8 fold better than random selection (Franke et al. 2006). This is not too impressive and part of the explanation may be that this report, as well as another recent one (Aerts et al. 2006) incorporate unscored human protein interaction data, and that none of them take advantage of cross-species integration of interaction data. Yet another recent report used only unscored binary interaction pairs to identify candidates (Oti et al. 2006). They identify new candidates and estimate that this corresponds to a 10-fold enrichment compared with positional information alone (Oti et al. 2006), the candidates and thereby the estimated enrichment, however, remains to be validated.

Approaches like these may be valuable in prioritizing candidate genes in linkage regions or from disease associated regions, in which the disease gene(s) are not known. And information of whether genes from the different loci observed, do interact at a functional level are potentially interesting.

Obviously, the input information is crucial for the success of such an approach. Studies will be biased by absence of complete functional information in databases of the majority of genes, and also interaction databases are far from complete. However, hypotheses generated with existing knowledge may be of value, and genes, that would otherwise not have been predicted to be involved in the disease in question, might be identified this way. Data amounts in databases are rapidly increasing. This include increased knowledge regarding genes, proteins, interactions among them, methods integrating high throughput genomic and proteomic approaches, as

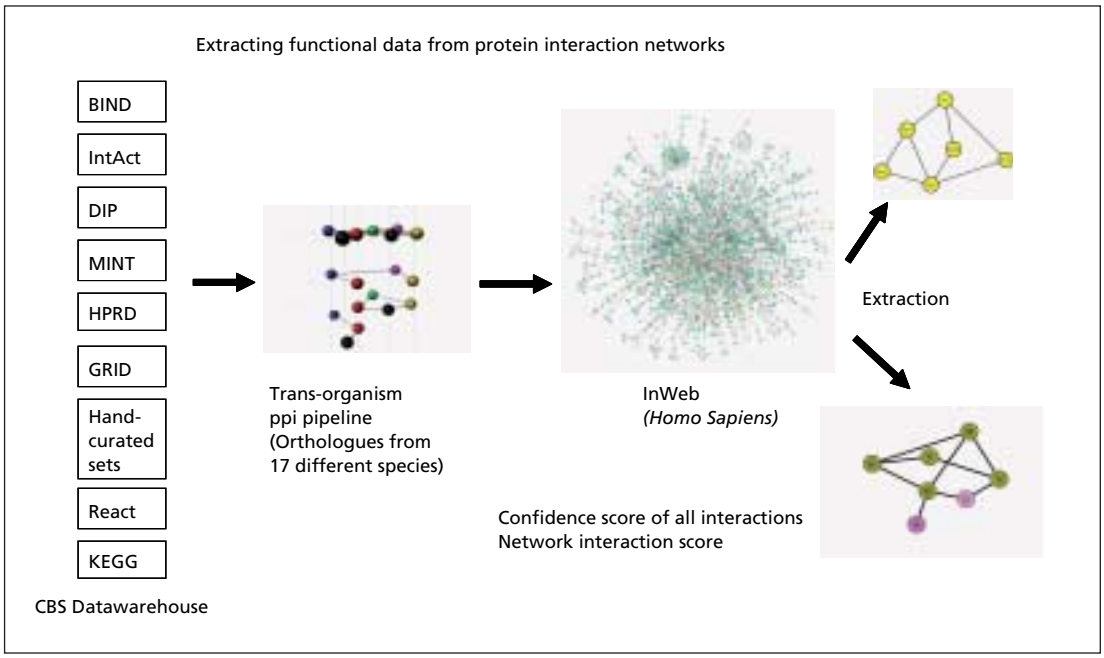

Figure 13. Schematical overview of the tool used for identifying high confidence protein-protein interactions. For the CBS (Center for Bio logical Sequence analysis) data warehouse information from the databases mentioned are extracted and data are integrated from 17 different species into an enormous human protein interaction network. From this InWeb sub networks can be extracted. Only interactions and networks exceeding predefined scores are allowed.

well as text mining methods extracting functional relationships from the literature. All information should be quality controlled, and preferably integrate information from other species.

In our own studies we searched for functionality for rules identified by data mining studies (Bergholdt et al. 2007b), in which several rules of genetic interaction were predicted. We used protein interaction networks to examine whether the gene products from the interacting genetic regions could also be shown to interact in biological pathways.

We used an analysis tool, recently developed at the Center for Biological Sequence Analysis, Danish Technical University, for integrating protein-protein interactions in generating protein interaction networks. At the Center for Biological Sequence Analysis a data warehouse has been developed, holding information from 17 different species and data pooled from all the largest interaction databases, Figure 13.

High standards of quality control and the highest confidence score level were set as a threshold and applied to all interactions included in this data warehouse (Lage et al. 2007a), as opposed to in the other reports mentioned (Aerts et al. 2006; Franke et al. 2006; Oti et al. 2006). This combined human protein network can then be queried and functional protein interaction networks, corresponding to the chromosomal regions in question, be extracted.

We queried this network for interactions corresponding to regions 5 Mb on each side of the two microsatellites in every rule of predicted genetic interaction. For the functional networks extracted, a further stringent criterion (network interaction score) for number of interactions allowed for each non-input protein, was applied, in order to reduce noise from highly interacting proteins (Bergholdt et al. 2007b).

We attempted generation of protein-protein interaction networks for all predicted interactions, using the highly stringent confidence score as well as a stringent network interaction score. The strategy used is outlined in Figure 14.

We demonstrated support for physical interactions at the protein level for all the predicted genetic interactions, demonstrating that genetic interaction is probably important in T1D susceptibility (Bergholdt et al. 2007b). The high degree of compatibility between genetic rules of interaction derived from data mining analysis, as well as the networks obtained, support that the genetic rules have predictive value for T1D and allow for identification of specific candidate genes/proteins of the involved genomic regions. The method developed and the analyses made represent a completely novel exploration of integrative genomics. We integrate genetic and physical protein interactions to identify protein networks that link genomic regions showing genetic interactions. The resulting networks point directly to novel candidates visualized in context of their interaction network, potentially providing even further biological insight. Furthermore, we integrated information from more than just the two directly genetically interacting regions and obtained some highly in-

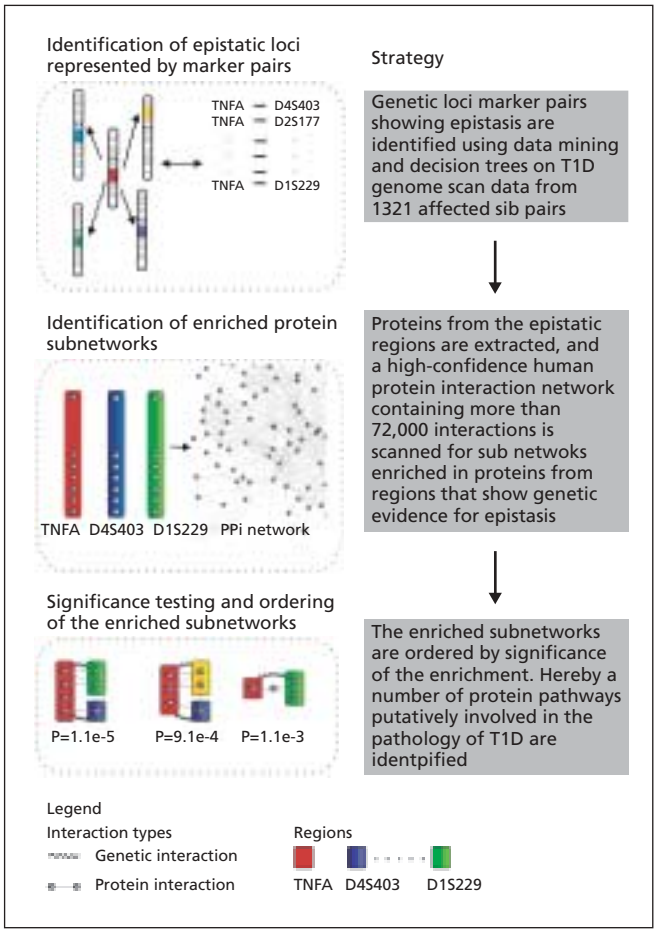

Figure 14. A schematical overview of the strategy used for the current study.

teresting and significant functional modules, Figure 15. By integrating e.g. all the regions demonstrated to in themselves genetically interact with the HLA-region, we can directly point at candidate genes for the underlying interaction. We applied a statistical test devised specifically for this work to evaluate the significance of the protein networks linking several genetic loci that have shown genetic interactions. The networks and functional modules looked promising, supported by the highly significant nature of many of them.

Findings are in line with current knowledge regarding mechanisms involved in T1D pathogenesis, since central proteins in the protein-protein interactions were proteins involved in signal transduction, kinase activity, regulation of transcription and apoptosis, reviewed in (Bergholdt et al. 2004b).

In several of the networks including *TNFA* as a marker for the MHC region, the same four MHC-region genes are present, *TUBB, RPS18, BAT1* and *ITPR3*, and responsible for the interaction observed, Figure 15. Of these genes, recent focus has been on the *ITPR3* gene, which demonstrated strong T1D association in two Swedish case-control cohorts (Roach et al. 2006), as described in chapter 4. It is of potential interest that we were able to pin-point the same gene as important also in putative gene-gene interactions. This may help putting this gene into a functional context of other interacting proteins, and we see the central position of this gene and other known HLA-region genes as an indication of the value of this method.

We examined whether additional information could be obtained by the use of a less stringent interaction score, allowing for more interactions for each non input-protein, but keeping the same confidence score. Interestingly, we did see the original networks being extended (Bergholdt et al. 2007b), and we hypothesized that this may add information of possible underlying biological pathways. As an example the protein-protein network between *TNFA*=1 and *D11S910* are shown with the stringent, as well as the less stringent interaction score, Figure 16.

The extended network for this interaction included genes and pathways of possible relevance to T1D, i.e. classical HLA-genes, his-

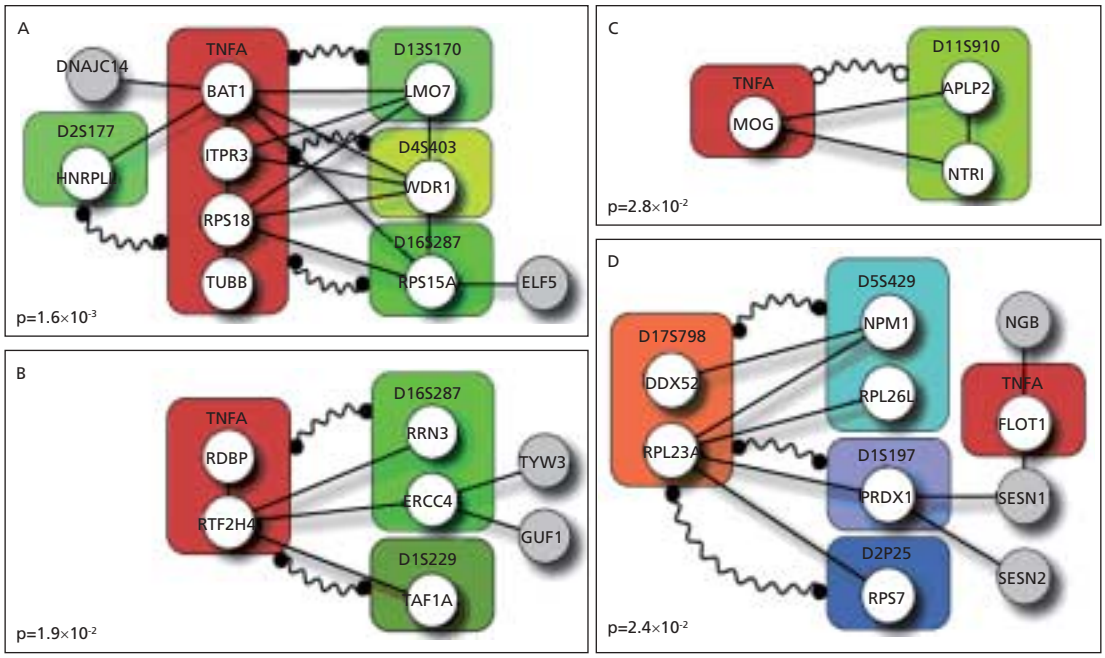

Figure 15. Significant functional modules. Straight lines represent validated protein-protein interactions, curved lines represent demonstrated genetic interactions (black bullets: predictive, white bullets: protective interactions). Circles with gene names represent the gene encoding the protein of the interaction. Boxes are the marker regions shown to be involved in the genetic interactions and in which the genes are located.

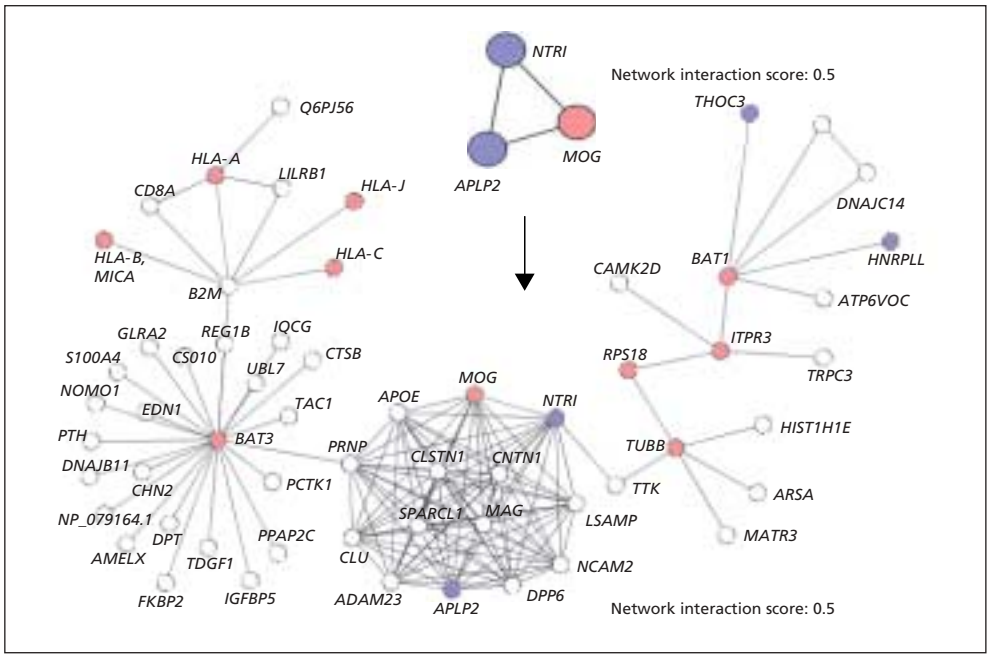

Figure 16. TNFA-D11S910 interaction network with network interaction scores of 0.5 and 0.15. It is obvious that the original stringent "candidate gene" network was largely extended.

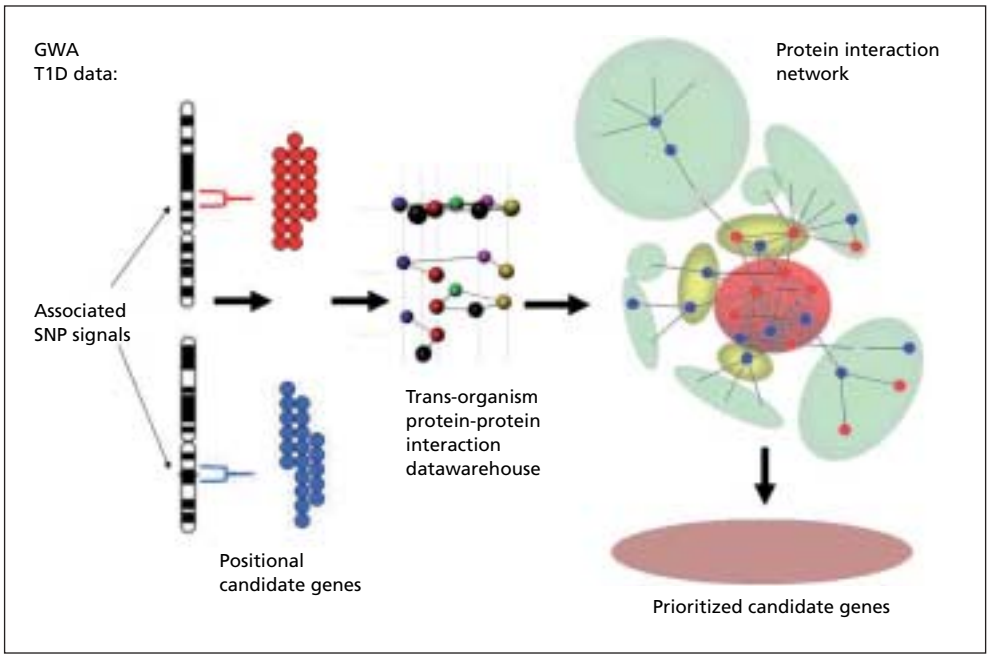

Figure 17. An example of how regions harboring SNPs demonstrating genome-wide association could be investigated for potential protein-protein interactions. This may be used for prioritizing most likely functional candidate genes in the corresponding regions, these genes being the central ones, i.e. with the highest stringency scores in the networks constructed.

tone genes, the *ITPR3* gene, a calcium-calmodulin associated protein kinase and genes encoding transcription factors. The combination of *TNFA* and *D11S910* was the only rule, which seemed to confer protection against T1D (Bergholdt et al. 2007b), and the networks constructed may provide insight into mechanisms underlying this.

We have identified a number of completely novel T1D candidate genes. Many of the networks are highly significant, showing that the identified candidates are strongly backed up by a combination of the human protein interaction data and the genetic interaction data in patients with T1D. We therefore think this may be a strong and valuable approach for identification of functional candidate genes.

With 30% or more of human genes lacking functional annotation, existing protein interaction databases and maps are still far from complete, and many of the protein interactions in databases have not been rigorously tested and validated. Our data are the most extended suggestions of genetic epistasis in a multi-factorial disease backed up by high-confidence protein network analyses with implications for functionality. Our data present a new approach to the analysis of complex genetic data, as well as a new comprehensive way of surveying and extracting data from protein-interaction databases, representing a methodological advance in genomics and bioinformatics. A paper by several of the co-authors of our paper was very recently published, in which the new method for elucidating protein-protein interaction networks was validated and thoroughly described (Lage et al. 2007a). The combination of these two approaches does reveal significant and interesting observations, and we propose this type of integrative analysis as a general method for elucidation of genes and networks involved in T1D and other complex diseases. Future studies and validation, of the interaction complexes we observed, are necessary to confirm their biological significance.

In summary, we used an interaction network method and confidence scoring system previously developed (Lage et al. 2007b) to make interaction networks linking regions shown to interact genetically (identified by the method described previously (Pociot et al. 2004)), using data that has not previously been investigated by this method. The dataset was furthermore much larger than used before. Hereby we integrated genetic and physical protein interactions to identify protein networks that link genomic regions showing genetic interactions, and point directly to novel candidates. We included a statistical test devised specifically for this work in order to evaluate the significance of protein networks linking several genetic loci. All the identified networks were highly significant. The analyses made and developed do represent a completely novel exploration of integrative genomics. Integrative genomics is an emerging and promising field in order to tackle complex disease.

Other applications for use of knowledge regarding functional human protein-protein interaction networks, could be to look for interactions among all known T1D linked regions or among regions identified in a genome-wide association study. A strategy for this is exemplified in Figure 17.

9. CONCLUSION AND FUTURE STRATEGIES

T1D, affecting 10-20 million people worldwide, is a multi-factorial disease of which the pathogenetic molecular mechanisms are still not fully understood. Despite the fact that many chromosomal regions and individual genes have been reported to be associated with T1D, only a few have been repeatedly demonstrated to show consistent and significant linkage and/or association to T1D. The risk of developing T1D is influenced by the individual amount of susceptibility and protective alleles inherited at a large, but unknown, number of variant sites, and by their expression in context of a diverse range of environmental exposures. Under such a model, it is inevitable that efforts to isolate the effects of a single gene or a single environmental exposure will present challenges in terms of power, interpretation and replication. Different analytical approaches for identification and/or clarification of the contributions of loci to T1D susceptibility have been used over the last decade. In our own studies we have performed classical candidate gene studies followed by functional analyses, linkage and fine mapping studies, epidemiological studies, as well as studies using new approaches to evaluate gene-gene interaction, protein-protein interaction, and transcriptional profiling in attempt to map T1D genes.

To fully understand complex genetic diseases, the best approach would be re-sequencing the complete genome in a large population of cases and controls, for each multi-factorial disease. This approach would not be limited by biased choices of candidate genes and would cover all coding and non-coding variants of importance. This is, however, not feasible at present, and would demand major breakthroughs in sequencing technology and lower costs.

The most effective strategy, for improving T1D gene identification, seems to be one of the most simple, i.e. increasing sample sizes. This can be accomplished by performing new larger studies and/or

by merging data from previous studies. Future collaborative efforts, in combination with new tools should increase the chances of identifying T1D susceptibility and protective loci and the molecular pathogenesis of T1D. Recent progress, e.g. the establishment of the Type 1 Diabetes Genetics Consortium, is likely to facilitate future studies in this field. Largely increased sample sizes and repositories will provide both initially robust detection of susceptibility loci for T1D and samples to permit replication of initial findings.

Genome-wide LD mapping may be more powerful than classical linkage analysis and candidate gene approaches. By determining the extended haplotypes at any given locus in a population, it is possible to identify which SNPs will be redundant and which will be essential for association studies. Resources like HapMap and SNP arrays, as well as genome-wide proteomic data, will and have lead to increased knowledge about genomic regions associated with multi-factorial diseases, and their molecular pathogenesis. Genome-wide SNP typing has now become feasible and it is now possible to efficiently and comprehensively test common genetic variation across the entire genome for a role in common disease. Genome wide association studies offer great promise and has just recently proven to be successful also in T1D (Hakonarson et al. 2007; Wellcome Trust Case Control Consortium 2007). Several new interesting T1D loci were suggested in these studies and known risk loci were confirmed (Hakonarson et al. 2007; Todd et al. 2007; Wellcome Trust Case Control Consortium 2007). However, for none the causal gene or variant is known. It is likely that whole genome association approaches will be able to lead to detection of T1D genes also of minor effects (e.g. with λ_S below 1.2).

Association signals can define regions of interest, and positional candidacy may highlight pathways and mechanisms of interest, but cannot provide identification of causal genes. Extensive re-sequencing and fine mapping, followed by targeted functional studies are essential in identifying causal genes and required to translate findings into robust statements about molecular mechanisms.

Multi-factorial diseases involve multiple interacting genetic determinants. Genetic interactions represent functional relationships between genes, in which the phenotypic effect of one gene is modified by another. However, statistical methods in current use for finding such genes essentially work under single gene models, thus not allowing identification of complex traits. In current practice, the majority of whole genome scans for complex trait loci are conducted by multipoint methods, assuming a single trait locus or multiple trait loci, unrelated to each other. Therefore, new strategies for detecting sets of marker loci, which are linked to multiple interacting disease genes are in demand. We have explored the use of data mining methods to evaluate genetic interaction. The importance of the predicted genetic interactions was strongly supported by comprehensive, high-confidence protein-protein interaction networks of the corresponding regions. This allowed identification of candidate genes of likely functional significance in T1D. We believe our study represents one of the most extensive suggestions of genetic epistasis in a multi-factorial disease supported by protein network analysis with implications for functionality. Another approach for selecting candidate genes of functional importance is transcriptional profiling, which we have used in further characterization of a linkage region on chromosome 21.

In addition to candidate-gene and genome-wide linkage or association studies, also other approaches may be useful in identifying new T1D susceptibility/resistance regions. Identification of syntenic regions, i.e. regions demonstrating correspondance in gene-order between the chromosomes of different species, and conferring susceptibility/resistance to diabetes in animal models for T1D, can be useful for identification of similar genes in man. Characterization of identical (overlapping) regions identified through genome scans of other autoimmune or inflammatory diseases is another approach. Occurrence of common features of autoimmune diseases and the existence of co-association of multiple autoimmune diseases in the same individual or family suggest common genetic factors predisposing to autoimmunity to be involved. Clustering of autoimmune susceptibility loci suggests that there may be related genetic backgrounds contributing to susceptibility of clinically distinct diseases.

Proteins exert their function within complex networks, and it is likely that disturbances of such networks, e.g. caused by genetic variation or changes in expression levels, may lead to instability and cause disease. Like any other cell, also the beta cell phenotype is maintained by multiple protein-protein and DNA-protein interactions in a state of dynamic stability, which is characterized by a capacity for robust, but limited, self regulation. If pushed outside this range by intra-cellular events or environmental insults, dynamic instability is produced and cascading catastrophic events for the cell may happen. Genetic variations producing protein expression changes or protein modifications may account for the different levels of risk for disease in different individuals, and suggesting that integrative methods for gene discovery are necessary. Exploiting protein-protein interactions can increase the likelihood of identifying positional candidate disease genes, and may on a larger scale predict novel candidate genes. In our own studies of genetic interactions, followed by construction of protein-protein interaction networks, we propose many novel candidate genes. We included an analysis of the statistical significance of the networks containing the candidates. The networks were highly significant, showing that the identified candidates are strongly backed up by a combination of human protein interaction data and the genetic interaction data. Our method proposes one possible solution to a difficult problem common to a wide range of researchers working in the broad field of complex diseases. Moving from genomes to interaction networks is important. We "know" many of the genes today, now the challenging task is to understand how they affect T1D risk. "Integrative genomics" is an emerging, promising field to tackle complex disease. It provides increased knowledge about functional mechanisms underlying disease and thereby an approach to increase our understanding of disease pathogenesis.

Much work in genomics and genetics in multi-factorial diseases still has to be carried out. Several regions have been implicated, but not yet confirmed. Following identification of genes there is a huge demand for functional genomics. The number of identified susceptibility genes may continue to grow, and the elucidation of their function in the pathogenesis of multi-factorial diseases, e.g. T1D, will be important for understanding their molecular pathogenesis. Approaches used will vary according to the function of the genes, but will include expression studies and generation of transgenic and knockout animal models.

We are still far from a comprehensive understanding of the molecular pathogenesis of T1D and other multi-factorial diseases. This makes it difficult to identify optimal strategies for intervention and treatment. Advances in genome informatics, genotyping technology, statistical methodology, as well as the availability of large-scale clinical material, however, holds promise for the future. The recent success of truly genome-wide association scans and the prospects for combining genetics with high-throughput genomics are additional sources of optimism for the future.

ENGLISH SUMMARY

The risk of developing type 1 diabetes (T1D) and other multi-factorial diseases is influenced by multiple genes, including genes conferring susceptibility as well as protection, and likely interacting in complex ways with each other and with environmental factors. Under such a model, it is inevitable that efforts to isolate the effects of a single gene or a single environmental exposure will present challenges in terms of power, interpretation and replication.

The aim of this thesis has been to review and discuss different approaches for identifying susceptibility genes, as well as to evaluate their evidence in multi-factorial diseases, using T1D as a model. More recent approaches integrate biological and genetic informa-

tion in gene mapping, and focus is increasingly on genome-wide approaches. The developments in gene identification have been tremendous. However, despite new methods and large studies, the overall impression still is that inconsistencies of data exist, and that novel approaches, integrating genetic and functional information, are needed. These may include novel analytical methods, such as global/genome-wide approaches, use of information from other autoimmune diseases and rare syndromes, functional approaches, as well as studies of interactions.

In the scientific work, which this thesis is based upon, we have used classical candidate gene studies, linkage and association analyses, comparative genetics, transcriptional profiling as well as studies of gene-gene and protein-protein interactions to map T1D genes. By these methods we have identified genes and genetic regions of importance in T1D, as well as elucidated interaction networks/pathways by development of new methods to functionally prioritize candidate genes for further investigation. Such "integrative genomics" provides increased knowledge about functional mechanisms underlying disease and thereby an approach to increase our understanding of disease pathogenesis. Much work in genomics and genetics in multi-factorial diseases still has to be carried out, and following identification of genes there is a huge demand for functional genomics. Today we "know" many of the genes, the challenging task is to understand how they affect T1D risk.

References

Abecasis G, Cookson W (2000) GOLD – Graphical Overview of Linkage Disequilibrium. Bioinformatics 16:182-183

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO (2001) Extent and distribution of linkage disequilibrium in three genomic regions. American Journal of Human Genetics 68:191-197

Abel K, Reneland R, Kammerer S, Mah S, Hoyal C, Cantor CR, Nelson MR, Braun A (2006) Genome-wide SNP association: identification of susceptibility alleles for osteoarthritis. Autoimmunity Reviews 5:258-263

Adorini L (2001) Interleukin 12 and autoimmune diabetes. Nature Genetics 27:131-132

Aerts S, Lambrechts D, Maity S, Van-Loo P, Coessens B, De-Smet F, Tranchevent LC, De-Moor B, Marynen P, Hassan B, Carmeliet P, Moreau Y (2006) Gene prioritization through genomic data fusion. Nature Biotechnology 24:537-544

Agardh D, Nilsson A, Carlsson A, Kockum I, Lernmark A, Ivarsson S (2002) Tissue transglutaminase autoantibodies and human leucocyte antigen in Down's syndrome patients with coeliac disease. Acta Paediatrica 91:34-38

Alberti K, Zimmet P (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabetic Medicine 15:539-553

Allison D, Cui X, Page G, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics 7:55-65

Allotey R, Mohan V, McDermott M, Deepa R, Premalatha G, Hassan Z, Cassell P, North B, Vaxillaire M, Mein C, Swan D, O'Grady E, Ramachandran A, Snehalatha C, Sinnot P, Hemmatpour S, Froguel P, Hitman G (2004) The EIF2AK3 gene region and type I diabetes in subjects from South India. Genes and Immunity 5:648-652

Altshuler D, Brooks L, Chakravarti A, Collins F, Daly M, Donnelly P (2005) A haplotype map of the human genome. Nature 437:1299-1320

Amano K, Sago H, Uchikawa C, Suzuki T, Kotliarova S, Nukina N, Epstein C, Yamakawa K (2004) Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. Human Molecular Genetics 13:1333-1340

Anwar A, Walker J, Frier B (1998) Type 1 diabetes mellitus and Down's syndrome: prevalence, management and diabetic complications. Diabetic Medicine 15:160-163

Ardlie KG, Kruglyak L, Seielstad M (2002a) Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics 3:299-309

Ardlie KG, Lunetta KL, Seielstad M (2002b) Testing for population subdivision and association in four case-control studies. American Journal of Human Genetics 71:304-311

Atkinson M, Eisenbarth G (2001) Type 1 diabetes: new perspectives on disease pathogenesis and treatment. Lancet 358:221-229

Awata T, Inoue K, Kurihara S, Ohkubo T, Inoue I, Abe T, Takino H, Kanazawa Y, Katayama S (2000) Missense variations of the gene responsible for Wolfram syndrome (WFS1/wolframin) in Japanese: possible contribution of the Arg456His mutation to type 1 diabetes as a nonautoimmune genetic basis. Biochemical and Biophysical Research Communications 268:612-616

Bachmaier K, Krawczyk C, Kozieradzki I, Kong YY, Sasaki T, Oliveira-dos-Santos A, Mariathasan S, Bouchard D, Wakeham A, Itie A, Le J, Ohashi PS, Sarosi I, Nishina H, Lipkowitz S, Penninger JM (2000) Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b. Nature 403:211-216

Bang-Berthelsen C, Bergholdt R, Heding P, Nerup J, Pociot F (2006) The role of CBR1 in the pathogenesis of type 1 diabetes. Diabetologia 49:A254

Barratt B, Payne F, Lowe C, Hermann R, Healy B, Harold D, Concannon P, Gharani N, McCarthy M, Olavesen M, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien D, Ronningen K, Gillespie K, Tuomilehto-Wolf E, Tuomilehto J, Bennett S, Clayton D, Cordell H, Todd J (2004) Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes 53:1884-1889

Barrett TG (2001) Mitochondrial diabetes, DIDMOAD and other inherited diabetes syndromes. Best Practice and Research Clinical Endocrinology and Metabolism 15:325-343

Barrett TG, Bundey SE, Macleod AF (1995) Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome. Lancet 346:1458-1463

Becker KG (1999) Comparative genetics of type 1 diabetes and autoimmune disease: common loci, common pathways? Diabetes 48:1353-1358

Becker KG (2004) The common variants/multiple disease hypothesis of common complex genetic disorders. Medical Hypotheses 62:309-317

Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, Trent JM (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. Proceedings of the National Academy of Sciences of the United States of America 95:9979-9984

Begovich A, Carlton V, Honigberg L, Schrodi S, Chokkalingam A, Alexander H, Ardlie K, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. American Journal of Human Genetics 75:330-337

Bell G, Horita S, Karam J (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176-183

Bellanne-Chantelot C, Clauin S, Chauveau D, Collin P, Daumont M, Douillard C, Dubois-Laforgue D, Dusselier L, Gautier JF, Jadoul M, Laloi-Michelin M, Jacquesson L, Larger E, Louis J, Nicolino M, Subra JF, Wilhem JM, Young J, Velho G, Timsit J (2005) Large genomic rearrangements in the hepatocyte nuclear factor-1β (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. Diabetes 54:3126-3132

Benichou J (2002) Attributable risk. In: Elston R, Olson J, Palmer L (eds). Biostatistical genetics and genetic epidemiology: Wiley Reference Series in Biostatistics. John Wiley and Sons Ltd., Chichester, pp 36-49

Bergholdt R, Eising S, Nerup J, Pociot F (2006) Increased prevalence of Down's syndrome in individuals with type 1 diabetes in Denmark: a nationwide population-based study. Diabetologia 49:1179-1182

Bergholdt R, Ghandil P, Johannesen J, Kristiansen OP, Kockum I, Luthman H, Ronningen KS, Nerup J, Julier C, Pociot F (2004a) Genetic and functional evaluation of an interleukin-12 polymorphism (IDDM18) in families with type 1 diabetes. J Med Genet 41:e39-

Bergholdt R, Heding P, Nielsen K, Nolsøe R, Sparre T, Størling J, Nerup J, Pociot F, Mandrup-Poulsen T (2004b) Type 1 diabetes mellitus: an inflammatory disease of the islet. Advances in Experimental Medicine and Biology 552:129-153

Bergholdt R, Karlsen AE, Hagedorn PH, Aalund M, Nielsen JH, Kruhoffer M, Orntoft T, Wang H, Wollheim CB, Nerup J, Pociot F (2007a) Transcriptional profiling of type 1 diabetes genes on chromosome 21 in a rat beta-cell line and human pancreatic islets. Genes and Immunity 8:232-238

Bergholdt R, Karlsen AE, Johannesen J, Hansen PM, Dinarello CA, Nerup J, Pociot F (1995) Characterization of polymorphisms of an Interleukin-1 receptor-type-1 gene (IL1RI) promotor region (p2) and their relation to insulin-dependent diabetes-mellitus (IDDM). Cytokine 7:727-733

Bergholdt R, Larsen ZM, Andersen NA, Johannesen J, Kristiansen OP, Mandrup-Poulsen T, Nerup J, Pociot F (2000) Characterization of new polymorphisms in the 5' UTR of the human Interleukin-1 receptor type 1 (IL1R1) gene: linkage to Type 1 diabetes and correlation to IL-1RI plasma level. Genes and Immunity 1:495-500

Bergholdt R, Nerup J, Pociot F (2005a) Fine mapping of a region on chromosome 21q21.11-q22.3 showing linkage to type 1 diabetes. Journal of Medical Genetics 42:17-25

Bergholdt R, Nielsen J, Nerup J, Pociot F (2005b) Expression profiling used for mapping susceptibility genes in a type 1 diabetes linked region on chromosome 21. Diabetologia 48:A109

Bergholdt R, Pociot F (2006) Type 1 and 2 Diabetes. In: Vandenbroeck K (ed) Cytokine gene polymorphisms in multi-factorial conditions. CRC Press, Taylor and Francis Group LLC, Boca Raton, Florida, USA, pp 305-319

Bergholdt R, Størling Z, Lage K, Karlberg E, Òlason P, Aalund M, Nerup J, Brunak S, Workman C, Pociot F (2007b) Integrative analysis for finding genes and networks involved in diabetes and other complex diseases. Genome Biology In press

Bergholdt R, Taxvig C, Eising S, Nerup J, Pociot F (2005c) CBLB variants in Type 1 Diabetes and their genetic interaction with CTLA4. Journal of Leukocyte Biology 77:579-585

Beyer A, Bandyopadhyay S, Ideker T (2007) Integrating physical and genetic maps: from genomes to interaction networks. Nature Reviews Genetics 8:699-710

Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC (1986) Evidence that autoimmunity in man is a Mendelian dominant trait. American Journal of Human Genetics 39:584-602

Bidwell J, Keen L, Gallagher G, Kimberly R, Huizinga T, McDermott M, Oksenberg J, McNicholl J, Pociot F, Hardt C (1999) Cytokine gene polymorphism in human disease: on-line databases. Genes and Immunity 1:3-19

Bidwell J, Keen L, Gallagher G, Kimberly R, Huizinga T, McDermott M, Oksenberg J, McNicholl J, Pociot F, Hardt C (2001) Cytokine gene polymorphism in human disease: on-line databases, supplement 1. Genes and Immunity 2:61-70

Bilbao J, Martin-Pagola A, Calvo B, Perez de Nanclares G, Gepv-N, Castano L (2002) Contribution of MIC-A Polymorphism to Type 1 Diabetes Mellitus in Basques. Annals of the New York Academy of Sciences 958:321-324

Bittles AH, Glasson EJ (2004) Clinical, social, and ethical implications of changing life expectancy in Down syndrome. Developmental Medicine and Child Neurology 46:282-286

Bjorses P, Halonen M, Palvimo JJ, Kolmer M, Aaltonen J, Ellonen P, Perheentupa J, Ulmanen I, Peltonen L (2000) Mutations in the AIRE gene: Effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. American Journal of Human Genetics 66:378-392

Boffelli D, Nobrega MA, Rubin EM (2004) Comparative genomics at the vertebrate extremes. Nature Reviews Genetics 5:456-465

Borch-Johnsen K (1989) The prognosis of insulin-dependent diabetes mellitus. An epidemiological approach. Danish Medical Bulletin 36:336-348

Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics 33 Suppl:228-237

Bottini N (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nature Genetics 36:337-338

Brassat D, Motsinger AA, Caillier SJ, Erlich HA, Walker K, Steiner LL, Cree BAC, Barcellos LF, Pericak-Vance MA, Schmidt S, Gregory S, Hauser SL, Haines JL, Oksenberg JR, Ritchie MD (2006) Multifactor dimensionality reduction reveals gene-gene interactions associated with multiple sclerosis susceptibility in African Americans. Genes and Immunity 7:310-315

Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A (2001) High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 98:581-584

Buhler J, Owerbach D, Schaffer A, Kimmel M, Gabbay K (1997) Linkage analyses in type-i diabetes-mellitus using caspar, a software and statistical program for conditional analysis of polygenic diseases. Human Heredity 47:211-222

Burch PR, Milunsky A (1969) Early-onset diabetes mellitus in the general and Down's syndrome populations. Genetics, aetiology, and pathogenesis. Lancet 1:554-558

Butler A, Janson J, Bonner-Weir S, Ritzel R, Rizza R, Butler P (2003) Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. Diabetes 52:102-110

Caillat-Zucman S, Daniel S, Djilali-Saiah I, Timsit J, Garchon H, Boitard C, Bach J (1995) Family study of linkage disequilibrium between TAP2 transporter and HLA class II genes. Absence of TAP2 contribution to association with insulin-dependent diabetes mellitus. Human Immunology 44:80-87

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nature reviews. Genetics 2:91-99

Carel JC, Boitard C, Bougnères PF (1993) Decreased insulin response to glucose in islet cell antibody-negative siblings of type 1 diabetic children. Journal of Clinical Investigation 92:509-513

Carlborg O, Haley C (2004) Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics 5:618-625

Carlborg O, Kerje S, Schütz K, Jacobsson L, Jensen P, Andersson L (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. Genome Research 13:413-421

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. American Journal of Human Genetics 74:106-120

Carlton VEH, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, Chang M, Catanese JJ, Leong DU, Ardlie KG, Kastner DL, Seldin MF, Criswell LA, Gregersen PK, Beasley E, Thomson G, Amos CI, Begovich AB (2005) PTPN22 genetic variation: Evidence for multiple variants associated with rheumatoid arthritis. American Journal of Human Genetics 77:567-581

Carthew RW (2006) Gene regulation by microRNAs. Current Opinion in Genetics & Development 16:203-208

Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, Balkau B, Charpentier G, Pattou F, Stetsyuk V, Scharfmann R, Staels B, Frühbeck G, Froguel P (2006) Transcription Factor TCF7L2 Genetic Study in the French Population: Expression in Human {beta}-Cells and Adipose Tissue and Strong Association With Type 2 Diabetes. Diabetes 55:2903-2908

Chao C, Sytwu H, Chen E, Toma J, McDevitt H (1999) The role of MHC class II molecules in susceptibility to type I diabetes: identification of peptide epitopes and characterization of the T cell repertoire. Proceedings of the National Academy of Sciences of the United States of America 96:9299-9304

Chapman JM, Cooper JD, Todd JA, Clayton DG (2003) Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Human Heredity 56:18-31

Chelala C, Duchatelet S, ML J, Bergholdt R, Dubois-Laforgue D, Ghandil P, Pociot F, Caillat-Zucman S, Timsit J, Julier C (2007) PTPN22 R620W Functional Variant in Type 1 Diabetes and Autoimmunity Related Traits. Diabetes 56:522-526

Chen CC, Huang CY, Huang FY, Dang CW, Lin CL, Lo FS, Lin CH, Tsang CC, Li HJ, Liu HF, Chu CC, Lin M, Lee YJ (2005) The CBLB gene and Graves' disease in children. Journal of Pediatric Endocrinology & Metabolism 18:1119-1126

Cheung V, Conlin L, Weber T, Arcaro M, Jen K, Morley M, Spielman R (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genetics 33:422-425

Chiang YJ, Kole HK, Brown K, Naramura M, Fukuhara S, Hu RJ, Jang IK, Gutkind JS, Shevach E, Gu H (2000) Cbl-b regulates the CD28 dependence of T-cell activation. Nature 403:216-220

Chikuba N, Akazawa S, Yamaguchi Y, Kawasaki E, Takino H, Takao Y, Maeda Y, Okuno S, Yamamoto H, Yokota A, et a (1992) Type 1 (insulin-dependent) diabetes mellitus with coexisting autoimmune thyroid disease in Japan. Internal medicine (Tokyo, Japan) 31:1076-1080

Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. Genetic Epidemiology 27:415-428

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305:869-872

Collier DA, Barrett TG, Curtis D, Macleod A, Arranz MJ, Maassen JA, Bundey S (1996) Linkage of Wolfram syndrome to chromosome 4p16.1 and evidence for heterogeneity. American Journal of Human Genetics 59:855-863

Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proceedings of the National Academy of Sciences of the United States of America 96:15173-15177

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

Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS (1998) A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nature genetics 19:292-296

Conrad D, Andrews T, Carter N, Hurles M, Pritchard J (2006) A high-resolution survey of deletion polymorphism in the human genome. Nature Genetics 38:75-81

Cordell H (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human Molecular Genetics 11:2463-2468

Cordell H, Todd J, Bennett S, Kawaguchi Y, Farrall M (1995) Two-locus maximum lod score analysis of a multi-factorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. American Journal of Human Genetics 57:920-934

Cordell H, Wedig G, Jacobs K, Elston R (2000) Multilocus linkage tests based on affected relative pairs. American Journal of Human Genetics 66:1273-1286

Cordell HJ, Todd JA, Hill NJ, Lord CJ, Lyons PA, Peterson LB, Wicker LS, Clayton DG (2001) Statistical Modeling of Interlocus Interactions in a Complex Disease: Rejection of the Multiplicative Model of Epistasis in Type 1 Diabetes. Genetics 158:357-367

Cornelis F, Faure S, Martinez M, Prud'homme J-F, Fritz P, Dib C, Alves H, et al. (1998) New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. Procedings of the National Academy of Science, USA 95:10746-10750

Cox N, Frigge M, Nicolae D, Concannon P, Hanis C, Bell G, Kong A (1999)

Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nature Genetics 21:213-215
Cox NJ, Wapelhorst B, Morrison VA, Johnson L, Pinchuk L, Spielman RS, Todd JA, Concannon P (2001) Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. American Journal of Human Genetics 69:820-830
Cruz T, Valdes A, Santiago A, Frazer de Llado T, Raffel L, Zeidler A, Rotter J, Erlich H, Rewers M, Bugawan T, Noble J (2004) DPB1 alleles are associated with type 1 diabetes susceptibility in multiple ethnic groups. Diabetes 53:2158-2163
Cucca F, Lampis R, Congia M, Angius E, Nutland S, Bain SC, Barnett AH, Todd JA (2001) A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. Human Molecular Genetics 10:2025-2037
Cucca F, Todd J (1996) HLA susceptibility to type 1 diabetes: methods and mechanisms. In: Browning M, McMichaels A (eds) HLA and MHC: Genes, molecules and function. BIOS Scientific Publishers Ltd., Oxford, pp 383-406
Curtis D, North B, Sham P (2001) Use of an artificial neural network to detect association between a disease and multiple marker genotypes. Annals of Human Genetics 65:95-107
Cusick M, Klitgord N, Vidal M, Hill D (2005) Interactome: gateway into systems biology. Human Molecular Genetics 14:R171-R181
Dahlman I, Eaves I, Kosoy R, Morrison V, Heward J, Gough S, Allahabadia A, Franklyn J, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield J, Wang W, Cordell H, Walker N, Todd J, Concannon P (2002) Parameters for reliable results in genetic association studies in common disease. Nature Genetics 30:149-150
Dahlquist G, Blom L, Tuvemo T, Nyström L, Sandström A, Wall S (1989) The Swedish childhood diabetes study--results from a nine year case register and a one year case-referent study indicating that type 1 (insulin-dependent) diabetes mellitus is associated with both type 2 (non-insulin-dependent) diabetes mellitus and autoimmune disorders. Diabetologia 32:2-6
Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nature Genetics 29:229-232
Damcott CM, Pollin TI, Reinhart LJ, Ott SH, Shen H, Silver KD, Mitchell BD, Shuldiner AR (2006) Polymorphisms in the Transcription Factor 7-Like 2 (TCF7L2) Gene Are Associated With Type 2 Diabetes in the Amish: Replication and Evidence for a Role in Both Insulin Secretion and Insulin Resistance. Diabetes 55:2654-2659
Daniels SE, Bhattacharrya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le-S.öuef PN, Lathrop GM, Musk AW, Cookson WO (1996) A genome-wide search for quantitative trait loci underlying asthma. Nature 383:247-250
Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA (1994) A genome-wide search for human type-1 diabetes susceptibility genes. Nature 371:130-136
Davoodi-Semiromi A, Yang JJ, She JX (2002) IL-12p40 is associated with type 1 diabetes in Caucasian- American families. Diabetes 51:2334-2336
de Bakker PIW, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature Genetics 38:1166-1172
De La Vega FM, Dailey D, Ziegle J, Williams J, Madden D, Gilbert DA (2002) New generation pharmacogenomic tools: a SNP linkage disequilibrium Map, validated SNP assay resource, and high-throughput instrumentation system for large-scale genetic studies. BioTechniques Suppl:48-50, 52, 54
Delepine M, Nicolino M, Barrett T, Golamaully M, Lathrop GM, Julier C (2000) EIF2AK3, encoding translation initiation factor 2-α kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nature Genetics 25:406-409
Delepine M, Pociot F, Habita C, Hashimoto L, Froguel P, Rotter J, Cambon-Thomsen A, Deschamps I, Djoulah S, Weissenbach J, Nerup J, Lathrop M, Julier C (1997) Evidence of a non-MHC susceptibility locus in type I diabetes linked to HLA on chromosome 6. American Journal of Human Genetics 60:174-187
Deruytter N, Boulard O, Garchon HJ (2004) Mapping non-class IIH2-linked loci for type 1 diabetes in nonobese diabetic mice. Diabetes 53:3323-3327
Deutsch S, Lyle R, Dermitzakis E, Subrahmanyan L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel C, Antonarakis S (2005) Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. Human Molecular Genetics 14:3741-3749
Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997-1004
DeWan A, Liu M, Hartman S, Zhang SS-M, Liu DTL, Zhao C, Tam POS, Chan WM, Lam DSC, Snyder M, Barnstable C, Pang CP, Hoh J (2006) HTRA1 Promoter Polymorphism in Wet Age-Related Macular Degeneration. Science 314:989-992
Donath M, J S, Maedler K, Mandrup-Poulsen T (2003) Inflammatory mediators and islet beta-cell failure: a link between type 1 and type 2 diabetes. Journal of Molecular Medicine 81:455-470
Dorman J, Bunker C (2000) HLA-DQ locus of the human leukocyte antigen complex and type 1 diabetes mellitus: a HuGE review. Epidemiologic Reviews 22:218-227
Duerr R, Taylor K, Brant S, Rioux J, Silverberg M, Daly M, Steinhart A, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta L, Kistner E, Schumm L, Lee A, Gregersen P, Barmada M, Rotter J, Nicolae D, Cho J (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314:1461-1463
Durocher F, Faure R, Labrie Y, Pelletier L, Bouchard I, Laframboise R (2006) A novel mutation in the EIF2AK3 gene with variable expressivity in two patients with Wolcott-Rallison syndrome. Clinical Genetics 70:34-38
Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, Armstrong H, et al. (1996) A full genome search in multiple sclerosis. Nature Genetics 13:472-476
Editorial (1999) Freely associating. Nature Genetics 22:1-2
Eftychi C, Howson JM, Barratt BJ, Vella A, Payne F, Smyth DJ, Twells RC, Walker NM, Rance HE, Tuomilehto-Wolf E, Tuomilehto J, Undlien DE, R.ønningen KS, Guja C, Ionescu-T.îrgoviste C, Savage DA, Todd JA (2004) Analysis of the type 2 diabetes-associated single nucleotide polymorphisms in the genes IRS1, KCNJ11, and PPARG2 in type 1 diabetes. Diabetes 53:870-873
Eizirik DL, Sandler S (1989) Function and metabolism of pancreatic beta-cells maintained in culture following experimentally induced damage. Pharmacology & toxicology 65:163-168
ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306:636-640
Erlich H, Rotter J, Chang J, Shaw S, Raffel L, Klitz W, Bugawan T, Zeidler A (1996) Association of HLA-DPB1*0301 with IDDM in Mexican-Americans. Diabetes 45:610-614
Esposito K, Nappo F, Marfella R, Giugliano G, Giugliano F, Ciotola M, Quagliaro L, Ceriello A, Giugliano D (2002) Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans: role of oxidative stress. Circulation 106:2067-2072
Farrall M (1997) Affected sibpair linkage tests for multiple linked susceptibility genes. Genetic Epidemiology 14:103-115
Federici M, Petrone A, Porzio O, Bizzarri C, Lauro D, D'Alfonso R, Patera I, Cappa M, Nistico L, Baroni M, Sesti G, Di-Mario U, Lauro R, Buzzetti R (2003) The Gly972fwdarwArg IRS-1 variant is associated with type 1 diabetes in continental Italy. Diabetes 52:887-890
Feingold E, Lamb NE, Sherman SL (1995) Methods for genetic-linkage analysis using trisomies. American Journal of Human Genetics 56:475-483
Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nature Reviews Genetics 7:85-97
Fingerlin T, Boehnke M, Abecasis G (2004) Increasing the power and efficiency of disease-marker case-control association studies through use of allele-sharing information. American Journal of Human Genetics 74:432-443
Fischer LJ, Hamburger SA (1980) Inhibition of alloxan action in isolated pancreatic islets by superoxide dismutase, catalase, and a metal chelator. Diabetes 29:213-216
Florez JC, Jablonski KA, Bayley N, Pollin TI, de-Bakker PI, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D, Diabetes Prevention Program Research Group (2006) TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. New England Journal of Medicine 355:241-250
Franke L, van Bakel H, Diosdado B, van Belzen M, Wapenaar M, Wijmenga C (2004) TEAM: a tool for the integration of expression, and linkage and association maps. European Journal of Human Genetics 12:633-638
Franke L, van-Bakel H, Fokkens L, de-Jong ED, Egmont-Petersen M, Wijmenga C (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. American Journal of Human Genetics 78:1011-1025
Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nature Reviews Genetics 8:657-662
Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JRB, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316:889-894
Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nature Genetics 36:388-393
Fuqua JS (2000) Wolfram syndrome: Clinical and genetic aspects. Endocrinologist 10:51-59
Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The Structure of Haplotype Blocks in the Human Genome. Science 296:2225-2229

Gaffney P, Kearns G, Shark K, Ortmann W, Selby S, Malmgren M, Rohlf K, Ockenden T, Messner R, King R, Rich S, Behrens T (1998) A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proceedings of the National Academy of Sciences of the United States of America 95:14875-14879

Gambelunghe G, Ghaderi M, Cosentino A, Falorni A, Brunetti P, Falorni A, Sanjeevi C (2000) Association of MHC Class I chain-related A (MIC-A) gene polymorphism with Type I diabetes. Diabetologia 43:507-514

Gandhi TKB, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, Nagaraju S, Periaswamy B, Mishra G, Nandakumar K, Shen B, Deshpande N, Nayak R, Sarker M, Boeke JD, Parmigiani G, Schultz J, Bader JS, Pandey A (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nature Genetics 38:285-293

Gillespie KM, Dix RJ, Williams AJK, Newton R, Robinson ZF, Bingley PJ, Gale EAM, Shield JPH (2006) Islet Autoimmunity in Children With Down's Syndrome. Diabetes 55:3185-3188

Ginn LR, Lin JP, Plotz PH, Bale SJ, Wilder RL, Mbauya A, Miller FW (1998) Familial autoimmunity in pedigrees of idiopathic inflammatory myopathy patients suggests common genetic risk factors for many autoimmune diseases. Arthritis and Rheumatism 41:400-405

Gitton Y, Dahmane N, Baik S, Ruiz i Altaba A, Neidhardt L, Scholze M, Herrmann B, Kahlem P, Benkahla A, Schrinner S, Yildirimman R, Herwig R, Lehrach H, Yaspo M, initiative Hem (2002) A gene expression map of human chromosome 21 orthologues in the mouse. Nature 420:586-590

Glaser RL, Ramsay JP, Morison IM (2006) The imprinted gene and parent-of-origin effect database now includes parental origin of de novo mutations. Nucleic Acids Research 34:D29-31

Goldstein DB (2001) Islands of linkage disequilibrium. Nature Genetics 29:109-111

Gottlieb PA, Eisenbarth GS (1998) Diagnosis and treatment of pre-insulin dependent diabetes. Annual Review of Medicine 49:391-405

Gourh P, Tan FK, Assassi S, Ahn CW, McNearney TA, Fischbach M, Arnett FC, Mayes MD (2006) Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. Arthritis and Rheumatism 54:3945-3953

Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, et al. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nature Genetics 38:320-323

Green A, Brutti G, Patterson CC, Dahlquist G, Soltesz G, Green A, Schober E, et al. (2000) Variation and trends in incidence of childhood diabetes in Europe. Lancet 355:873-876

Grennan DM, Sanders PA, Thomson W, Dyer PA (1986) Rheumatoid arthritis: inheritance and association with other autoimmune diseases. Disease Markers 4:157-162

Gretarsdottir S, Sveinbjornsdottir S, Jonsson H, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, et al. (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. American Journal of Human Genetics 70:593-603

Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods in Molecular Biology 342:129-138

Grimble R (2002) Inflammatory status and insulin resistance. Current opinion in clinical nutrition and metabolic care 5:551-559

Groves CJ, Zeggini E, Minton J, Frayling TM, Weedon MN, Rayner NW, Hitman GA, Walker M, Wiltshire S, Hattersley AT, McCarthy MI (2006) Association Analysis of 6,736 U.K. Subjects Provides Replication and Confirms TCF7L2 as a Type 2 Diabetes Susceptibility Gene With a Substantial Effect on Individual Risk. Diabetes 55:2640-2644

Gupta M, Nikitina-Zake L, Zarghami M, Landin-Olsson M, Kockum I, Lernmark A, Sanjeevi C (2003) Association between the transmembrane region polymorphism of MHC class I chain related gene-A and type 1 diabetes mellitus in Sweden. Human Immunology 64:553-561

Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, PericakVance MA, et al. (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. Nature Genetics 13:469-471

Hakonarson H, Grant SFA, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu H-Q, Polychronakos C (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448:591-594

Hansen L, Eiberg H, Barrett T, Bek T, Kjaersgaard P, Tranebjaerg L, Rosenberg T (2005) Mutation analysis of the WFS1 gene in seven Danish Wolfram syndrome families; four new mutations identified. European Journal of Human Genetics 13:1275-1284

Hardy C, Khanim F, Torres R, Scott-Brown M, Seller A, Poulton J, Collier D, Kirk J, Polymeropoulos M, Latif F, Barrett T (1999) Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in WFS1. American Journal of Human Genetics 65:1279-1290

Harjutsalo V, Podar T, Tuomilehto J (2005) Cumulative Incidence of Type 1 Diabetes in 10,168 Siblings of Finnish Young-Onset Type 1 Diabetic Patients. Diabetes 54:563-569

Hartland S, Newton JL, Griffin SM, Donaldson PT (2004) A functional polymorphism in the interleukin-1 receptor-1 gene is associated with increased risk of Helicobacter pylori infection but not with gastric cancer. Digestive Diseases and Sciences 49:1545-1550

Hasegawa K, Martin F, Huang G, Tumas D, Diehl L, Chan A (2004) PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. Science 303:685-689

Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon-Thomsen A, Deschamps I, Rotter JI, Djoulah S, James MR (1994) Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371:161-164

Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, et al. (2000) The DNA sequence of human chromosome 21. Nature 405:311-319

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani N, et al. (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nature Genetics 36:233-239

Helms C, Cao L, Krueger JG, Wijsman EM, Chamian F, Gordon D, Heffernan M, Daw JAW, Robarge J, Ott J, Kwok PY, Menter A, Bowcock AM (2003) A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nature Genetics 35:349-356

Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu XF, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF (2006) A common genetic variant is associated with adult and childhood obesity. Science 312:279-283

Herr M, Dudbridge F, Zavattari P, Cucca F, Guja C, March R, Campbell R, Barnett A, Bain S, Todd J, Koeleman B (2000) Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of IDDM1 in the HLA region on chromosome 6p21. Human Molecular Genetics 9:1291-1301

Hinds D, Kloek A, Jen M, Chen X, Frazer K (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nature Genetics 38:82-85

Hirschhorn J (2003) Genetic epidemiology of type 1 diabetes. Pediatric Diabetes 4:87-100

Hirschhorn J, Daly M (2005) Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6:95-108

Hirschhorn J, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genetics in Medicine 4:45-61

Hoheisel J (2006) Microarray technology: beyond transcript profiling and genotype analysis. Nature Reviews Genetics 7:200-210

Hollegaard MV, Bidwell JL (2006) Cytokine gene polymorphism in human disease: on-line databases, Supplement 3. Genes and Immunity 7:269-276

Holm P, Luthman H, Kockum I (2003) No evidence for linkage in Swedish multiplex T1DM families to IL12B on chromosome 5q33-34. Annals of the New York Academy of Sciences 1005:352-355

Horikawa Y, Oda N, Cox NJ, Li XQ, Orho-Melander M, Hara M, Hinokio Y, et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nature Genetics 26:163-175

Houldsworth A, Metzner M, Rossol S, Shaw S, Kaminski E, Demaine AG, Cramp ME (2005) Polymorphisms in the IL-12B gene and outcome of HCV infection. Journal of Interferon and Cytokine Research 25:271-276

Huang D, Cancilla MR, Morahan G (2000) Comlete primary structure, chromosomal localisation, and definition of polymorphisms of the gene encoding the human interleukin-12 p40 subunit. Genes and Immunity 1:515-520

Hug BA, Ahmed N, Robbins JA, Lazar MA (2004) A chromatin immunoprecipitation screen reveals protein kinase C beta as a direct RUNX1 target gene. Journal of Biological Chemistry 279:825-830

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599-603

Hugot J-P, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas J-L, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel J-F, Thomas G (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 379:821-823

Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J (2003) Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes 52:1052-1055

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nature Genetics 36:949-951

Inoue H, Tanizawa Y, Wasson J, Behn P, Kalidas K, Bernal-Mizrachi E,

Mueckler M, Marshall H, Donis-Keller H, Crock P, Rogers D, Mikuni M, Kumashiro H, Higashi K, Sobue G, Oka Y, Permutt MA (1998) A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nature Genetics 20:143-148

Ishihara H, Takeda S, Tamura A, Takahashi R, Yamaguchi S, Takei D, Yamada T, Inoue H, Soga H, Katagiri H, Tanizawa Y, Oka Y (2004) Disruption of the WFS1 gene in mice causes progressive β -cell loss and impaired stimulus – Secretion coupling in insulin secretion. Human Molecular Genetics 13:1159-1170

Ivarsson S, Ericsson U, Gustafsson J, Forslund M, Vegfors P, Anneren G (1997) The impact of thyroid autoimmunity in children and adolescents with Down syndrome. Acta paediatrica (Oslo, Norway 86:1065-1067

Iyer S, Korada M, Rainbow L, Kirk J, Brown RM, Shaw N, Barrett TG (2004) Wolcott-Rallison syndrome: A clinical and genetic study of three children, novel mutation in EIF2AK3 and a review of the literature. Acta Paediatrica, International Journal of Paediatrics 93:1195-1201

Jagiello P, Aries P, Arning L, Wagenleiter SEN, Csernok E, Hellmich B, Gross WL, Epplen JT (2005) The PTPN22 620W allele is a risk factor for Wegener's granulomatosis. Arthritis and Rheumatism 52:4039-4043

Jawaheer D, Seldin M, Amos C, Chen W, Shigeta R, Etzel C, Damle A, et al. (2003) Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis and Rheumatism 48:906-916

Jeremiah DE, Leyshon GE, Rose T, Francis HW, Elliott RW (1973) Down's syndrome and diabetes. Psychological medicine 3:455-457

Johansen A, Jensen DP, Bergholdt R, Mortensen HB, Pociot F, Nerup J, Hansen T, Pedersen O (2006) IRS1, KCNJ11, PPARgamma2 and HNF-1alpha: do amino acid polymorphisms in these candidate genes support a shared aetiology between type 1 and type 2 diabetes? Diabetes, Obesity and Metabolism 8:75-82

Johansson S, Lie BA, Thorsby E, Undlien DE (2001) The polymorphism in the 3' untranslated region of IL12B has a negligible effect on the susceptibility to develop type 1 diabetes in Norway. Immunogenetics 53:603-605

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human MicroRNA targets. PLoS biology 2:e363

John B, Sander C, Marks DS (2006) Prediction of human microRNA targets. Methods in Molecular Biology 342:101-113

Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, Di-Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nature Genetics 29:233-237

Kahlem P, Sultan M, Herwig R, Steinfath M, Balzereit D, Eppens B, Saran N, Pletcher M, South S, Stetten G, Lehrach H, Reeves R, Yaspo M (2004) Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome. Genome Research 14:1258-1267

Karasik A, O'Hara C, Srikanta S, Swift M, Soeldner JS, Kahn CR, Herskowitz RD (1989) Genetically programmed selective islet beta-cell loss in diabetic subjects with Wolfram's syndrome. Diabetes Care 12:135-138

Karlsson B, Gustafsson J, Hedov G, Ivarsson S, Anneren G (1998) Thyroid dysfunction in Down's syndrome: relation to age and thyroid autoimmunity. Archives of disease in childhood 79:242-245

Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J (2000) Incidence of childhood type 1 diabetes worldwide. Diabetes Care 23:1516-1526

Ke X, Cardon LR (2003) Efficient selective screening of haplotype tag SNPs. Bioinformatics 19:287-288

Kim F, Gallis B, Corson M (2001a) TNF-alpha inhibits flow and insulin signaling leading to NO production in aortic endothelial cells. American journal of physiology. Cell physiology 280:C1057-C1065

Kim J, Sen S, Avery C, Simpson E, Chandler P, Nishina P, Churchill G, Naggert J (2001b) Genetic analysis of a new mouse model for non-insulin- dependent diabetes. Genomics 74:273-286

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385-389

Knip M, Veijola W, Virtanen SM, Hyoty H, Vaarala O, Akerblom HK (2005) Environmental triggers and determinants of type 1 diabetes. Diabetes 54:S125-S136

Koeleman B, Lie B, Undlien D, Dudbridge F, Thorsby E, de Vries R, Cucca F, Roep B, Giphart M, Todd J (2004) Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. Genes and Immunity 5:381-388

Kono DH, Burlingame RW, Owens DG, Kuramochi A, Balderas RS, Balomenos D, Theofilopoulos AN (1994) Lupus susceptibility loci in New Zealand mice. Proceedings of the National Academy of Sciences of the United States of America 91:10168-10172

Kosoy R, Yokoi N, Seino S, Concannon P (2004) Polymorphic variation in the CBLB gene in human type 1 diabetes. Genes and Immunity 5:232-235

Kristiansen OP, Larsen ZM, Pociot F (2000a) CTLA-4 in autoimmune diseases – a general susceptibility gene to autoimmunity? Genes and Immunity 1:170-184

Kristiansen OP, Pociot F, Johannesen J, Bergholdt R, Dinarello CA, Nerup J, Mandrup-Poulsen T (2000b) Linkage disequilibrium testing of four interleukin-1 gene- cluster polymorphisms in Danish multiplex families with insulin-dependent diabetes mellitus. Cytokine 12:171-175

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genetics 22:139-144

Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nature Genetics 27:234-236

Kubaszek A, Pihlajamaki J, Komarovski V, Lindi V, Lindstrom J, Eriksson J, Valle T, Hamalainen H, Ilanne-Parikka P, Keinanen-Kiukaanniemi S, Tuomilehto J, Uusitupa M, Laakso M, Study FDP (2003) Promoter polymorphisms of the TNF-alpha (G-308A) and IL-6 (C-174G) genes predict the conversion from impaired glucose tolerance to type 2 diabetes: the Finnish Diabetes Prevention Study. Diabetes 52:1872-1876

Kubisch HM, Wang J, Bray TM, Phillips JP (1997) Targeted overexpression of Cu/Zn superoxide dismutase protects pancreatic beta-cells against oxidative stress. Diabetes 46:1563-1566

Kuida S, Beier D (2000) Genetic localization of interacting modifiers affecting severity in a murine model of polycystic kidney disease. Genome Research 10:49-54

Kuokkanen S, Sundvall M, Terwilliger JD, Tienari PJ, Wikstrom J, Holmdahl R, Pettersson U, Peltonen L (1996) A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus Eae2. Nature Genetics 13:477-480

Kyogoku C, Langefeld C, Ortmann W, Lee A, Selby S, Carlton V, Chang M, Ramos P, Baechler E, Batliwalla F, Novitzke J, Williams A, Gillett C, Rodine P, Graham R, Ardlie K, Gaffney P, Moser K, Petri M, Begovich A, Gregersen P, Behrens T (2004) Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. American Journal of Human Genetics 75:504-507

Lage K, Karlberg E, Størling Z, Olason P, Pedersen A, Rigina O, Hinsby A, Tümer Z, Pociot F, Tommerup N, Moreau Y, Brunak S (2007a) A human phenome-interactome network of protein complexes implicated in genetic disorders. Nature Biotechnology 25:309-316

Lage K, Karlberg E, Størling Z, Òlason P, Pedersen A, Rigina O, Hinsby A, Tümer Z, Pociot F, Tommerup N, Moreau Y, Brunak S (2007b) A human phenome-interactome network of protein complexes in genetic disorders. Nature Biotechnology In press

Laird N, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nature Reviews Genetics 7:385-394

Lambert A, Gillespie K, Thomson G, Cordell H, Todd J, Gale E, Bingley P (2004) Absolute risk of childhood-onset type 1 diabetes defined by human leukocyte antigen class II genotype: a population-based study in the United Kingdom. Journal of Clinical Endocrinology and Metabolism 89:4037-4043

Lamont AG, Adorini L (1996) IL-12 – a key cytokine in immune regulation. Immunology Today 17:214-217

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature genetics 11:241-247

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037-2048

Lang J, Bellgrau D (2004) Animal models of type 1 diabetes: genetics and immunological function. Advances in Experimental Medicine and Biology 552:91-9116

Larsen ZM, Johannesen J, Kristiansen OP, Nerup J, Pociot F, Danish I, Epidemiology and Genetics Group., Danish SGoIiC (2004) Evidence for linkage on chromosome 4p16.1 in Type 1 diabetes Danish families and complete mutation scanning of the WFS1 (Wolframin) gene. Diabetic medicine 21:218-222

Latek R, Suri A, Petzold S, Nelson C, anagawa O, Unanue E, Fremont D (2000) Structural basis of peptide binding and presentation by the type I diabetes-associated MHC class II molecule of NOD mice. Immunity 12:699-710

Lee AT, Li W, Liew A, Bombardier C, Weisman M, Massarotti EM, Kent J, Wolfe F, Begovich AB, Gregersen PK (2005) The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. Genes and Immunity 6:129-133

Lee B, Gai W, Laychock SG (2001a) Proteasomal activation mediates down-regulation of inositol 1,4,5-trisphosphate receptor and calcium mobilization in rat pancreatic islets. Endocrinology 142:1744-1751

Lee K, Wucherpfennig K, Wiley D (2001b) Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type I diabetes. Nature Immunology 2:501-507

Leiter EH, von-Herrath M (2004) Animal models have little to teach us about type 1 diabetes: 2. In opposition to this proposal. Diabetologia 47:1657-1660

Li H, Lindholm E, Almgren P, Gustafsson A, Forsblom C, Groop L, Tuomi T (2001) Possible Human Leukocyte Antigen-Mediated Genetic Interaction between Type 1 and Type 2 Diabetes. Journal of Clinical Endocrinology and Metabolism 86:574-582

Li J, Burmeister M (2005) Genetical genomics: combining genetics with gene expression analysis. Human Molecular Genetics 14:R163-R169
Liblau RS, Singer SM, McDevitt HO (1995) Th1 and Th2 CD4(+)T-cells in the pathogenesis of organ- specific autoimmune-diseases. Immunology Today 16:34-38
Lie B, Akselsen H, Joner G, Dahl-Jorgensen K, Ronningen K, Thorsby E, Undlien D (1997) HLA associations in insulin-dependent diabetes mellitus: no independent association to particular DP genes. Human Immunology 55:170-175
Lohmueller K (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 33:177-182
Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288:136-140
Lorini R, d'Annunzio G, Vitali L, Scaramuzza A (1996) IDDM and autoimmune thyroid disease in the pediatric age group. Journal of Pediatric Endocrinology and Metabolism 9:89-94
Lucek P, Hanke J, Reich J, Solla S, Ott J (1998) Multi-locus nonparametric linkage analysis of complex trait loci with neural networks. Human Heredity 48:275-284
Maedler K, Sergeev P, Ris F, Oberholzer J, Joller-Jemelka H, Spinas G, Kaiser N, Halban P, Donath M (2002) Glucose-induced beta cell production of IL-1beta contributes to glucotoxicity in human pancreatic islets. Journal of Clinical Investigation 110:851-860
Malek TR, Bayer AL (2004) Tolerance, not immunity, crucially depends on IL-2. Nature Reviews Immunology 4:665-674
Mandrup-Poulsen T (1996) The role of interleukin-1 in the pathogenesis of IDDM. Diabetologia 39:1005-1029
Manetti R, Parronchi P, Giudizi MG, Piccinni MP, Maggi E, Trinchieri G, Romagnani S (1993) Natural-killer-cell stimulatory factor (Interleukin-12 [IL- 12]) induces T-helper type-1 (Th1)-specific immune-responses and inhibits the development of IL-4-producing Th2 cells. Journal of Experimental Medicine 177:1199-1204
Maraganore DM, De-Andrade M, Lesnick TC, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DC (2005) High-resolution whole-genome association study of Parkinson disease. American Journal of Human Genetics 77:685-693
Marinov M, Weeks D (2001) The complexity of linkage analysis with neural networks. Human Heredity 51:169-176
Martin S, Kardorf J, Schulte B, Lampeter EF, Gries FA, Melchers I, Wagner R, Bertrams J, Roep BO, Pfützner A (1995) Autoantibodies to the islet antigen ICA69 occur in IDDM and in rheumatoid arthritis. Diabetologia 38:351-355
Mathews CE (2005) Utility of murine models for the study of spontaneous autoimmune type 1 diabetes. Pediatric Diabetes 6:165-177
Mathews CE, Dunn BD, Hannigan MO, Huang CK, Leiter EH (2002) Genetic control of neutrophil superoxide production in diabetes-resistant ALR/Lt mice. Free Radical Biology and Medicine 32:744-751
Mathews CE, Leiter EH (1999) Constitutive differences in antioxidant defense status distinguish alloxan-resistant and alloxan-susceptible mice. Free Radical Biology and Medicine 27:449-455
Mathis D, Vence L, Benoist C (2001) Beta-Cell death during progression to diabetes. Nature 414:792-798
Matthews D, Fry L, Powles A, Weber J, McCarthy M, Fisher E, Davies K, Williamson R (1996) Evidence that a locus for familial psoriasis maps to chromosome 4q. Nature Genetics 14:231-233
McCarroll S, Hadnott T, Perry G, Sabeti P, Zody M, Barrett J, Dallaire S, Gabriel S, Lee C, Daly M, Altshuler D, Consortium IH (2006) Common deletion polymorphisms in the human genome. Nature Genetics 38:86-92
McCarthy JJ, Meyer J, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ, Gen Quest Multicenter Study (2003) Evidence for substantial effect modification by gender in a large-scale genetic association study of the metabolic syndrome among coronary heart disease patients. Human Genetics 114:87-98
McCarthy MI, Groop PH, Hansen T (2005) Making the right associations. Diabetologia 48:1241-1243
McCormack RM, Maxwell AP, Carson DJ, Patterson CC, Middleton D, Savage DA (2002) The IL12B 3' untranslated region DNA polymorphism is not associated with early-onset type 1 diabetes. Genes and Immunity 3:433-435
McGinnis R, Shifman S, Darvasi A (2002) Power and efficiency of the TDT and case-control design for association scans. Behavior Genetics 32:135-144
Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Pritchard LE, Cucca F, Barnett AH, Bain SC, Todd JA (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nature genetics 19:297-300
Melanitou E (2005) The autoimmune contrivance: Genetics in the mouse model. Clinical Immunology 117:195-206
Metcalfe KA, Hitman GA, Pociot F, Bergholdt R, Tuomilehtowolf E, Tuomilehto J, Viswanathan M, et al. (1996) An association between type-1 diabetes and the interleukin-1 receptor-type-1 gene. Human Immunology 51:41-48
Meyer G, Badenhoop K (2002) Autoimmune regulator (AIRE) gene on chromosome 21: Implications for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) any-more common manifestations of endocrine autoimmunity. Journal of Endocrinological Investigation 25:804-811
Meyer G, Donner H, Herwig J, Bohles H, Usadel KH, Badenhoop K (2001) Screening for an AIRE-1 mutation in patients with Addison's disease, type 1 diabetes, Graves' disease and Hashimoto's thyroiditis as well as in APECED syndrome. Clinical Endocrinology 54:335-338
MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921-923
Milunsky A, Neurath PW (1968) Diabetes mellitus in Down's Syndrome. Archives of environmental health 17:372-376
Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, Barrett T, Frayling TM (2002) Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. Diabetes 51:1287-1290
Moffatt MF, Traherne JA, Abecasis GR, Cookson WOCM (2000) Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. Human Molecular Genetics 9:1011-1019
Mølbak AG, Christau B, Marner B, Borch-Johnsen K, Nerup J (1994) Incidence of insulin-dependent diabetes mellitus in age groups over 30 years in Denmark. Diabetic Medicine 11:650-655
Momiyama Y, Ohmori R, Nagano M, Kato R, Taniguchi H, Egashira T, Nakamura H, Ohsuzu F (2005) Polymorphism of the 3′-untranslated region of interleukin-12 p40 gene is not associated with the presence or severity of coronary artery disease. Circulation Journal 69:793-797
Morahan G, Boutlis CS, Huang D, Pain A, Saunders JR, Hobbs MR, Granger DL, Weinberg JB, Peshu N, Mwaikambo ED, Marsh K, Roberts DJ, Anstey NM (2002a) A promoter polymorphism in the gene encoding interleukin-12 p40 (IL12B) is associated with mortality from cerebral malaria and with reduced nitric oxide production. Genes and Immunity 3:414-418
Morahan G, Huang DX, Wu M, Holt BJ, White GP, Kendall GE, Sly PD, Holt PG (2002b) Association of IL12B promoter polymorphism with severity of atopic and non-atopic asthma in children. Lancet 360:455-459
Morahan G, Huang DX, Ymer SI, Cancilla MR, Stephen K, Dabadghao P, Werther G, Tait BD, Harrison LC, Colman PG (2001) Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. Nature Genetics 27:218-221
Mordes JP, Bortell R, Blankenhorn EP, Rossini AA, Greiner DL (2004) Rat models of type 1 diabetes: genetics, environment, and autoimmunity. ILAR journal / National Research Council, Institute of Laboratory Animal Resources 45:278-291
Morel L, Rudofsky UH, Longmate JA, Schiffenbauer J, Wakeland EK (1994) Polygenic control of susceptibility to murine systemic lupus erythematosus. Immunity 1:219-229
Morley M, Molony C, Weber T, Devlin J, Ewens K, Spielman R, Cheung V (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430:743-747
Morrison VA, Onengut-Gumuscu S, Concannon P (2004) A functional variant of IRS1 is associated with type 1 diabetes in families from the US and UK. Molecular Genetics and Metabolism 81:291-294
Moser KL, Neas BR, Salmon JE, Yu H, Gray-McGuire C, Asundi N, Bruner GR, Fox J, Kelly J, Henshall S, Bacino D, Dietz M, Hogue R, Koelsch G, Nightingale L, Shaver T, Abdou NI, Albert DA, Carson C, Petri M, Treadwell EL, James JA, Harley JB (1998) Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. Proceedings of the National Academy of Sciences of the United States of America 95:14869-14874
Muller S, Martin S, Koenig W, Hanifi-Moghaddam P, Rathmann W, Haastert B, Giani G, Illig T, Thorand B, Kolb H (2002) Impaired glucose tolerance is associated with increased serum concentrations of interleukin 6 and co-regulated acute-phase proteins but not TNF-alpha or its receptors. Diabetologia 45:805-812
Naik RG, Palmer JP (2003) Latent autoimmune diabetes in adults (LADA). Reviews in Endocrine & Metabolic Disorders 4:233-241
Nakamura A, Shimizu C, Nagai S, Taniguchi S, Umetsu M, Atsumi T, Wada N, Yoshioka N, Ono Y, Tanizawa Y, Koike T (2006) A novel mutation of WFS1 gene in a Japanese man of Wolfram syndrome with positive diabetes-related antibodies. Diabetes Research and Clinical Practice 73:215-217
Namkung J, Kim Y, Park T (2005) Whole-genome association studies of alcoholism with loci linked to schizophrenia susceptibility. BMC Genetics 6:S9
Nerup J, Mandrup-Poulsen T, Helqvist S, Andersen HU, Pociot F, Reimers JI, Cuartero BG, Karlsen AE, Bjerre U, Lorenzen T (1994) On the pathogenesis of IDDM. Diabetologia 37:S82-S89
Nerup J, Platz P, Andersen O, Christy M, Lyngsoe J, Poulsen J, Ryder L, Nielsen L, Thomsen M, Svejgaard A (1974) HL-A antigens and diabetes mellitus. The Lancet 2:864-866
Nerup J, Pociot F, European Consortium for IDDM genome studies. Members: Holm.P JC, Kochum.I, Senee.V, Blanc.H, Papp.J, Åkesson.K, Bartso-

cas.C, deLeiva.A, Dahlquist.G, Rønningen.KS, Lathrop.M, Luthman.H, Pociot.F, Nerup.J. (2001) A genomewide scan for Type 1-diabetes susceptibility in Scandinavian families: Identification of new loci with evidence of interactions. American Journal of Human Genetics 69:1301-1313

Nguyen C, Limaye N, Wakeland E (2002) Susceptibility genes in the pathogenesis of murine lupus. Arthritis Research 4 Suppl 3:S255-S263

Nielsen C, Hansen D, Husby S, Jacobsen BB, Lillevang ST (2003) Association of a putative regulatory polymorphism in the PD-1 gene with susceptibility to type 1 diabetes. Tissue Antigens 62:492-497

Nielsen K, Kruhoffer M, Orntoft T, Sparre T, Wang H, Wollheim C, Jorgensen M, Nerup J, Karlsen A (2004) Gene expression profiles during beta cell maturation and after IL-1 beta exposure reveal important roles of Pdx-1 and Nkx6.1 for IL-1 beta sensitivity. Diabetologia 47:2185-2199

Nistico L, Buzzetti R, Pritchard L, Van der Auwera B, Giovannini C, Bosi E, Larrad M, Rios M, Chow C, Cockram C, Jacobs K, Mijovic C, Bain S, Barnett A, Vandewalle C, Schuit F, Gorus F, Tosi R, Pozzilli P, Todd J (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Human Molecular Genetics 5:1075-1080

Nistico L, Giorgi G, Giordano M, Galgani A, Petrone A, D'Alfonso S, Federici M, Di-Mario U, Pozzilli P, Buzzetti R, Cascino I (2002) IL12B polymorphism and type 1 diabetes in the Italian population – A case-control study. Diabetes 51:1649-1650

Nithiyananthan R, Heward JM, Allahabadia A, Barnett AH, Franklyn JA, Gough SCL (2000) A heterozygous deletion of the autoimmune regulator (AIRE1) gene, autoimmune thyroid disease, and type 1 diabetes: No evidence for association. Journal of Clinical Endocrinology and Metabolism 85:1320-1322

Noble J, Valdes A, Thomson G, Erlich H (2000) The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. Diabetes 49:121-125

Oberley LW (1988) Free radicals and diabetes. Free radical biology & medicine 5:113-124

O'Doherty A, Ruf S, Mulligan C, Hildreth V, Errington ML, Cooke S, Sesay A, Modino S, Vanes L, Hernandez D, Linehan JM, Sharpe PT, Brandner S, Bliss TVP, Henderson DJ, Nizetic D, Tybulewicz VLJ, Fisher EMC (2005) An Aneuploid Mouse Strain Carrying Human Chromosome 21 with Down Syndrome Phenotypes. Science 309:2033-2037

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411:603-606

Olivier M, Bustos VI, Levy MR, Smick GA, Moreno I, Bushard JM, Almendras AA, Sheppard K, Zierten DL, Aggarwal A, Carlson CS, Foster BD, Vo N, Kelly L, Liu X, Cox DR (2001) Complex high-resolution linkage disequilibrium and haplotype patterns of single-nucleotide polymorphisms in 2.5 Mb of sequence on human chromosome 21. Genomics 78:64-72

Olson L, Roper R, Baxter L, Carlson E, Epstein C, Reeves R (2004) Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. Developmental Dynamics 230:581-589

Onengut-Gumuscu S, Buckner JH, Concannon P (2006) A Haplotype-Based Analysis of the PTPN22 Locus in Type 1 Diabetes. Diabetes 55:2883-2889

Onengut-Gumuscu S, Ewens KG, Spielman RS, Concannon P (2004) A functional polymorphism (1858C/T) in the PTPN22 gene is linked and associated with type I diabetes in multiplex families. Genes and Immunity 5:678-680

Onkamo P, Karvonen M, Tuomilehto J (1999) Worldwide increase in incidence of Type I diabetes--the analysis of the data on published incidence trends. Diabetologia 42:1395-1403

Orozco G, Gonzalez-Gay MA, Paco L, Lopez-Nevot MA, Guzman M, Pascual-Salcedo D, Balsa A, Martin J (2005a) Interleukin 12 (IL12B) and interleukin 12 receptor (IL12RB1) gene polymorphisms in rheumatoid arthritis. Human Immunology 66:711-715

Orozco G, Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, Ortego-Centeno N, Jimenez-Alonso J, Pascual-Salcedo D, Balsa A, de-Pablo R, Nunez-Roldan A, Gonzalez-Escribano MF, Martin J (2005b) Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis and Rheumatism 52:219-224

Oti M, Snel B, Huynen MA, Brunner HG (2006) Predicting disease genes using protein-protein interactions. Journal of Medical Genetics 43:691-698

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nature Genetics 32:650-654

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, C.R. H, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BTN, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SPA, Cox DR (2001) Blocks of limited haplotype diversity revealed by high- resolution scanning of human chromosome 21. Science 294:1719-1723

Patterson D, Costa ACS (2005) Down syndrome and genetics – A case of linked histories. Nature Reviews Genetics 6:137-147

Payami H, Joe S, Thomson G (1989) Autoimmune thyroid disease in type I diabetic families. Genetic Epidemiology 6:137-141

Payne F, Smyth DJ, Pask R, Barratt BJ, Cooper JD, Twells RCJ, Walker NM, Lam AC, Smink LJ, Nutland S, Rance HE, Todd JA (2004) Haplotype tag single nucleotide polymorphism analysis of the human orthologues of the rat type 1 diabetes genes Ian4 (Lyp/Iddm1) and Cblb. Diabetes 53:505-509

Pearce SHS, Cheetham T, Imrie H, Vaidya B, Barnes ND, Bilous RW, Carr D, Meeran K, Shaw NJ, Smith CS, Toft AD, Williams G, Kendall-Taylor P (1998) A common and recurrent 13-bp deletion in the autoimmune regulator gene in British kindreds with autoimmune polyendocrinopathy type 1. American Journal of Human Genetics 63:1675-1684

Peng JC, Abu-Bakar S, Richardson MM, Jonsson JJ, Frazer IH, Nielsen LK, Morahan G, Thomas R (2006) IL10 and IL12B polymorphisms each influence IL-12p70 secretion by dendritic cells in response to LPS. Immunology and Cell Biology 84:227-232

Pennacchio LA, Rubin EM (2001) Genomic strategies to identify mammalian regulatory sequences. Nature Reviews Genetics 2:100-109

Perheentupa J (1996) Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (apeced). Hormone and Metabolic Research 28:353-356

Pickup J (2004) Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. Diabetes Care 27:813-823

Pickup J, Mattock M, Chusney G, Burt D (1997) NIDDM as a disease of the innate immune system: association of acute-phase reactants and interleukin-6 with metabolic syndrome X. Diabetologia 40:1286-1292

Pickup J, Williams G (1991) Textbook of Diabetes. Vol 1. Blackwell Scientific Publications

Pietropaolo M, Barinas-Mitchell E, Pietropaolo S, Kuller L, Trucco M (2000) Evidence of islet cell autoimmunity in elderly patients with type 2 diabetes. Diabetes 49:32-38

Pociot F (1996) Insulin-dependent diabetes-mellitus – a polygenic disorder. Danish Medical Bulletin 43:216-248

Pociot F (2002) CTLA4 gene and autoimmune endocrinopathies: A new marker? Journal of Endocrinological Investigation 25:1001-1005

Pociot F (2004) CTLA-4 in Autoimmune Disease. Landes Bioscience, Georgetown, Texas, USA.

Pociot F, Karlsen AE, Pedersen CB, Aalund M, Nerup J (2004) Novel analytical methods applied to type 1 diabetes genome scan data. American Journal of Human Genetics 74:647-660

Pociot F, McDermott MF (2002) Genetics of Type 1 diabetes mellitus. Genes and Immunity 3:235-249

Pociot F, Molvig J, Wogensen L, Worsaae H, Dalboge H, Baek L, Nerup J (1991) A tumour necrosis factor beta gene polymorphism in relation to monokine secretion and insulin-dependent diabetes mellitus. Scandinavian journal of immunology 33:37-49

Pociot F, Wilson AG, Nerup J, Duff GW (1993) No independent association between a tumor necrosis factor-alpha promotor region polymorphism and insulin-dependent diabetes mellitus. European journal of immunology 23:3050-3053

Polak M, Shield J (2004) Neonatal and very-early-onset diabetes mellitus. Seminars in Neonatology 9:59-65

Polymeropoulos MH, Swift RG, Swift M (1994) Linkage of the gene for Wolfram syndrome to markers on the short arm of chromosome 4. Nature Genetics 8:95-97

Porter JR, Barrett TG (2005) Monogenic syndromes of abnormal glucose homeostasis: clinical review and relevance to the understanding of the pathology of insulin resistance and beta cell failure. Journal of Medical Genetics 42:893-902

Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, Stoffel M (2004) A pancreatic islet-specific microRNA regulates insulin secretion. Nature 432:226-230

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics 65:220-228

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. American Journal of Human Genetics 67:170-181

Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, Brookes AJ, Tentler D, Kristjansdottir H, Grondal G, Bolstad AI, Svenungsson E, Lundberg I, Sturfelt G, Jonssen A, Truedsson L, Lima G, Alcocer-Varela J, Jonsson R, Gyllensten UB, Harley JB, Alarcon-Segovia D, Steinsson K, Alarcon-Riquelme ME (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nature Genetics 32:666-669

Przeworski M, Hudson RR, Di-Rienzo A (2000) Adjusting the focus on human variation. Trends in Genetics 16:296-302

Pugliese A, Zeller M, Fernandez A, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD (1997) The insulin gene is transcribed in the human thymus and transcription levels correlate with

allelic variation at the ins vntr-iddm2 susceptibility locus for type-1 diabetes. Nature Genetics 15:293-297
Rabinovitch A, Suarez-Pinzon WL (1998) Cytokines and their roles in pancreatic islet beta-cell destruction and insulin-dependent diabetes mellitus. Biochemical Pharmacology 55:1139-1149
Rando TA, Horton JC, Layzer RB (1992) Wolfram syndrome: evidence of a diffuse neurodegenerative disease by magnetic resonance imaging. Neurology 42:1220-1224
Randolph AG, Lange C, Silverman EK, Lazarus R, Silverman ES, Raby B, Brown A, Ozonoff A, Richter B, Weiss ST (2004) The IL12B gene is associated with asthma. American Journal of Human Genetics 75:709-715
Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199-204
Reich DE, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. Genetic Epidemiology 20:4-16
Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. Nature Genetics 32:135-142
Reik W, Walter J (2001) Genomic imprinting: Parental influence on the genome. Nature Reviews Genetics 2:21-32
Remmers EF, Longman RE, Du Y, O'Hare A, Cannon GW, Griffiths MM, Wilder RL (1996) A genome scan localizes five non-MHC loci controlling collagen-induced arthritis in rats. Nature Genetics 14:82-85
Reymond A, Marigo V, Yaylaoglu M, Leoni A, Ucla C, Scamuffa N, Cacciappoli C, Dermitzakis E, Lyle R, Banfi S, Eichele G, Antonaraki sS, Ballabio A (2002) Human chromosome 21 gene expression atlas in the mouse. Nature 420:582-586
Rich S (1990) Mapping genes in diabetes. Genetic epidemiological perspective. Diabetes 39:1315-1319
Richards EJ (2006) Inherited epigenetic variation – Revisiting soft inheritance. Nature Reviews Genetics 7:395-401
Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. American journal of human genetics 40:1-14
Risch N, Merikangas K (1996) The future of genetic-studies of complex human-diseases. Science 273:1516-1517
Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847-856
Ritchie M, White B, Parker J, Hahn L, Moore J (2003a) Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. BMC bioinformatics [electronic resource] 4:28
Ritchie MD, Hahn LW, Moore JH (2003b) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genetic epidemiology 24:150-157
Roach JC, Deutsch K, Li S, Siegel AF, Bekris LM, Einhaus DC, Sheridan CM, Glusman G, Hood L, Lernmark A, Janer M, Group SCDSGatDIiSS (2006) Genetic Mapping at 3-Kilobase Resolution Reveals Inositol 1,4,5-Triphosphate Receptor 3 as a Risk Factor for Type 1 Diabetes in Sweden. American Journal of Human Genetics 79:614-627
Robbins MJ, Sharp RA, Slonim AE, Burr IM (1980) Protection against streptozotocin-induced diabetes by superoxide dismutase. Diabetologia 18:55-58
Roep BO (1996) T-cell responses to autoantigens in IDDM. The search for the Holy Grail. Diabetes 45:1147-1156
Ronningen K, Keiding N, Green A, Group. EAS (2001) Correlations between the incidence of childhood-onset type I diabetes in Europe and HLA genotypes. Diabetologia 44 Suppl 3:B51-B59
Rose NR (1997) Autoimmune diseases: Tracing the shared threads. Hospital Practice 32:147-154
Rudd MF, Webb EL, Matakidou A, Sellick GS, Williams RD, Bridle H, Eisen T, Houlston RS, *-G.E.L.C.A.P.S. C (2006) Variants in the GH-IGF axis confer susceptibility to lung cancer. Genome Research 16:693-701
Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-933
Salomon B, Lenschow DJ, Rhee L, Ashourian N, Singh B, Sharpe A, Bluestone JA (2000) B7/CD28 costimulation is essential for the homeostasis of the CD4+CD25+ immunoregulatory T cells that control autoimmune diabetes. Immunity 12:431-440
Sanjeevi CB, Kanungo A, Berzina L, Shtauvere-Brameus A, Ghaderi M, Samal KC (2002) MHC Class I Chain-Related Gene A Alleles Distinguish Malnutrition-Modulated Diabetes, Insulin-Dependent Diabetes, and Non-Insulin- Dependent Diabetes Mellitus Patients from Eastern India. Annals of the New York Academy of Sciences 958:341-344
Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, Lathrop GM, Bell JI, Jewell DP (1996) Two stage genome-wide search inflammatory bowel disease provides evidence for susceptibility loci on chromosome 3, 7 and 12. Nature Genetics 14:199-202
Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, Clayton D, Goodfellow PN, Compston A (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nature Genetics 13:464-468
Saxena R, Voight BF, Lyssenko V, Burtt NP, De-Bakker PIW, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331-1336
Schadt E, Monks S, Drake T, Lusis A, Che N, Colinayo V, Ruff T, Milligan S, Lamb J, Cavet G, Linsley P, Mao M, Stoughton R, Friend S (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297-302
Schranz DB, Lernmark A (1998) Immunology in diabetes: an update. Diabetes/Metabolism Reviews 14:3-29
Scott HS, Heino M, Peterson P, Mittaz L, Lalioti MD, Betterle C, Cohen A, Seri M, Lerone M, Romeo G, Collin P, Salo M, Metcalfe R, Weetman A, Papasavvas MP, Rossier C, Nagamine K, Kudoh J, Shimizu N, Krohn KJE, Antonarakis SE (1998) Common mutations in autoimmune polyendocrinopathy-candidiasis ectodermal dystrophy patients of different origins. Molecular Endocrinology 12:1112-1119
Scott LJ, Bonnycastle LL, Willer CJ, Sprau AG, Jackson AU, Narisu N, Duren WL, Chines PS, Stringham HM, Erdos MR, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Collins FS, Boehnke M (2006) Association of Transcription Factor 7-Like 2 (TCF7L2) Variants With Type 2 Diabetes in a Finnish Sample. Diabetes 55:2649-2653
Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, et al. (2007) A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. Science 316:1341-1345
Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, M.ånér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528
Seegers D, Zwiers A, Strober W, Peña AS, Bouma G (2002) A TaqI polymorphism in the 3'UTR of the IL-12 p40 gene correlates with increased IL-12 secretion. Genes and Immunity 3:419-423
Senee V, Vattem KM, Delepine M, Rainbow LA, Haton C, Lecoq A, Shaw NJ, Robert J-J, Rooman R, Diatloff-Zito C, Michaud JL, Bin-Abbas B, Taha D, Zabel B, Franceschini P, Topaloglu AK, Lathrop GM, Barrett TG, Nicolino M, Wek RC, Julier C (2004) Wolcott-Rallison Syndrome: Clinical, Genetic, and Functional Study of EIF2AK3 Mutations and Suggestion of Genetic Heterogeneity. Diabetes 53:1876-1883
Serrano-Rios M, Mato J, Oya M, Larrodera L, Hawkins F, Escobar F (1973) Failure to find Ig-G insulin antibodies in Down's syndrome. Hormone and Metabolic Research 5:57-58
Sharfe N, Dadi HK, Shahar M, Roifman CM (1997) Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. Proceedings of the National Academy of Sciences of the United States of America 94:3168-3171
Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. American Journal of Human Genetics 77:78-88
She JX (1996) Susceptibility to type I diabetes: HLA-DQ and DR revisited. Immunology Today 17:323-329
Shield JPH, Wadsworth EJK, Hassold TJ, Judis LA, Jacobs PA (1999) Is disomic homozygosity at the APECED locus the cause of increased autoimmunity in Down's syndrome? Archives of Disease in Childhood 81:147-150
Siebold C, Hansen B, Wyer J, Harlos K, Esnouf R, Svejgaard A, Bell J, Strominger J, Jones E, Fugger L (2004) Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proceedings of the National Academy of Sciences of the United States of America 101:1999-2004
Skorka A, Bednarczukt T, Bar-Andziak E, Nauman J, Ploski R (2005) Lymphoid tyrosine phosphatase (PTPN22/LYP) variant and Graves' disease in a Polish population: Association and gene dose-dependent correlation with age of onset. Clinical Endocrinology 62:679-682
Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson T, Montpetit A, Pshezhetsky A, Prentki M, Posner B, Balding D, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881-885
Smith CJA, Crock PA, King BR, Meldrum CJ, Scott RJ (2004) Phenotype-genotype correlations in a series of Wolfram syndrome families. Diabetes Care 27:2003-2009
Smyth D, Cooper J, Collins J, Heward J, Franklyn J, Howson J, Vella A, Nutland S, Rance H, Maier L, Barratt B, Guja C, Ionescu-Tirgoviste C, Savage D, Dunger D, Widmer B, Strachan D, Ring S, Walker N, Clayton D, Twells R, Gough S, Todd J (2004) Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. Diabetes 53:3020-3023
Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C,

Ionescu-Tirgoviste C, Widmer B, Dunger DB, Savage DA, Walker NM, Clayton DG, Todd JA (2006a) A genome-wide association study of non-synonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nature Genetics 38:617-619

Smyth DJ, Howson JMM, Payne F, Maier LM, Bailey R, Holland K, Lowe CE, Cooper JD, Hulme JS, Vella A, Dahlman I, Lam AC, Nutland S, Walker NM, Twells RCJ, Todd JA (2006b) Analysis of polymorphisms in 16 genes in type 1 diabetes that have been associated with other immune-mediated diseases. BMC Medical Genetics 7:NIL_1-NIL_12

Sparre T, Christensen UB, Larsen PM, Fey SJ, Wrzesinski K, Roepstorff P, Mandrup-Poulsen T, Pociot F, Karlsen AE, Nerup J (2002) IL-1 beta induced protein changes in diabetes prone BB rat islets of Langerhans identified by proteome analysis. Diabetologia 45:1550-1561

Spector TD, Reneland RH, Mah S, Valdes AM, Hart DJ, Kammerer S, Langdown M, Hoyal CR, Atienza J, Doherty M, Rahman P, Nelson MR, Braun A (2006) Association between a variation in LRCH1 and knee osteoarthritis – A genome-wide single-nucleotide polymorphism association study using DNA pooling. Arthritis and Rheumatism 54:524-532

Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. American Journal of Human Genetics 59:983-989

Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. American journal of human genetics 62:450-458

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium – the insulin gene region and insulin-dependent diabetes-mellitus (IDDM). American Journal of Human Genetics 52:506-516

Spranger J, Kroke A, Mohlig M, Hoffmann K, Bergmann M, Ristow M, Boeing H, Pfeiffer A (2003) Inflammatory cytokines and the risk to develop type 2 diabetes: results of the prospective population-based European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Diabetes 52:812-817

Stanilova S, Miteva L (2005) Taq-I polymorphism in 3'UTR of the IL-12B and association with IL-12p40 production from human PBMC. Genes and Immunity 6:364-366

Starr R, Willson T, Viney E, Murray L, Rayner J, Jenkins B, Gonda T, Alexander W, Metcalf D, Nicola N, Hilton D (1997) A family of cytokine-inducible inhibitors of signalling. Nature 387:917-921

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, et al. (2002) Neuregulin 1 and susceptibility to schizophrenia. American Journal of Human Genetics 71:877-892

Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, et al. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nature Genetics 39:770-775

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489-493

Stranger B, Forrest M, Clark A, Minichiello M, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis S, Tavaré S, Deloukas P, Dermitzakis E (2005) Genome-wide associations of gene expression variations in humans. PLOS Genetics 1:e78

Stratmann T, Apostolopoulos V, Mallet-Designe V, Corper A, Scott C, Wilson I, Kang A, Teyton L (2000) The I-Ag7 MHC class II molecule linked to murine diabetes is a promiscuous peptide binder. TheJournal of Immunology 165:3214-3225

Strom TM, H.örtnagel K, Hofmann S, Gekeler F, Scharfe C, Rabl W, Gerbitz KD, Meitinger T (1998) Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. Human Molecular Genetics 7:2021-2028

Svensson J, Carstensen B, Molbak A, Christau B, Mortensen H, Nerup J, Borch-Johnsen K (2002) Increased risk of childhood type 1 diabetes in children born after 1985. Diabetes Care 25:2197-2201

Tabor HK, Risch NJ, Myers RM (2002) Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics 3:391-397

Tait KF, Marshall T, Berman J, Carr-Smith J, Rowe B, Todd JA, Bain SC, Barnett AH, Gough SCL (2004) Clustering of autoimmune disease in parents of siblings from the Type 1 diabetes Warren repository. Diabetic Medicine 21:358-362

Temelkova-Kurktschiev T, Henkel E, Koehler C, Karrei K, Hanefeld M (2002a) Subclinical inflammation in newly detected Type II diabetes and impaired glucose tolerance. Diabetologia 45:151

Temelkova-Kurktschiev T, Siegert G, S B, Henkel E, Koehler C, Jaross W, Hanefeld M (2002b) Subclinical inflammation is strongly related to insulin resistance but not to impaired insulin secretion in a high risk population for diabetes. Metabolism 51:743-749

Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788-793

Thomsen M, Molvig J, Zerbib A, de-Preval C, Abbal M, Dugoujon J, Ohayon E, Svejgaard A, Cambon-Thomsen A, Nerup J (1988) The susceptibility to insulin-dependent diabetes mellitus is associated with C4 allotypes independently of the association with HLA-DQ alleles in HLA-DR3,4 heterozygotes. Immunogenetics 28:320-327

Thomsen M, Platz P, Andersen OO, Christy M, Lyngsøoe, Nerup J, Rasmussen K, Ryder LP, Nielsen LS, Svejgaard A (1975) MLC typing in juvenile diabetes mellitus and idiopathic Addison's disease. Transplantation Reviews 22:125-147

Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, Barbosa J, Rich SS, Bertrams J, Baur MP, etal. (1988) Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. American Journal of Human Genetics 43:799-816

Thorsby E (1997) Invited anniversary review: HLA associated diseases. Human Immunology 53:1-11

Tian J, Olcott AP, Hanssen LR, Zekzer D, Middleton B, Kaufman DL (1998) Infectious Th1 and Th2 autoimmunity in diabetes-prone mice. Immunological Reviews 164:119-127

Tica V, Nikitina-Zake L, Donadi E, Sanjeevi CB (2003) MIC-A Genotypes 4/5.1 and 9/9 Are Positively Associated with Type 1 Diabetes Mellitus in Brazilian Population. Annals of the New York Academy of Sciences 1005:310-313

Todd J, Walker N, Cooper J, Smyth D, Downes K, Plagnol V, Bailey R, et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nature Genetics 39:857-864

Todd JA, Wicker LS (2001) Genetic protection from the inflammatory disease type 1 diabetes in humans and animal models. Immunity 15:387-395

Tokuhiro S, Yamada R, Chang XT, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K (2003) An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nature Genetics 35:341-348

Tomfohrde J, Silverman A, Barnes R, Fernandez-Vina MA, Young M, Lory D, Morris L, Wuepper KD, Stastny P, Menter A, etal. (1994) Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. Science 264:1141-1145

Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, Stewart CA, Ashurst JL, Atrazhev AM, Coggill P, Palmer S, Almeida J, Sims S, Wilming LG, Rogers J, De-Jong PJ, Carrington M, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. PLoS Genetics 2:81-92

Trembleau S, Penna G, Bosi E, Mortara A, Gately MK, Adorini L (1995) Interleukin-12 administration induces T-helper type-1 cells and accelerates autoimmune diabetes in NOD mice. Journal of Experimental Medicine 181:817-821

Tsukahara F, Hattori M, Muraki T, Sakaki Y (1996) Identification and cloning of a novel cDNA belonging to tetratricopeptide repeat gene family from Down syndrome-critical region 21q22.2. Journal of biochemistry 120:820-827

Tsunoda T, Lathrop GM, Sekine A, Yamada R, Takahashi A, Ohnishi Y, Tanaka T, Nakamura Y (2004) Variation of gene-based SNPs and linkage disequilibrium patterns in the human genome. Human Molecular Genetics 13:1623-1632

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE (2005) Fine-scale structural variation of the human genome. Nature Genetics 37:727-732

Ueda H, Howson JMM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, et al. (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423:506-511

Undlien D, Akselsen H, Joner G, Dahl-Jorgensen K, Sovik O, Ronningen K, Thorsby E (1997) No independent associations of LMP2 and LMP7 polymorphisms with susceptibility to develop IDDM. Diabetes 46:307-312

Undlien DE, Lie BA, Thorsby E (2001) HLA complex genes in type 1 diabetes and other autoimmune diseases. Which genes are involved? Trends in Genetics 17:93-100

Valdes A, Noble J, Clerget-Darpoux F, Erlich H, Thomson G (2001) Modeling of HLA class II susceptibility to Type I diabetes reveals an effect associated with DPB1. Genetic Epidemiology 21:212-223

van Endert P, Liblau R, Patel S, Fugger L, Lopez T, Pociot F, Nerup J, McDevitt H (1994) Major histocompatibility complex-encoded antigen processing gene polymorphism in IDDM. Diabetes 43:110-117

van Heel DA, Dechairo BM, Dawson G, McGovern DPB, Negoro K, Carey AH, Cardon LR, Mackay I, Jewell DP, Lench NJ (2003) The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. Human Molecular Genetics 12:2569-2575

Van-Eerdewegh P, Little RD, Dupuis J, Del-Mastro RG, Falls K, Simon J, Torrey D, et al. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426-430

Van-Goor JC, Massa GG, Hirasing R (1997) Increased incidence and prevalence of diabetes mellitus in Down's syndrome. Archives of disease in childhood 77:186-PY
Vaxillaire M, Benmezroua Y, Durand E, Vasseur F, Froguel P (2001) No evidence for diabetes-associated mutations of PEK/EIF2AK3 gene in French patients with early-onset type II diabetes. Diabetologia 44:786
Velaga M, Wilson V, Jennings C, Owen C, Herington S, Donaldson P, Ball S, James R, Quinton R, Perros P, Pearce S (2004) The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. Journal of Clinical Endocrinology and Metabolism 89:5862-5865
Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, Widmer B, Jones R, Ring SM, McArdle W, Pembrey ME, Strachan DP, Dunger DB, Twells RCJ, Clayton DG, Todd JA (2005) Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. American Journal of Human Genetics 76:773-779
Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA (2004) Loss of Functional Suppression by CD4+CD25+ Regulatory T Cells in Patients with Multiple Sclerosis. Journal of Experimental Medicine 199:971-979
Waldhör T, Schober E, Rami B, Tuomilehto J (1999) The prevalence of IDDM in the first degree relatives of children newly diagnosed with IDDM in Austria--a population-based study. Austrian Diabetes Incidence Study Group. Experimental and clinical endocrinology & diabetes 107:323-327
Wang W, Barratt B, Clayton D, Todd J (2005) Genome-wide association studies: Theoretical and practical concerns. Nature Reviews Genetics 6:109-118
Warrington JA, Bengtsson U (1994) High-resolution physical mapping of human 5q31-q33 using 3 methods – radiation hybrid mapping, interphase fluorescence in-situ hybridization, and pulsed-field gel-electrophoresis. Genomics 24:395-398
Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678
Wicker LS, Clark J, Fraser HI, Garner VES, Gonzalez-Munoz A, Healy B, Howlett S, Hunter K, Rainbow D, Rosa RL, Smink LJ, Todd JA, Peterson LB (2005) Type 1 diabetes genes and pathways shared by humans and NOD mice. Journal of Autoimmunity 25:29-33
Wilkin T (2001) The accelerator hypothesis: weight gain as the missing link between Type I and Type II diabetes. Diabetologia 44:914-922
Windsor L, Morahan G, Huang D, McCann V, Jones T, James I, Christiansen FT, Price P (2004) Alleles of the IL12B 3′UTR associate with late onset of type 1 diabetes. Human Immunology 65:1432-1436
Wolcott CD, Rallison ML (1972) Infancy-onset diabetes mellitus and multiple epiphyseal dysplasia. Journal of Pediatrics 80:292-297
Wolfram D, Wagener H (1938) Diabetes mellitus and simple optic atrophy among siblings: report of four cases. Mayo Clinical Proceedings 9:715–718
Wu RL, Lin M (2006) Opinion – Functional mapping – how to map and study the genetic architecture of dynamic complex traits. Nature Reviews Genetics 7:229-237
Xiong MM, Zhao JY, Boerwinkle E (2003) Haplotype bolck linkage disequilibrium mapping. Frontiers in Bioscience 8:A85-A93
Yamanouchi J, Rainbow D, Serra P, Howlett S, Hunter K, Garner VES, Gonzalez-Munoz A, Clark J, Veijola R, Cubbon R, Chen SL, Rosa R, Cumiskey AM, Serreze DV, Gregory S, Rogers J, Lyons PA, Healy B, Smink LJ, Todd JA, Peterson LB, Wicker LS, Santamaria P (2007) Interleukin-2 gene variation impairs regulatory T cell function and causes autoimmunity. Nature Genetics 39:329-337
Yamato E, Ikegami H, Kawaguchi Y, Fujisawa T, Hamada Y, Ueda H, Shintani M, Ogihara T (1997) Insulin-dependent diabetes mellitus associated with autoimmune thyroiditis and rheumatoid arthritis. American Journal of the Medical Sciences 313:64-66
Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D, Cardon L, Takazoe M, Tanaka T, Ichimori T, Saito S, Sekine A, Iida A, Takahashi A, Tsunoda T, Lathrop M, Nakamura Y (2005) Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. Human Molecular Genetics 14:3499-3506
Yang JM, Nagasaka S, Yatagai T, Nakamura T, Kusaka I, Ishikawa SE, Saito T, Ishibashi S (2006) Interleukin-12p40 gene (IL-12B) polymorphism and Type 1 diabetes mellitus in Japanese: possible role in subjects without having high-risk HLA haplotypes. Diabetes Research and Clinical Practice 71:164-169
Yang Y, Santamaria P (2006) Lessons on autoimmune diabetes from animal models. Clinical Science 110:627-639
Ying SY, Chang DC, Miller JD, Lin SL (2006) MicroRNA protocols. Perspectives. Methods in Molecular Biology 342:351-358
Yokoi N, Kanazawa M, Kitada K, Tanaka A, Kanazawa Y, Suda S, Ito H, Serikawa T, Komeda K (1997) A non-MHC locus essential for autoimmune type I diabetes in the Komeda Diabetes-Prone rat. Journal of clinical investigation 100:2015-2021
Yokoi N, Komeda K, Wang H, Yano H, Kitada K, Saitoh Y, Seino Y, Yasuda K, Serikawa T, Seino S (2002) Cblb is a major susceptibility gene for rat type 1 diabetes mellitus. Nature Genetics 31:391-394
Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336-1341
Zhang C, Qi L, Hunter DJ, Meigs JB, Manson JE, van-Dam RM, Hu FB (2006) Variant of Transcription Factor 7-Like 2 (TCF7L2) Gene and the Risk of Type 2 Diabetes in Large Cohorts of U.S. Women and Men. Diabetes 55:2645-2648
Zhang K, Calabrese P, Nordborg M, Sun F (2002) Haplotype block structure and its applications to association studies: power and study designs. American Journal of Human Genetics 71:1386-1394