Published in final edited form as:

Diabetes Obes Metab. 2009 February ; 11(Suppl 1): 60–66. doi:10.1111/j.1463-1326.2008.01004.x.

Identification of T1D susceptibility genes within the MHC region by combining protein interaction networks and SNP genotyping data

C. Brorsson¹, N. T. Hansen², K. Lage², R. Bergholdt¹, S. Brunak², F. Pociot^{1,3}, and Type 1 Diabetes Genetics Consortium

¹Steno Diabetes Center, Gentofte, Denmark

²Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark

³Department of Clinical Sciences, University Hospital Malmö, Lund University, Malmö, Sweden

Abstract

Aim—To develop novel methods for identifying new genes that contribute to the risk of developing type 1 diabetes within the Major Histocompatibility Complex (MHC) region on chromosome 6, independently of the known linkage disequilibrium (LD) between human leucocyte antigen (*HLA*)-*DRB1*, *-DQA1*, *-DQB1* genes.

Methods—We have developed a novel method that combines single nucleotide polymorphism (SNP) genotyping data with protein–protein interaction (ppi) networks to identify disease-associated network modules enriched for proteins encoded from the MHC region. Approximately 2500 SNPs located in the 4 Mb MHC region were analysed in 1000 affected offspring trios generated by the Type 1 Diabetes Genetics Consortium (T1DGC). The most associated SNP in each gene was chosen and genes were mapped to ppi networks for identification of interaction partners. The association testing and resulting interacting protein modules were statistically evaluated using permutation.

Results—A total of 151 genes could be mapped to nodes within the protein interaction network and their interaction partners were identified. Five protein interaction modules reached statistical significance using this approach. The identified proteins are well known in the pathogenesis of T1D, but the modules also contain additional candidates that have been implicated in β -cell development and diabetic complications.

Conclusions—The extensive LD within the MHC region makes it important to develop new methods for analysing genotyping data for identification of additional risk genes for T1D. Combining genetic data with knowledge about functional pathways provides new insight into mechanisms underlying T1D.

Keywords

genetic association; integrative genomics; major histocompatibility complex; protein interaction networks; type 1 diabetes

Correspondence: Flemming Pociot, Steno Diabetes Center, Niels Steensensvej 2, DK-2820 Gentofte, Denmark. fpoc@steno.dk.

Conflict of Interest: The authors declare that they have no conflicts of interest in publishing this article.

Introduction

Type 1 diabetes (T1D) is a complex disease believed to be influenced by many interacting genetic and environmental factors. The Major Histocompatibility Complex (MHC) region on chromosome 6p21 has been robustly replicated as a major risk locus in many immune-mediated diseases, including T1D [1]. Within the MHC region, certain haplotypes of human leucocyte antigen (*HLA*) genes are known to confer the main genetic risk or protection. For T1D, these are defined by a complex epistasis between *HLA-DRB1*, *-DQA1*, *-DQB1* alleles within the MHC class II region [2]. Some evidence has emerged that not all genetic risk carried by the MHC region can be explained by these genes and that there could be additional genes secondary to *HLA DR-DQ* within the MHC region that provides risk for T1D [3,4]. However, because of the extensive linkage disequilibrium (LD) between genes in this region, the identification of possible additional genes is not straightforward. In an integrative genomics approach, we have developed a novel method for identifying risk genes for T1D within the MHC region that is independent of LD by applying protein–protein interaction (ppi) networks in the analysis of single nucleotide polymorphism (SNP) association data. We used a ppi platform based on high-confidence human protein interactions generated by extensive data integration across several model organisms [5]. The ppi network method has recently been successfully used on T1D data [6]. Genetic epistasis analyses were performed on genome-wide linkage data, suggesting a set of genetic interactions that also interact physically in protein complexes. The resulting interaction networks were statistically evaluated and functional support for the genetic interactions were thereby demonstrated [6].

In the present study, we analysed SNP association data for the proteins encoded by the MHC region using ppi networks. By analyzing interaction partners of candidate genes in these modules additional candidate genes might be identified based on their functional role in biologically relevant pathways. The combination of ppi networks and SNP association signals is a new approach for identifying additional risk genes within the MHC region, which to our knowledge has never been performed before. In this study, we present preliminary results demonstrating the feasibility of using this approach on data from a large family-based association study for SNPs in the MHC region.

Materials and Methods

Samples and Markers

The T1DGC MHC data set (February 2007 release) containing families from five different Caucasoid cohorts (Human Biological Data Interchange (HBDI), British Diabetes Association (BDA), Joslin (JOS), UK and Danish (DAN)) were typed for 2957 SNPs using Illumina panels. Only SNPs with a call rate above 95% and individuals that passed a genotyping success rate of 90% were considered for further analysis. SNPs that did not pass Hardy–Weinberg equilibrium in founders ($p > 0.001$) had a minor allele frequency of less than 1% or that showed Mendelian errors were removed. This resulted in 974 affected offspring trios that were analysed for 2441 SNPs.

Association Study

Association with T1D was analysed using the transmission disequilibrium test (TDT). A total of 1526 SNPs that were analyzed for association could be mapped to gene products (5 kb upstream from transcription start site to stop codon sequence) within the 4 Mb MHC region present in the ppi network. The SNP with the highest $-\log(10)$ TDT value was assigned to each respective gene/protein.

Protein–protein Interaction Networks

A second order ppi network was generated from the genes located in the 4 Mb MHC region. A total of 151 genes in the genotyped region could be mapped to nodes in the Inweb protein interaction network developed by Lage *et al.* [5]. The $-\log(10)$ TDT value for the highest associated SNP within each gene or its promoter region was assigned to each node in the network. All genes in the MHC region and their interaction partners were used as bait proteins in virtual pull-downs from the second order network, generating an inventory of modules that potentially could be enriched for genes associated with T1D [5,6] (figure 1).

As the association data only covered ~6% of the proteins in the ppi network, each pull-down was transformed into a pseudo-complex only including genes with association data. The average association of the proteins in each reduced virtual pull-down was compared with a distribution generated from 10 000 randomly sampled pull-downs of the same pseudo-size. This resulted in a p value for each network that reflected the likelihood of selecting a subset of the genes in the extended MHC region with the average association found in the network. From the permutation-based p values, the most significantly enriched modules were selected for a thorough study.

Results

The quality thresholds chosen for SNPs and individuals resulted in 2443 SNPs that were analysed in 974 affected offspring trios. The associated interaction modules were statistically validated using permutation. Five interaction modules had a permuted p value below 0.05, which was chosen as the cut-off for calling significance (table 1). Each of these is briefly described.

Associated Modules

HLA-DQA2—The network module with the lowest p value contained HLA-DQA2 interacting with two preys outside of the MHC region (figure 1A). The preys were identified as T-cell antigen *CD4* (cluster of differentiation 4) and insulin (*INS*). The importance of the T-cell receptor (TCR) and its co-receptor (CD4) interacting with MHC molecules bound to antigenic peptides is well known for an adaptive immune response [7]. Wang *et al.* [8] excluded a direct TCR–CD4 interaction but found evidence for the TCR and CD4 signalling being co-ordinated around the antigenic MHC class II complex. Insulin has been suggested as an important autoantigen in T1D and its interaction with MHC class II molecules has been studied using X-ray crystallography [9]. The strongest genetic association to T1D is conferred by HLA class II genes and particularly there is strong evidence for the involvement of *HLA-DQ*, and *-DR* genes in the risk of developing the disease [2].

Advanced Glycosylation End Product–Specific Receptor—The gene for advanced glycosylation end product–specific receptor (*AGER* or *RAGE*) is located in the MHC class III region. This bait protein was identified in a network module with six interaction partners, all encoded outside of the MHC region (figure 1B). Four of these belong to the S100 family of calcium-binding proteins. The other two preys were high mobility group protein B1 (*HMGB1*) and transthyretin precursor (*TTR*). *AGER* is a member of the immunoglobulin superfamily of receptors and is known to interact with multiple ligands, such as advanced glycosylation end products, S100/calgranulins and *HMGB1* [10–13].

Advanced glycosylation end products have been implicated in various diseases such as diabetes [14] and Alzheimer disease [15]. In diabetes, activation of *AGER* signalling has been primarily implicated in diabetic complications such as neuropathy through activation of nuclear factor kappa B and interleukin (IL)-6 [16]. A polymorphism in *AGER* has been associated with T1D;

interestingly, the association was HLA-DQB1 dependent [17]. The same study also showed that the same polymorphism was associated with diabetic nephropathy and retinopathy in T1D individuals. Members of the S100 calcium-binding protein superfamily are suggested to function as cytokines and mediators of inflammation. Their activation of *AGER* has also been implicated in T1D and diabetic complications [18–20].

HMGB1 is a chromatin-binding protein that is released from necrotic cells. Steer *et al.* [21] found that IL-1 β stimulated the release of *HMGB1* from rat β cells and rat islets, indicating a necrotic response to IL-1. In rat RINm5F cells, *HMGB1* has been found to bind to the promoter of the inducible nitric oxide synthase, a potential mediator of cytokine-induced β -cell dysfunction in T1D [22]. Transthyretin functions as a transport protein for thyroxine and retinol (vitamin A), together with retinol-binding protein. Plasma levels of TTR together with retinol and retinol-binding protein have been shown to be reduced in individuals with T1D in several studies [23,24]. TTR has been shown to interact with *AGER* in a study of familial amyloidotic polyneuropathy, a neurodegenerative disorder characterized by TTR amyloid fibrils [25].

Cluster of Differentiation 4 (CD4)—The network module with CD4 as bait had seven interaction partners, which were located in the MHC region (figure 1C). The preys were identified as six different HLA-DR, -DP, -DQ molecules and β -tubulin (*TUBB*). The role of MHC class II molecules in antigen presentation and their combination with CD4 for a T-cell response is again accentuated. An association of CD4 and components of the cytoskeleton has been reported [26,27]. CD4 and *TUBB* were also found co-purified in a study of CD4/lck interaction partners using sodium dodecyl sulphate–polyacrylamide gel electrophoresis and mass spectrometry [28]. Polymorphisms in the promoter region of the human *CD4* gene and their association to T1D have been analysed in a Danish family material [29,30]. Evidence for association to T1D was initially found for a variable number of tandem repeat (VNTR) polymorphism and later linkage and association was demonstrated for a haplotype including the VNTR and three SNPs in the *CD4* promoter. This haplotype was also associated with constitutively high activity of the *CD4* promoter [30].

Apolipoprotein M—The gene for apolipoprotein M (*APOM*) is located between *BAT3* and *BAT4* within the MHC region (figure 1D). ApoM is a component of HDL particles and has been shown to have an important function in the formation of pre- β -HDL and cholesterol efflux in mice [31,32]. Wolfrum *et al.* [32] also showed that apoM expression can inhibit atherosclerotic lesions in LDL-receptor knockout mice fed a cholesterol-rich diet.

Individuals with maturity-onset diabetes in the young type 3 (MODY3) have reduced plasma levels of apoM compared with normal and control diabetics [33]. Coronary heart disease has been shown to be more common in MODY3 subjects compared with T1D individuals, thereby indicating that low levels of apoM might increase susceptibility for atherosclerosis in humans [32]. As apoM is associated with HDL, it would be possible that this protein is only important in metabolic disorders such as obesity and type 2 diabetes. However, Xu *et al.* [34] investigated the expression and secretion of apoM in alloxan-treated mice and demonstrated that apoM was significantly reduced in alloxan-induced diabetic mice compared with controls and that administration of insulin could partially normalize apoM plasma levels and mRNA levels. They concluded that apoM was related to diabetes in this mouse model but the mechanism of the dysfunction is unclear. In the associated network module, apoM was found to interact with Kruppel-like factor 6 (*KLF6*) and C8orf30A. None of these proteins has been described in T1D. The gene encoding *KLF6* is located on chromosome 10p15 and the protein has mainly been described as a tumour suppressor for prostate cancers [35]. C8orf30A is a protein of unknown function also known as Brain protein 16, which is encoded from chromosome 8q24.

Allograft Inflammatory Factor 1—Allograft inflammatory factor 1 (AIF1) is also known as ionized calcium-binding adaptor molecule 1 (figure 1E). The gene is located in the MHC class III region adjacent to *BAT2*. It was originally cloned from a rat model of cardiac allograft rejection and high levels were demonstrated in infiltrating macrophages [36]. AIF1 has been demonstrated to affect glucose-induced insulin release in T1D rat models *in vivo* and AIF1-positive cells were highly present in the prediabetic pancreas, which indicates a function in cell-mediated autoimmune responses [37]. Polymorphisms in introns of *AIF1* were tested for association with T1D in Japanese individuals but association could not be demonstrated [38]. In the identified network module, AIF1 interacted with lymphocyte cytosolic protein 1 (LCP1 or L-plastin). The LCP1 protein was found to be upregulated in a proteomic analysis in a model of pancreatic regeneration in rats, indicating a role in β -cell differentiation [39].

Discussion

To our knowledge this is the first study where SNP association data have been combined with ppi networks for gene products from the MHC region to assess additional risk genes for T1D. Despite the well-known importance of HLA for the risk of developing T1D, the exact combination of causal genes has not yet been identified. This can be explained, in part, by the extensive LD between *HLA* and other genes in this region. Analyses stratified for *HLA* haplotypes have had limited success in identifying additional independent genes, highlighting the importance for completely new approaches for analysing association data. Identifying protein interaction partners to known proteins or candidate gene products could aid in understanding the biological pathways and functions that might be involved in disease pathogenesis.

We believe that this new integrative genomics approach gives the opportunity to look for new candidate genes without having to account for LD between HLA genes. We identified several interaction modules enriched for proteins with SNP association signals that were also statistically validated using permutation. As a proof of concept two of the modules, CD4 and HLA-DQA2, included additional proteins that are already well known in the pathogenesis of T1D. However, the other associated modules also included proteins that were found to be previously associated with T1D in various ways.

The network module containing *AGER* identified several proteins involved in inflammation and that also have been linked to the development of diabetic complications. The inflammatory factor AIF1 was found to interact with a protein that has been implicated in β -cell differentiation and AIF1 itself, apart from having a role in inflammation, has been shown to affect insulin release in rat models of T1D. One network module, containing ApoM, identified additional interacting proteins that have not previously been linked to T1D. However, ApoM, as a component of HDL particles, has a likely role in metabolic disorders but has also shown reduced plasma levels in a mouse model of T1D. Additional work is needed to characterize the interacting proteins KLF6 and c8orf30 and to elucidate their possible role in T1D. However, this highlights the strength of the associated modules approach for detecting additional candidate genes that could be involved in the pathogenesis of the disease.

The developed approach provides a promising base from which further studies can be performed. The family collection provided by the T1DGC is genetically well characterized which gives the opportunity to perform analyses on subsets of the data, for example, stratified by *HLA* genotypes. This would have the power to highlight differences between these groups of patients and the significant network modules identified for each group. The current study proved the feasibility of this novel approach in predicting candidate genes for complex diseases.

Acknowledgments

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. For more information on the Type 1 Diabetes Genetics Consortium please see <http://www.t1dgc.org>. Caroline Brorsson was funded by a research grant from EFSD/JDRF/NovoNordisk and Regine Bergholdt was funded by the Danish Medical Research Council (271-05-0672).

References

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678. [PubMed: 17554300]
2. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002;3:235–249. [PubMed: 12140742]
3. Lie BA, Todd JA, Pociot F, et al. The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet* 1999;64:793–800. [PubMed: 10053014]
4. Johansson S, Lie BA, Todd JA, et al. Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1. *Genes Immun* 2003;4:46–53. [PubMed: 12595901]
5. Lage K, Karlberg EO, Stirling ZM, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 2007;25:309–316. [PubMed: 17344885]
6. Bergholdt R, Stirling ZM, Lage K, et al. Novel human protein networks of genetic interactions in type 1 diabetes. *Genome Biol* 2007;8:R253. [PubMed: 18045462]
7. Sette A, Buus S, Colon S, Smith JA, Miles C, Grey HM. Structural characteristics of an antigen required for its interaction with Ia and recognition by T cells. *Nature* 1987;328:395–399. [PubMed: 3497349]
8. Wang JH, Meijers R, Xiong Y, et al. Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc Natl Acad Sci U S A* 2001;98:10799–10804. [PubMed: 11535811]
9. Lee KH, Wucherpfennig KW, Wiley DC. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat Immunol* 2001;2:501–507. [PubMed: 11376336]
10. Schmidt AM, Vianna M, Gerlach M, et al. Isolation and characterization of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface. *J Biol Chem* 1992;267:14987–14997. [PubMed: 1321822]
11. Nepper M, Schmidt AM, Brett J, et al. Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J Biol Chem* 1992;267:14998–15004. [PubMed: 1378843]
12. Hofmann MA, Drury S, Fu C, et al. RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* 1999;97:889–901. [PubMed: 10399917]
13. Hori O, Brett J, Slaterry T, et al. The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphotericin B. Mediation of neurite outgrowth and co-expression of RAGE and amphotericin B in the developing nervous system. *J Biol Chem* 1995;270:25752–25761. [PubMed: 7592757]
14. Brownlee M. Advanced protein glycosylation in diabetes and aging. *Annu Rev Med* 1995;46:223–234. [PubMed: 7598459]
15. Smith MA, Taneda S, Richey PL, et al. Advanced Maillard reaction end products are associated with Alzheimer disease pathology. *Proc Natl Acad Sci U S A* 1994;91:5710–5714. [PubMed: 8202552]
16. Haslbeck KM, Bierhaus A, Erwin S, et al. Receptor for advanced glycation endproduct (RAGE)-mediated nuclear factor-kappaB activation in vasculitic neuropathy. *Muscle Nerve* 2004;29:853–860. [PubMed: 15170618]
17. Lindholm E, Bakhtadze E, Sjogren M, et al. The -374 T/A polymorphism in the gene encoding RAGE is associated with diabetic nephropathy and retinopathy in type 1 diabetic patients. *Diabetologia* 2006;49:2745–2755. [PubMed: 16969646]

18. Zimmer DB, Chessher J, Wilson GL, Zimmer WE. S100A1 and S100B expression and target proteins in type I diabetes. *Endocrinology* 1997;138:5176–5183. [PubMed: 9389498]
19. Strachan MW, Abraha HD, Sherwood RA, et al. Evaluation of serum markers of neuronal damage following severe hypoglycaemia in adults with insulin-treated diabetes mellitus. *Diabetes Metab Res Rev* 1999;15:5–12. [PubMed: 10398541]
20. McIntyre EA, Abraha HD, Perros P, Sherwood RA. Serum S-100beta protein is a potential biochemical marker for cerebral oedema complicating severe diabetic ketoacidosis. *Diabet Med* 2000;17:807–809. [PubMed: 11131106]
21. Steer SA, Scarim AL, Chambers KT, Corbett JA. Interleukin-1 stimulates beta-cell necrosis and release of the immunological adjuvant HMGB1. *PLoS Med* 2006;3:e17. [PubMed: 16354107]
22. Darville MI, Terryn S, Eizirik DL. An octamer motif is required for activation of the inducible nitric oxide synthase promoter in pancreatic beta-cells. *Endocrinology* 2004;145:1130–1136. [PubMed: 14630716]
23. Itoh N, Hanafusa T, Miyagawa J, et al. Transthyretin (prealbumin) in the pancreas and sera of newly diagnosed type I (insulin-dependent) diabetic patients. *J Clin Endocrinol Metab* 1992;74:1372–1377. [PubMed: 1592883]
24. Basu TK, Basualdo C. Vitamin A homeostasis and diabetes mellitus. *Nutrition* 1997;13:804–806. [PubMed: 9290094]
25. Monteiro FA, Cardoso I, Sousa MM, Saraiva MJ. In vitro inhibition of transthyretin aggregate-induced cytotoxicity by full and peptide derived forms of the soluble receptor for advanced glycation end products (RAGE). *FEBS Lett* 2006;580:3451–3456. [PubMed: 16716307]
26. Thuillier L, Hivroz C, Fagard R, Andreoli C, Mangeat P. Ligation of CD4 surface antigen induces rapid tyrosine phosphorylation of the cytoskeletal protein ezrin. *Cell Immunol* 1994;156:322–331. [PubMed: 8025951]
27. Kinch MS, Sanfridson A, Doyle C. The protein tyrosine kinase p56lck regulates cell adhesion mediated by CD4 and major histocompatibility complex class II proteins. *J Exp Med* 1994;180:1729–1739. [PubMed: 7964457]
28. Bernhard OK, Cunningham AL, Sheil MM. Analysis of proteins copurifying with the CD4/lck complex using one-dimensional polyacrylamide gel electrophoresis and mass spectrometry: comparison with affinity-tag based protein detection and evaluation of different solubilization methods. *J Am Soc Mass Spectrom* 2004;15:558–567. [PubMed: 15047060]
29. Kristiansen OP, Zamani M, Johannesen J, et al. Linkage and association between a CD4 gene polymorphism and IDDM in Danish IDDM patients The Danish IDDM Epidemiology and Genetics Group, and The Danish Study Group of Diabetes in Childhood. *Diabetes* 1998;47:281–283. [PubMed: 9519727]
30. Kristiansen OP, Karlsen AE, Larsen ZM, Johannesen J, Pociot F, Mandrup-Poulsen T. Identification of a type 1 diabetes-associated CD4 promoter haplotype with high constitutive activity. *Scand J Immunol* 2004;59:582–591. [PubMed: 15182254]
31. Xu N, Dahlback B. A novel human apolipoprotein (apoM). *J Biol Chem* 1999;274:31286–31290. [PubMed: 10531326]
32. Wolfrum C, Poy MN, Stoffel M. Apolipoprotein M is required for prebeta-HDL formation and cholesterol efflux to HDL and protects against atherosclerosis. *Nat Med* 2005;11:418–422. [PubMed: 15793583]
33. Richter S, Shih DQ, Pearson ER, et al. Regulation of apolipoprotein M gene expression by MODY3 gene hepatocyte nuclear factor-1alpha: haploinsufficiency is associated with reduced serum apolipoprotein M levels. *Diabetes* 2003;52:2989–2995. [PubMed: 14633861]
34. Xu N, Nilsson-Ehle P, Ahren B. Suppression of apolipoprotein M expression and secretion in alloxan-diabetic mouse: partial reversal by insulin. *Biochem Biophys Res Commun* 2006;342:1174–1177. [PubMed: 16516154]
35. Narla G, Friedman SL, Martignetti JA. Kruppel cripples prostate cancer: KLF6 progress and prospects. *Am J Pathol* 2003;162:1047–1052. [PubMed: 12651597]
36. Utans U, Arceci RJ, Yamashita Y, Russell ME. Cloning and characterization of allograft inflammatory factor-1: a novel macrophage factor identified in rat cardiac allografts with chronic rejection. *J Clin Invest* 1995;95:2954–2962. [PubMed: 7769138]

37. Chen ZW, Ahren B, Ostenson CG, et al. Identification, isolation, and characterization of daintain (allograft inflammatory factor 1), a macrophage polypeptide with effects on insulin secretion and abundantly present in the pancreas of prediabetic BB rats. *Proc Natl Acad Sci U S A* 1997;94:13879–13884. [PubMed: 9391121]
38. Nishimura M, Obayashi H, Mizuta I, et al. TNF, TNF receptor type 1, and allograft inflammatory factor-1 gene polymorphisms in Japanese patients with type 1 diabetes. *Hum Immunol* 2003;64:302–309. [PubMed: 12559634]
39. Yang M, Liu W, Wang CY, et al. Proteomic analysis of differential protein expression in early process of pancreatic regeneration in pancreatectomized rats. *Acta Pharmacol Sin* 2006;27:568–578. [PubMed: 16626512]

Fig. 1.

The significantly associated protein modules are shown and proteins encoded from the major histocompatibility complex regions are highlighted in red. First, a virtual pull-down of each candidate (bait) was performed querying the interaction network for all interactions (preys) of the protein (and subsequently all interactions between the interacting proteins) and only retaining the interactions over a given threshold as defined by the Bayesian predictor developed [5,6]. Only modules remaining following statistical assessment (10 000 random queries) are shown in figs 1A-1E.

Table 1
The five interaction modules identified as significant after permutation are listed

Bait	Size	Avg(−log(tdt))	Pseudo-size	Avg(−log(tdt))	p value
HLA-DQA2	3	52.83	1	79.25	0.0176
AGER	7	22.18	1	77.65	0.0213
CD4	97	2.96	7	41.04	0.0317
APOM	4	34.60	1	69.19	0.0410
AIF1	2	68.84	1	68.84	0.0426

AIF1, allograft inflammatory factor 1; AGER, advanced glycosylation end product–specific receptor; APOM, apolipoprotein M; CD4, cluster of differentiation 4; HDL, human leucocyte antigen; TDT, transmission disequilibrium test.

Size reflects the total number of interacting proteins in the virtual pull-down, whereas pseudo-size is the number of interacting proteins in the pseudo-complex encoded from the major histocompatibility complex region and thus having association data from the transmission disequilibrium test (TDT) analysis. The p values were generated by comparing the average −log(10) TDT value for each pseudo-complex to a random distribution after 10 000 permutations for pull-downs of the same pseudo-size.