Int J Clin Exp Med (2008) 1, 98-116
www.ijcem.com/IJCEM801004

Review Article

Biomarkers for Type 1 Diabetes

Sharad Purohit[1] and Jin-Xiong She[1,2]

[1]Center for Biotechnology and Genomic Medicine, [2]Department of Pathology, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA

Received January 8, 2008; accepted February 20, 2008; available online February 29, 2008

Abstract: Type 1 diabetes (T1D) is an autoimmune disorder characterized by the immune destruction of the insulin producing β cells of the pancreatic islets. Autoimmunity towards pancreatic antigens results from complex interactions between multiple genes, environmental factors and the immune system. The autoimmune process may occur many years before the onset of clinical diabetes and this long asymptomatic period provides excellent opportunities for the prediction and prevention of the disease. Research in past four decades has identified a number of risk factors including susceptibility genes, gene and protein expression changes, cellular changes as well as environmental triggers, which may serve as excellent biomarkers for risk assessment. Furthermore, demographic and clinical parameters such as age and family history of diabetes and other autoimmune diseases are also important for risk assessment. Despite the identification of multiple useful biomarkers, the existing tests for T1D prediction are still imperfect and earlier biomarkers are also urgently needed. Because of the insufficient predictive power of individual risk factors, future biomarkers with better predictive power will most likely take advantage of the combinatorial power of multiple biomarkers of different nature and the integration of various biomarkers and demographic/clinical information will be the key to success.

Key Words: Type-1 diabetes, biomarkers, genomics, proteomics, bioinformatics, prediction

Introduction

Autoimmune diseases, as a whole, affect a significant proportion of the population, especially in elderly women (http://www.wrongdiagnosis.com/a/ai/prevalence.htm). It is now recognized that autoimmune diseases should be considered as a related group of diseases because they share common genetic and immunological mechanisms[1]. Type 1 diabetes (T1D), also known as insulin dependent diabetes mellitus (IDDM), is primarily a childhood autoimmune disease with selective destruction of the pancreatic β-cells, leading to insulin insufficiency over time [2]. Genetic susceptibility in synergistic combination with environmental triggers leads to the development of immune response towards self antigens expressed by the pancreatic β-cells that produce insulin, resulting over time in the loss of β-cell mass and finally loss of glucose homeostasis. The incidence of T1D varies in different countries and populations [3,4] with Finland, Scandinavia and Sardinia having the highest incidences (30-50/100,000 per year). The incidence is much lower in Asian countries (1-2/100,000 per year), while it is 12-15/100,000 per year in the US Caucasians [3-7]. Approximately 90% of cases are sporadic, occurring in individuals with no family history of T1D. However, first degree relatives (FDR) of patients with T1D are at increased risk compared to the general population. In Caucasians, the risk of T1D in the general population is 0.4% and the risk of siblings of affected individuals is about 6%, approximately 15 fold higher than the general population [8-10].

It is believed that genetic susceptibility is a prerequisite for the development of T1D; however, not all genetically predisposed individuals do develop clinical disease. The vast majority (~90%) of the T1D patients develop autoantibodies against pancreatic β-cells before the clinical onset [11-15]. Although the time period between the

Table 1. Known T1D susceptibility genes and strong candidate genes

Locus	Chromosome	Genes or candidates	Function/Mechanism
IDDM1	6p21	DRB1, DQA1, DQB1	Antigen presentation
IDDM2	11p15	INS	Ag-specific T cell selection
IDDM12	2q33	CTLA4	Down-regulation of T cell activation
	1p13	PTPN22	Down-regulation of T cell activation
	12q24	PTPN11 (?)	Down-regulation of T cell activation
IDDM10	10p11-q11	IL-2Rβ	T cell activation
	12q13	IKZF4 (?)	T cell function
IDDM5	6q25	SUMO4	Control NFKB & cytokines
IDDM18	5q31-33	IL12B	Proinflammatory cytokines
	2q24	IFIH1/ GCA / KCNH7 (?)	Innate immune response / antigen presentation / insulin secretion
	6p21	ITPR3	Insulin secretion

appearance of autoantibodies and clinical onset varies greatly, it usually takes years for the clinical disease to occur [15]. Furthermore, only a proportion of the autoantibody-positive individuals will progress to clinical diabetes. This lengthy asymptomatic period, from genetic predisposition to prediabetes marked by autoimmunity (autoantibodies and cellular immunity) and finally to clinical disease, provides excellent opportunities for disease prevention. However, prevention for human T1D is still not available today for many different reasons including the difficulties of accurately identifying sufficient number of high risk population, our inability of conducting large numbers of clinical trials, and heterogeneous and poorly understood etiology of the disease. Therefore, prevention tailored for the whole at-risk population may not be effective and personalized prevention strategies based on one's own risk and etiology may prove to be more efficient. To achieve these ambitious goals, biomarkers for the disease process are urgently needed for both risk assessment and more importantly for tailoring and monitoring therapies. In this review, we will focus on the existing knowledge and recent development in the area for T1D biomarkers and discuss pitfalls of previous studies and potential solutions.

Susceptibility genes

The increased risk in siblings versus the general population ($\lambda s = 15$) as well as the high concordance rate in identical twins (~50%) are indications of the importance of genetic factors in T1D pathogenesis [16-19]. Search for genetic factors implicated in T1D started in the 1970s and the effort intensified in the 1990s. The search for genetic factors in all common diseases including T1D had three distinct stages, each corresponding to technological advances and new theoretical realization. The initial phase (1970s – 1991) consisted of testing well known candidate genes. Initially, a strict case control study design was employed and later study design has incorporated family-based association/linkage approach [20-23]. Candidate gene analysis was made possible by the discovery of protein-based polymorphisms such as the human leukocyte antigens (HLA) and DNA polymorphisms such as the variable number of tandem repeats (VNTR) in the insulin (INS) gene [23-25]. This strategy was initially quite successful and has been attempted for over 100 T1D candidate genes from the very beginning till this day. All but one confirmed T1D genes currently known (Table 1) have been discovered using this candidate gene approach. The first T1D susceptibility region, e.g., HLA (IDDM1), was initially discovered in the early 1970s using case control studies [20-23]. Genes within the HLA have since been extensively studied in multiple populations [26-36]. It turned out that the HLA is the most important susceptibility gene for T1D and many other autoimmune diseases. The region harbors multiple classical and non-classical HLA genes implicated in susceptibility and protection of T1D. Insulin (IDDM2) is the second susceptibility gene discovered for T1D. It was initially found to be associated with T1D in the early 1980s using case control studies and

then confirmed using family-based studies [23-25,37,38]. The third confirmed T1D gene was not discovered until the late 1990s when another well known candidate gene, CTLA4, was tested by association studies using both population and family-based study designs [38-42]. More recently, candidate genes including PTPN22 [43-46] and IL2RA [47,48] have been shown to be T1D genetic factors (Table 1).

Despite the initial and more recent success of the candidate gene approach, it does not allow the systematic discovery of all genetic factors until all genes have been tested. The discovery of large number of genetic polymorphisms, namely the microsatellite markers that are abundant and distributed across the whole genome, allowed the implementation of a more systematic approach to investigate the genetic factors for common diseases. The approach chosen by all investigators was to search for linkage in the entire genome using patient pedigrees, an approach that was widely successful for single gene disorders and lead to the discovery of a large number of Mendelian disease genes since the late 1980s. Because the limited availability of extended pedigrees for common diseases like T1D, affected sibpair analysis was the popular method of choice for linkage analysis [23]. Research in the 1990s in multiple laboratories has lead to the identification of over 20 suggestive linkage intervals including the regions for several known susceptibility genes [23]. Despite the localization of these linkage intervals, the studies were all underpowered and many linkage intervals may not true and others may escape detection. The disease genes and etiological mutations have not been identified in the new linkage intervals with one notable exception, e.g., IDDM5. IDDM5 was mapped to the 6q25 region. A combination of linkage, association and functional studies using genetic markers in the region identified SUMO4 as the susceptibility gene [49]. This is the only T1D susceptibility gene identified using linkage/positional cloning techniques. The gene turned out to be relatively strong and uniform in the Asian populations; however, the association between SUMO4 and T1D in the Caucasian populations is inconsistent and much weak than in the Asian populations [36,50-54]. The lack of success with linkage analysis and positional cloning can be attributed to many factors including the tedious nature of positional cloning, small effect size of the gene and lack of sufficient number of families required for such studies. However, the lack of progress in individual research laboratories lead funding agencies (NIH and JDRF) to create the worldwide T1D Genetics Consortium (T1DGC), which has the goal of collecting thousands of sibpair and simplex families with T1D from all over the world [55]. The T1DGC collection should have been a valuable resource for the studies of T1D genetics. However, its potential has not been realized. This is partly because linkage analysis has proven not to be an ideal method for studying complex disease genes and the field of complex disease genetics has shifted its focus to the third stage of analytical approaches, genome-wide association (GWA) [56,57].

As a by product of the human genome project, a huge number of single nucleotide polymorphisms (SNPs) have been identified in the human population. These SNPs provided a new set of tools critical for genetic analysis of complex traits. In addition, the development of affordable high throughput technologies to rapidly analyze large numbers of SNPs has revolutionized how complex diseases are studied. These technologies allowed the analyses of 0.5 – 1 million SNPs in the whole human genome and thousands of patients and controls. Two whole genome scans [56,57]and one scan using a large number of non-synonymous SNPs have been reported for T1D [58]. These studies identified several new regions associated with T1D. As cases and controls are much easier to ascertain, the studies can be quickly confirmed in other populations and ethnic groups. Identification of associations also speeds up the identification of the etiological mutations as association only occurs when the genetic marker is very close to the etiological mutation. It is hoped that a number of new T1D genes will be identified in the near future and functional studies will be carried out to elucidate the molecular mechanisms underlying the disease.

Identification of complex disease genes is complicated by many factors. Genetic heterogeneity is a major issue. Due to differences in gene frequencies and patterns of linkage disequilibrium in different ethnic groups or populations, associations may be found in one study but not another

[36,42,50,54,59-61]. Therefore, studies of multiple populations and ethnic groups are required to identify all susceptibility genes. Trans-racial studies are also a powerful tool for identifying etiological mutations as illustrated by the studies of the HLA genes. Gene-gene interaction is another important issue. Attempts to identify susceptibility loci that, on their own, have marginal effects by use of gene-gene interaction tests have increased in popularity. The results obtained from analyses of epitasis are, however performed at small scale and difficult to interpret. Gene-gene interaction, albeit only marginally significant, has recently been reported for the interleukin-4 and interleukin-13 genes (IL4 and IL13) with the interleukin-4 receptor A gene (IL4RA), contributing to the susceptibility of T1D [62-64]. There is still uncertainty concerning the joint action of the two established T1D susceptibility loci, the HLA class II genes (IDDM1) and the insulin gene (IDDM2) [65].

Unlike single gene disorders, complex diseases are influenced by environmental (or non-genetic) factors. As a result, the correlation between genotype and phenotype is not perfect, with disease outcome being determined by interactions between susceptibility genes and environmental determinants. Identification of susceptibility genes will aid the identification of environmental factors, which will in return aid the discovery of other susceptibility genes through analysis of gene-environment interactions. The TEDDY (The Environmental Determinants of Diabetes in the Young) study was designed precisely with these goals in mind and should result in the discovery and confirmation of both genetic and environmental factors [66,67].

Both genome-wide linkage and association scans indicated that the most important T1D genes reside within the HLA region. The HLA region may confer up to 50% of the total genetic risk according to some estimates, but much lower according others [32]. Irrespective of the precise percentages, only the HLA class II genes have been used for assessing T1D risk at this time. Genotyping the HLA class II loci has been used in several large population-screening programs including the PANDA, DIPP, DAISY, Dewit, and DiPiS studies [68]. More recently, TEDDY has also adopted the strategy to screen for high risk subjects using HLA class II genes in order to identify the environmental triggers of T1D through long term fellow up studies of the genetically at-risk cohort [66,67]. The HLA genes in T1D have been extensively reviewed elsewhere and the details will not be discussed here [8,9,69]. We should emphasize that the HLA-based test has very low specificity. For example, the HLA genotype with the highest risk (DR3/4 with DQB1*0302) has only a life-time risk of about 7% for a subject without a diabetic FDR and 20-30% of risk for a FDR of a T1D patient [8]. This genotype only identifies 30-50% of the T1D patients depending on the populations [8]. Four HLA genotypes from the general population are eligible for TEDDY follow-up: DR3/4, DR4/4, DR3/3 and DR4/8 (DR3 haplotype contains DQA1*0501-DQB1*0201, DR4 haplotypes contains DQA1*030x-DQB1*0302 and DR8 haplotype contains DQA1*0401-DQB1*0402). Adding additional genotypes will undoubtedly increase the sensitivity to identify T1D patient but will decrease the specificity of the test. The frequencies of these high risk genotypes in the general population and T1D patients vary significantly among different ethnic groups as well as geographic populations. For example, these four genotypes can identity approximately 62% of the Caucasian patients but only 32% of African American patients in Georgia (unpublished data). Therefore, it is essential to design population-specific inclusion criteria for HLA-based screening programs.

The specificity and sensitivity as well as the positive predictive value need to be improved for the test to become clinically useful. Adding non-HLA genes to the test should enhance the value of genetic testing; however, a common feature for the “non-HLA genes” is their small effect on the disease. Therefore, any single gene does not offer great value for disease prediction. It is hopeful that their predictive value can be realized when multiple genes are used in combination. One of the main areas of future genetic studies should focus on figuring out how to use the HLA and non-HLA genes for risk assessment.

Gene expression

The human genome project has lead to the identification of some 32,000 genes in human cells. The expression levels for this complete set of genes can now be assessed using microarray technology. This advancement has

fundamentally changed how investigators approach biomedical questions and provides unparalleled opportunities for biomarker discovery. Multiple microarray platforms including the Affymetrix and Ilumina systems have been developed and are widely used for gene expression profiling studies. Dissection of global changes in gene expression during pre-disease states, disease progression, and following clinical treatment can provide great insight into disease mechanism and treatment management. Microarray has been applied to many cancer studies with great success. For example, early investigations distinguished acute myeloid and acute lymphoblastic leukemia cells using gene expression profiling [70]. Subsequent studies have used microarray technology to predict outcomes in breast and ovarian cancers [71,72]. Additionally, classification of diffuse large B-cell lymphomas on the basis of gene expression profiles can identify clinically significant subtypes of cancer and the new classification has significant prognostic implications [73]. Microarray has also been applied to the studies of several autoimmune diseases. Examination of systemic lupus erythematosus (SLE) using microarray technology identified a subgroup of patients who may benefit from new therapeutic options [74,75]. Novel treatments for diseases, such as multiple sclerosis, have also been suggested by gene expression profiling [76,77]. Our group has extensively used microarray to study T1d in both animal models and human subjects. A number of these genes are differentially expressed during the progression to disease in the NOD mice [78]. Similarly, over 100 genes are up-regulated in T1D subjects [79]. Most of these genes are also up-regulated in prediabetic subjects, suggesting that they may be useful predictive markers. Many of the differentially expressed genes were found to be involved in important immunological functions including antigen processing and presentation, cytotoxicity and apoptosis, and immune regulation [79]. Upregulation of several proinflammatory mediators and markers was found between diabetic and prediabetic subjects [79]. The expression profile of several groups of genes, important for cell proliferation, transcription and translation, as well as mitochondrial genes was consistent with the earlier reports about cellular activation in T1D and prediabetic subjects, as expected in an autoimmune state due to increased cellular activation and proliferation [79]. Some of the genes were also identified by a microarray study in rheumatoid arthritis patients [80], suggesting that autoimmune diseases may share common expression profiles for certain genes and genetic pathways. One challenge for T1D studies in human subjects is the limited access to the pathologic tissues where the molecular and cellular events take place. Although changes in the peripheral blood may exist, the extent of difference is usually small as suggested by our global gene expression studies. Further, there is large individual variations in both patient and control samples. Therefore, large number of samples may have to be analyzed to identify and confirm the differences. Our experience indicates that at least several hundred subjects in each group must be studied to validate the expression differences that we have seen in the microarray datasets.

Islet autoantibodies and autoreactive T cells

The presence of autoantibodies against islet antigens is a hallmark for the development of T1D. Several autoantibodies have been detected before or at the onset of the disease [81]. Four islet autoantibodies appear to be the most useful T1D markers. These are islet cell autoantibodies (ICA), autoantibodies to insulin (IAA), autoantibodies to the tyrosine phosphatases IA-2 and IA-2β and autoantibodies to glutamate decarboxylase (GADA)[82-86] Other β-cell autoantigens include carboxypeptidase-H, islet cell antigen (ICA)-69, GM Gangliosides, a 38 kd autoantigen and SOX13. Novel autoantibodies are still being discovered [87], for example, the zinc transporter (ZnT8) which is the most recently discovered T1D autoantigen [87]. ICA is the first major autoantibody identified for T1D. At the onset of T1D more than 70% of the patients are positive for ICA [83]. ICA measurement is technically challenging due to the semi-quantitative indirect fluorescent assay systems, which are difficult to standardize despite of significant improvements. IAA is normally among the first autoantibody that appears in young children[86]. At disease onset IAA is found in 35-60% of young children but their frequency is less in teenagers and adults[86]. IAA by itself is not highly predictive but can be used in combination with other islet autoantibodies for prediction [11,12,86]. Autoantibodies to GAD have been proven to be valuable markers for

T1D and are detectable many years before the clinical onset of the disease and are seen in both ICA positive and ICA-negative FDR subjects. Approximately 70-80% of newly diagnosed T1D patients and 3-5% of FDR have autoantibody to GAD. IA-2A and IA-2B autoantibodies are detected in more than 55-75% of newly diagnosed T1D patients [82,84-86]. These autoantibodies normally appear after IAA and/or GADA [88].

The zinc transporter autoantibodies (ZnT8A) were detected in 60-80% of new-onset T1D compared with <2% of controls and <3% type 2 diabetes patients and in up to 30% of patients with other autoimmune disorders with a T1D association. ZnT8 autoantibodies (ZnT8A) were found in 26% of T1D subjects classified as autoantibody-negative on the basis of existing markers (GADA, IA-2A, IAA, and ICA). ZnT8A were detected as early as 2 years of age and increasing levels and prevalence persisting to disease onset in prospective monitoring. ZnT8A generally emerged later than GADA and IAA in prediabetes, although not in a strict order. The combined measurement of ZnT8A, GADA, IA2A, and IAA raised autoimmunity detection rates to 98% at disease onset [87].

T1D prediction is greatly enhanced by islet autoantibodies [14,89-91]. Although the pathogenic implications of circulating autoantibodies are not fully understood, their use as indicators of islet cell destruction and impending clinical disease allow for the reasonable identification of individuals at increased risk for developing T1D, especially in conjunction with high-risk genetic factors and/or a family history of the disease. Many studies have shown that the presence of autoantibodies against islet antigens is very useful for T1D prediction in the FDR of diabetic patients as well as in the general population [14,81,90,92]. The risk of developing diabetes is strongly correlated to the number of autoantibody markers, that is, the presence of two or more autoantibodies gives a higher probability of developing the disease than the presence of single antibody [81,93,94]. The autoantibody assays have constantly been improved and the performance in most laboratories is quite excellent [95-97]. However, it is still very challenging to obtain consistent results across different laboratories for the low titer autoantibodies near the cutoff and a good proportion of the low titer autoantibodies may be false-positives. However, most of the discrepancies may be resolve when the autoantibody titer increases. Despite the utility of the autoantibodies in T1D prediction, they have several serious limitations. First, the appearance of autoantibodies marks a relatively late stage of the autoimmune process and therefore is not suitable for early disease intervention. Second, only a subset of the autoantibody-positive subjects will progress to clinical diabetes and therefore it would be useful to have biomarkers that allow the distinction of the progressors versus non-progressors. Third, autoantibodies are not useful as biomarkers for therapeutic outcomes.

Insulitis, infiltration of lymphocytes into pancreatic islets, evolves through several discrete stages that culminate in β-cell death. In the first stage, antigenic epitopes of β-cell-specific peptides are processed by antigen presenting cells in local lymph nodes, and autoreactive lymphocyte clones are propagated. Subsequently, cell-mediated and direct cytokine-mediated reactions are generated against β-cells, and the β-cells are sensitized to apoptosis. Antigen specific immune reactions are believed to be involved in the destruction of pancreatic β-cells. Pancreatic β-cell autoantigens are the targets of immune mediated destruction. Autoreactive T-cells directed against β-cell autoantigens should be excellent markers for T1D and other autoimmune diseases [98,99]. However, currently available T cell assays are highly variable and lack sensitivity [100-103].

Cytokines, chemokines and other serum proteins

Cytokines and chemokines are important mediators of immune responses due to their ability to recruit and activate leucocytes and other immune cells. Cytokines have been proposed as inducers of β-cell damage in human T1D via the generation of NO [104,105]. Type-1 cytokines such as IL-2, IFN-γ and TNF-α dominate over an immunoregulatory subset of cytokines viz., IL-4, IL-5 and IL-13, leading to imbalance between these two subsets. This allows the type-1 cytokines to initiate a cascade of immune-inflammatory processes in the islet, which includes activating macrophages to produce proinflammatory cytokines. The

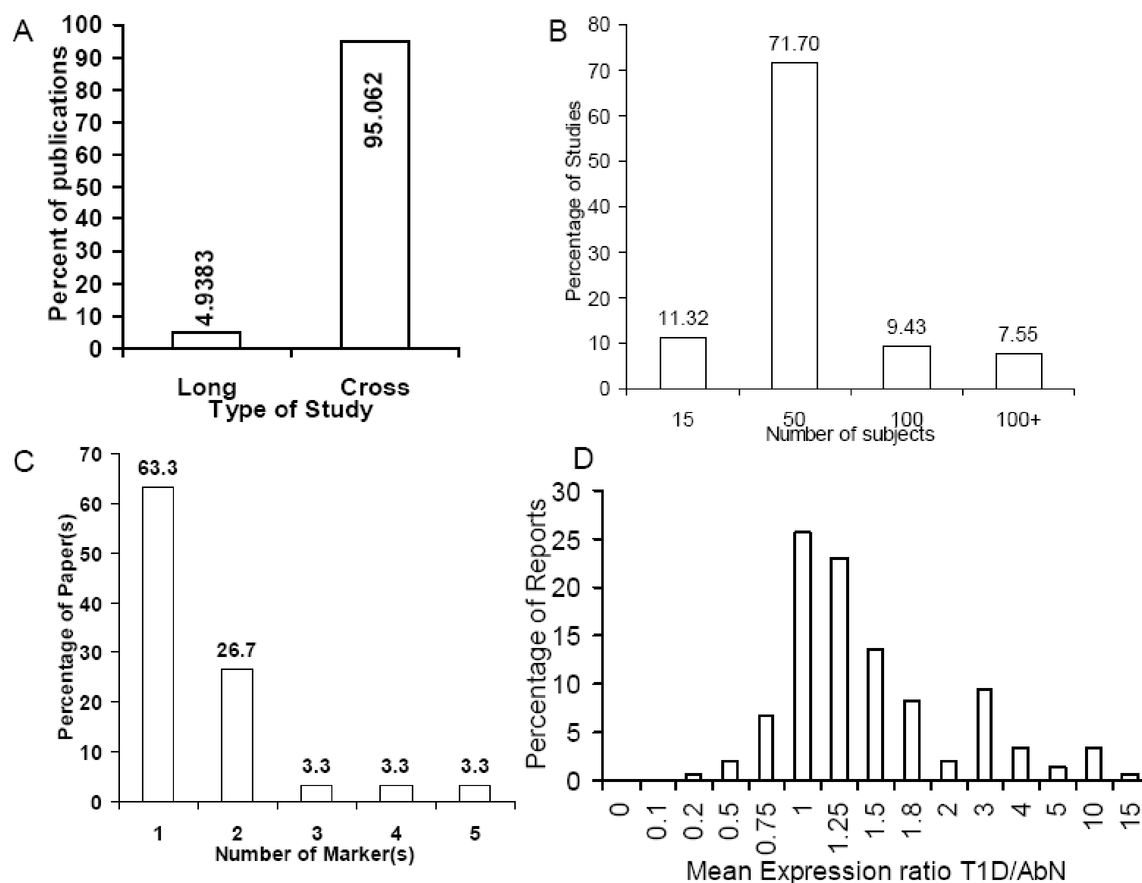

Figure 1: Literature survey on systemic levels of cytokines, chemokines and other serum proteins. A: type of studies performed, B: Number of subjects involved in the study including controls and patients, C: Number of markers analyzed in each study, D: Mean expression differences observed between the controls and T1D patients.

proinflammatory cytokines IL-1β, IL-6 and TNF-α have cytotoxic, cytostatic, or cytocidal actions on pancreatic islets by inducing NO production [106-108]. Recent reports suggest that the pancreas participates in TNF-α production during stress and the islets are predominantly responsible for this synthesis. IL-1β and TNF-α are important for the β-cell lyses in T1D, while IL-1 receptor antagonist (IL-1Ra) is considered protective by blocking the effects of IL-1. *In vitro* TNF-α and IL-1β inhibit insulin release from β-cells [104,105]. It appears that the process of autoimmune aggression against β-cells and its effect on insulin release and glucose homeostasis is a slow and chronic process [109]. In some studies performed with newly diagnosed T1D patients, production of IL-1 was found to be increased significantly when compared with chronic T1D patients and healthy controls [110]. Circulating concentrations of IL-1Ra in chronic T1D patients was increased; with no changes in TNF synthesis [111,112]. A proinflammatory imbalance in T1D patients may play an important role in β-cell loss.

Local generation of chemokines by islet cells may be important in the initiation and regulation of inflammatory processes during insulitis. This hypothesis is supported by several reports that demonstrated that high levels of MCP-1, IP-10 are released by islets cells during autoimmune attack [113-116]. IP-10 is a member of the CXC family of chemokines [116]. It attracts activated T-helper 1 (Th1) and natural killer (NK) cells expressing the CXCR3 receptor [116]. Serum concentration of IP-10 has been found to be elevated in new onset T1D patients and in autoantibody-positive relatives [117,118].

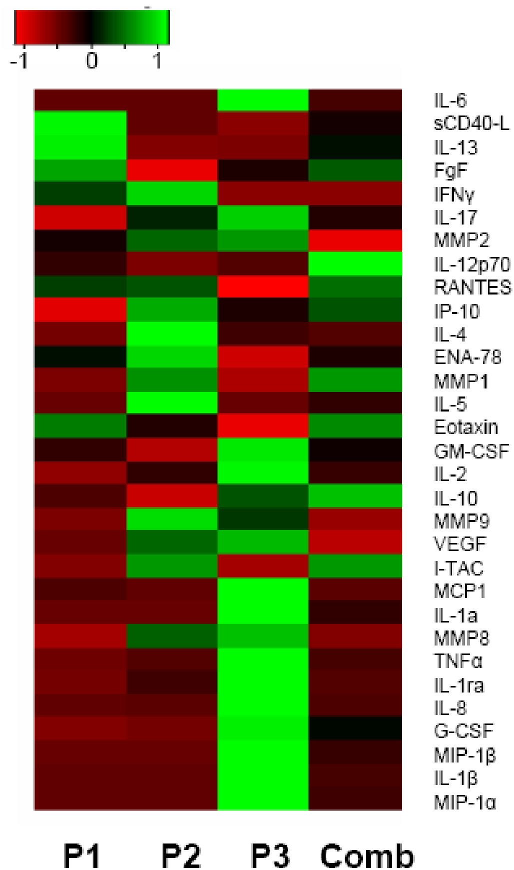

Figure 2: Heatmap showing the mean expression ratios between T1D and age-matched controls for serum levels of cytokines, chemokines and MMPs. Each plate consists of 40 patients and 40 controls. P1: Plate 1, P2: Plate 2, P3: Plate 3, Comb: combined data from all three plates.

Circulating levels of soluble adhesion molecules have been studied in T1D patients but the results are usually inconsistent. Several studies have found an increase in serum concentrations; however, other studies reported no difference or even decreased levels [119-125]. Whether the increased concentrations of soluble adhesion molecules represent spillover from an active destructive process or a compensatory mechanism by which the immune system tries to protect the target tissue against destruction remains an open question. The physiological role of soluble adhesion molecules is unknown but considering the crucial impact of cell adhesion molecules in lymphoid-endothelial interactions increased concentrations in circulation, shed from the cell surface, could be an epiphenomenon of immune activation and thus might provide a useful monitor of disease activity in inflammatory disorders. So far the studies on the role of soluble adhesion molecules in T1D have been cross-sectional, with no proper data on the dynamics of these molecules in pre-clinical diabetes. The only large study published to date is the EURODIAB prospective complications study group that analyzed 540 cross-sectional subjects. In this study a positive relationship was reported between unadjusted values of sVCAM-1 and sE-selectin with non-proliferative and proliferative retinopathy, micro- and macroalbuminuria and CVD. After adjusting for age, sex, duration of diabetes, BMI, and other complications strong significant associations were found between sVCAM-1 and macroalbuminuria [124].

A PubMed search on serum proteins under category of cytokines, chemokines, matrix metalloproteases, acute phase proteins and other soluble proteins; results in a total of 270 reports. We analyzed these reports in terms of number of subjects involved in the study, type of study, number of markers and mean differences between the groups (**Figure 1**). All published literature reported positive relationships between various cytokines, chemokines and adhesion molecules in T1D patients. It is apparent that there studies have many limitations. The vast majority of these studies (95%) are of cross-sectional design (**Figure 1A**). Therefore, it is difficult to know whether the serum protein changes are involved in the disease process or consequence of the metabolic changes of diabetes. Another major drawback is the small number of subjects, resulting in low study power. As shown in **Figure 1B**, 80% of the studies were performed with ~50 subjects including patients and controls. In studies with small number of subjects, many of the observations could be due to random variation, rendering the observations difficult to replicate. On the other hand, true differences may be missed because of insufficient power to detect smaller differences with small sample size. Therefore, all studies with small sample size would not allow any firm conclusion, whether positive or negative. To illustrate this point, we performed an extensive study of serum proteins using the LUMINEX assays for 31 serum proteins and

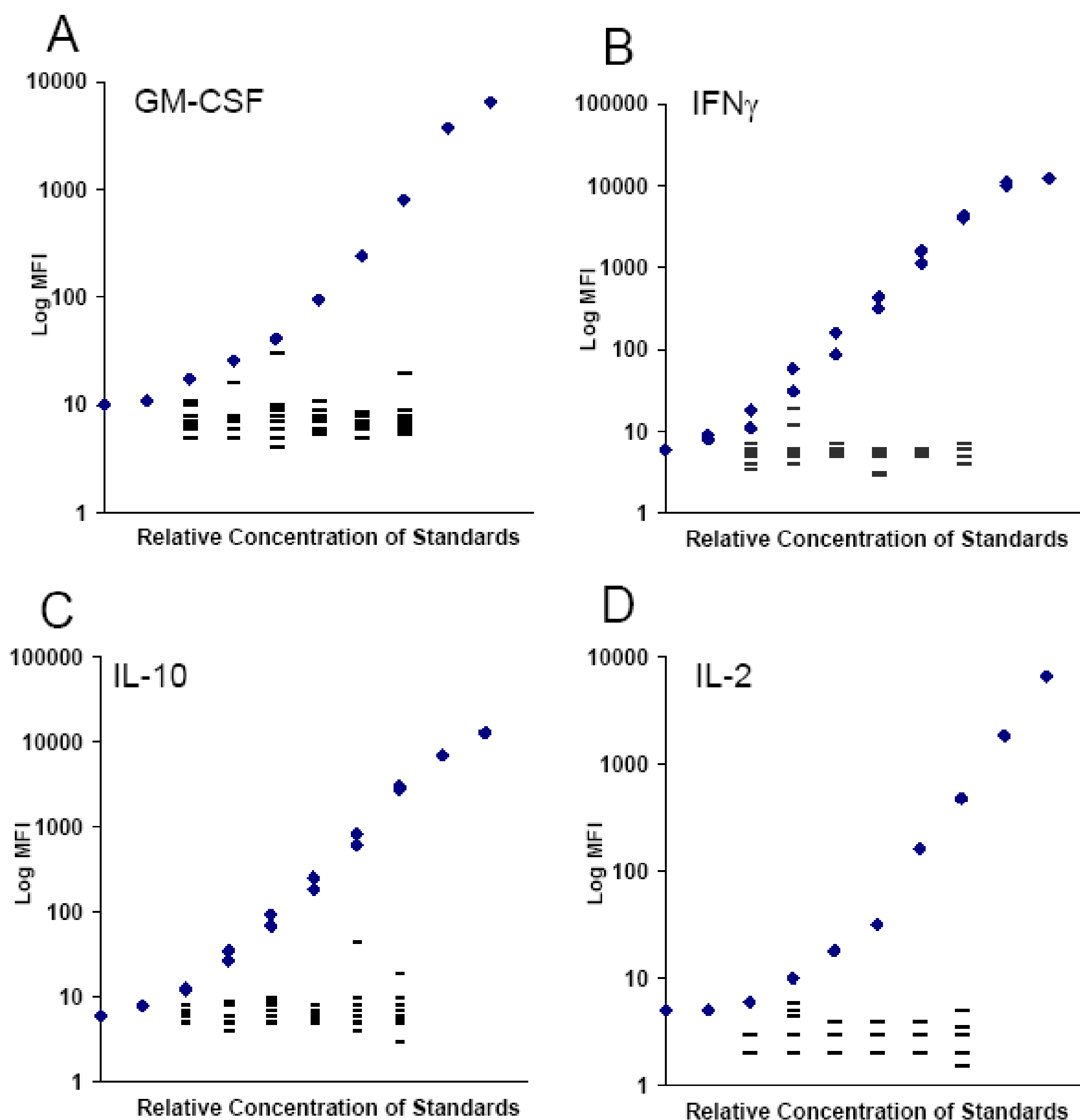

Figure 3: Representative charts showing distribution of mean fluorescent intensities (MFI) for selected proteins relative to their standard curves. Standards were serially diluted 1:3 to create 9 point standard curve(s). Each concentration was assayed in duplicates (blue diamonds). MFI's of the samples (dashes) were then plotted along these standards to display distribution of MFI's across the standard curve(s). Most samples have MFI's below the lower end of the linear range of their respective standard curve.

240 samples (120 T1D patients and 120 controls) randomized on three 96-well plates. Each plate contains 40 patients and 40 controls. **Figure 2** presents the T1D/control ratios for each serum protein and each plate. It is immediately apparent that the ratios estimated using small sample sizes are highly variable. Based on our experience, hundreds to thousands of samples may have to be analyzed to estimate the true mean concentrations and variations for most serum proteins.

A third conclusion from the literature review is the small degree of differences between T1D and controls. The highest difference is 1.5 fold

or lower in more than 50% of studies (**Figure 1D**). Although a few studies observed large differences in the means between T1D and controls, the differences are most likely caused by the extreme expression levels in a small number of subjects. In these cases, the median expression levels are usually not significantly different. A fourth conclusion from the literature review is the small number of molecules analyzed in each report (90% of the studies assessed only one or two molecules) (**Figure 1C**).

Serum proteins that may be used as biomarkers for the disease process are usually present in sub picomolar range. With the complexity of the serum proteome and the very high dynamic range, one has to pay attention to the sensitivity of the assay. Sensitivity of the assay is the capability of the assay to detect low abundance proteins. Assay sensitivity is especially critical for high throughput assays. Unfortunately, most assays available today are not ideal for serum analysis. As shown in **Figure 3**, the serum concentrations for many serum proteins such as GM-CSF, IFNγ and interleukins are below the linear range of the assay. Measurements for such proteins are not reliable and can lead to wrongful conclusions. For generation of quality data it is necessary that most samples have concentrations within the linear range of the standard curve. The sample volume should be adjusted to achieve good results. Unfortunately, better assays have to be developed for some proteins if their concentration is below the detection limit of the currently available assays.

Proteomics

Human serum contains thousands of small- to medium-sized peptides, as a result of various cellular activities, known as the serum proteome/peptidome. These peptides and proteins may provide valuable information about the health status of the individual or disease state and thus have a huge potential in discovery of new markers for T1D. The initial efforts in proteomics have focused on protein identification. Recent mass spectrometry (MS)-based technology developments have provided useful platforms for both protein identification and quantification. Quantitative analysis of global protein levels, termed 'quantitative proteomics', is important for the system-based understanding of the molecular function of each protein component and is expected to provide insights into molecular mechanisms of various biological systems. Several methods are widely used to generate global quantitative protein profiles, including two dimensional (2D) gel electrophoresis followed by MS analysis, stable isotope labeling-based quantification, MS signal intensity-based quantification and protein array-based quantification [126]. The mass pattern from MS analysis, which does not require a high end mass spectrometer, provides the global changes in the protein profiles in health and disease states [127-129]. Pattern-based analysis was once a method of choice in the recent past for prediction purposes and was touted highly for biomarker discovery. The MS patterns do not provide the sequence identity of the proteins for the development of more reliable assays and hence validation of pattern-based biomarkers is very difficult. Similar approach was applied to T1D studies using SELDI-TOF in our lab [130]. It was shown that a large number of serum proteins may differ between T1D and healthy controls. However, no single protein is able to distinguish T1D from controls but the use of multiple proteins hold promise as potential biomarkers for T1D [130]. However, the proteins were not identified and follow-up studies were not feasible because of drift of the MS profiles over time.

Proteomic studies are still in the infancy stage. Many issues remain to be solved before complex proteomes like the serum can be fully analyzed. A common feature of biological samples is their extraordinary complexity, which is a result of the high multidimensionality of their protein constituents. These proteins differ in their cellular and sub-cellular distribution; their occurrence in complexes; their charge, molecular mass and hydrophobicity; and their expressed levels and post-translational modifications (PTM). Due to the high dynamic range in protein concentration, the proteins that are readily analyzable by currently available methods are limited to the abundant proteins and low abundance proteins that are promising biomarkers are difficult to detect and quantify. This difficulty is particularly noticeable in body fluids such as serum, where more than 99% of the protein complement consists of serum albumin and globulins [131]. Analyzing biological fluid proteomes with a vast dynamic range of 1012- 1015 of

their protein abundance and occurrence of multiple protein isoforms, presents a major challenge for proteomic studies. Hence, a major effort in modern proteomics focuses on the development and application of complimentary fractionation/separation strategies that increase the detection and quantification of low abundance proteins. Most fractionation strategies utilize the chemical and physical properties, as well as post transcriptional modifications of proteins for development of orthogonal separation strategies. Immuno-affinity based capture and/or depletion represents another well established approach to enrich protein subsets of interest. Multi modular combinations of liquid chromatography provide options for deconvulation of complex mixtures of proteins. Given the complex nature of mammalian proteomes, all these complimentary separation and enrichment techniques will facilitate the discovery of biomarkers. Despite these recent advances, proteomic technologies still need significant improvement in several areas to become a powerful tool for biomarker discovery that it is expected to be.

Computational technologies and multivariate models for prediction

Previous and ongoing studies indicate that a large number of molecules may differ between T1D and controls; however, the degree of differences is too small for each single molecule to be an ideal biomarker. The accuracy of predictions can be significantly improved by using multiple molecules simultaneously. In this regard, new computational approaches are being developed for selecting an optimal subset (model) or subsets of predictive molecules and assessing the prediction value of the models. A number of methods can be used for model selection based on classification of subjects into known classes. Discriminant analysis, a set of multivariate techniques, can be used to classify samples into known categories. Many different models can be used in the discrimination including parametric (linear and quadratic discriminators) and nonparametric (e.g., kernel based discriminators, k-nearest neighbor discriminators) discriminators. Logistic regression is related to the parametric discriminators and logistic regression could be used to estimate the probability that an individual would get a complex disease [132]. However, logistic regression suffers from the inability to accurately estimate the needed parameters when the two groups are perfectly separated based on the variables included in the model. This problem could potentially be overcome using exact logistic regression [133]. However, genomic and proteomic data usually have a large number of variables relative to the number of study subjects resulting in sparse data, which may result in inaccurate estimates of the parameters needed to predict the status of new subjects. Principal components [134] and Classification and Regression Trees (CART) [135] are also widely used methods that can be used to classify subjects and predict new subjects' status [136,137]. Neural networks and support vector machines are more complicated methods requiring more "training" than the other methods. All of these methods have been used at least once with proteomic or microarray data; although only one study has compared these methods using the same data, finding that relatively simple discriminators tended to perform best [138].

Irrespective of the statistical method, it is desirable to use only a subset of all possible molecules because of several considerations: (1) overfitting is a serious issue when a large number of molecules are used; (2) fewer molecules often give better predictions; and (3) it is more economic to assay for fewer molecules. The simplest solution to this problem has been to focus more on those variables that have the largest differences between the groups being examined [134,138]. However, such an approach ignores an important nature of the multivariate data; that is, molecules exhibiting small univariate differences can have a large contribution in a multivariate analysis. Models with different numbers and combinations of variables must be examined to find an optimal set of molecules for multivariate models. However, examining all possible models would require evaluating 2p-1 models, where p is the number of variables included in the study (e.g., for a dataset with 100 proteins, the total number of models is approximately 1030) [79]. Such a task is not feasible when hundreds or thousands proteins are studied. One approach is to systematically search for an optimal set of molecules. In stepwise procedures, variables are added and/or removed at each step depending on a significance test or some measure of

information contributed by that variable to the difference between the groups. This procedure continues until no variables can be added or removed. Although stepwise procedures rely on tests or information for a single variable, all decisions are based on multivariate analyses [79].

Regardless of the selection method, the selected models should have low prediction error rates, which are often evaluated based on cross-validation error rates. The actual prediction error rate is likely larger than the estimated rate because the estimates of the prediction error rate tend to have a high variance, especially leave-one-out cross-validation error rates. The real predictive value of a model has to be tested with future samples and ideally in a prospective setting. Because the larger the number of molecules included the more likely that noise has been modeled rather than real differences between groups, models with smaller number of molecules are preferred if good prediction value can be achieved. In our studies of T1D [79,130] we have followed the approach along the lines of Random Forests [139] and Stochastic Discrimination [140,141] to select multiple models by randomly searching for models with very small number (<10) of molecules and low estimated prediction error rates. Such models may not have perfect prediction but the predictions can be improved using averaging of models through plurality voting. This approach may use more total number of molecules than single models, but each model does not overfit as smaller number of molecules is used and the average prediction may give excellent error rates that may be validated by future studies.

Integration of different types of biomarkers

As discussed in the above sections, a number of molecules (DNA, RNA, proteins and potentially others) as well as other information (age, family history of diabetes and other autoimmune diseases, environmental factors) can influence the development of T1D and potentially the therapeutic response. However, each of these risk factors alone does not determine the outcome of disease because of the multifactorial nature of the disease. How these risk factors interact to cause the disease is not well understood at this time as investigators are still focusing on the elucidation of individual risk factors. It is hoped that the identification of one category of risk factors should aid the discovery of other risk factors through studies of interactions between different risk factors. Precise prediction of disease can only be achieved by integrating all risk factors and biomarkers.

Acknowledgements

This work was partially supported by grants from the National Institutes of health (2RO1HD37800, 4R33DK069878, 4R33HD050196, 5U24DK58778, 2P01AI-42288) and Juvenile Diabetes Research Foundation International (JDRF 1-2004-661) to JXS. SP is recipient of Advanced Postdoctoral Fellowship from JDRF, New York (10-2006-792).

Address correspondence to: Jin-Xiong She, PhD, Center for Biotechnology and Genomic Medicine, Medical College of Georgia, 1120 15th St, August, GA 30912, Fax: (706) 721-3688, Email: jshe@mail.mcg.edu.

References

[1] Anaya JM, gado-Vega AM and Castiblanco J. Genetic basis of Sjogren's syndrome. How strong is the evidence? Clin Dev Immunol 2006; 13: 209-222.

[2] Atkinson MA and Maclaren NK. The pathogenesis of insulin-dependent diabetes mellitus. N Engl J Med 1994; 331: 1428-1436.

[3] Karvonen M, Tuomilehto J, Libman I and LaPorte R. A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. World Health Organization DIAMOND Project Group. Diabetologia 1993; 36: 883-892.

[4] Tuomilehto J, Lounamaa R, Tuomilehto-Wolf E, Reunanen A, Virtala E, Kaprio EA and Akerblom HK. Epidemiology of childhood diabetes mellitus in Finland--background of a nationwide study of type 1 (insulin-dependent) diabetes mellitus. The Childhood Diabetes in Finland (DiMe) Study Group. Diabetologia 1992; 35: 70-76.

[5] Gorodezky C, Alaez C, Murguia A, Rodriguez A, Balladares S, Vazquez M, Flores H and Robles C. HLA and autoimmune diseases: Type 1 diabetes (T1D) as an example. Autoimmun Rev 2006; 5: 187-194.

[6] Awdeh ZL, Yunis EJ, Audeh MJ, Fici D, Pugliese A, Larsen CE and Alper CA. A genetic explanation for the rising incidence of type 1 diabetes, a polygenic disease. J Autoimmun 2006; 27: 174-181.

[7] Kawasaki E, Matsuura N and Eguchi K. Type 1 diabetes in Japan. Diabetologia 2006; 49: 828-836.
[8] She JX. Susceptibility to type I diabetes: HLA-DQ and DR revisited. Immunol Today 1996; 17: 323-329.
[9] She JX and Marron MP. Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. Curr Opin Immunol 1998; 10: 682-689.
[10] Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 1987; 40: 1-14.
[11] Bingley PJ, Bonifacio E, Williams AJ, Genovese S, Bottazzo GF and Gale EA. Prediction of IDDM in the general population: strategies based on combinations of autoantibody markers. Diabetes 1997; 46: 1701-1710.
[12] Bingley PJ, Christie MR, Bonifacio E, Bonfanti R, Shattock M, Fonte MT, Bottazzo GF and Gale EA. Combined analysis of autoantibodies improves prediction of IDDM in islet cell antibody-positive relatives. Diabetes 1994; 43: 1304-1310.
[13] Bonifacio E, Genovese S, Braghi S, Bazzigaluppi E, Lampasona V, Bingley PJ, Rogge L, Pastore MR, Bognetti E, Bottazzo GF and . Islet autoantibody markers in IDDM: risk assessment strategies yielding high sensitivity. Diabetologia 1995; 38: 816-822.
[14] Verge CF, Gianani R, Kawasaki E, Yu L, Pietropaolo M, Jackson RA, Chase HP and Eisenbarth GS. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. Diabetes 1996; 45: 926-933.
[15] Harrison LC. Risk assessment, prediction and prevention of type 1 diabetes. Pediatr Diabetes 2001; 2: 71-82.
[16] Barnett AH, Eff C, Leslie RD and Pyke DA. Diabetes in identical twins. A study of 200 pairs. Diabetologia 1981; 20: 87-93.
[17] Kumar D, Gemayel NS, Deapen D, Kapadia D, Yamashita PH, Lee M, Dwyer JH, Roy-Burman P, Bray GA and Mack TM. North-American twins with IDDM. Genetic, etiological, and clinical significance of disease concordance according to age, zygosity, and the interval after diagnosis in first twin. Diabetes 1993; 42: 1351-1363.
[18] Olmos P, A'Hern R, Heaton DA, Millward BA, Risley D, Pyke DA and Leslie RD. The significance of the concordance rate for type 1 (insulin-dependent) diabetes in identical twins. Diabetologia 1988; 31: 747-750.
[19] Kyvik KO, Green A and Beck-Nielsen H. Concordance rates of insuling dependent diabetes mellitus: a population based suty of young Danish twins. BMJ 1995; 311: 913-917.
[20] Cudworth AG and Woodrow JC. Letter: HL-A antigens and diabetes mellitus. Lancet 1974; 2: 1153.
[21] Singal DP and Blajchman MA. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes 1973; 22: 429-432.
[22] Nerup J, Platz P, Andersen OO, Christy M, Lyngsoe J, Poulsen JE, Ryder LP, Nielsen LS, Thomsen M and Svejgaard A. HL-A antigens and diabetes mellitus. Lancet 1974; 2: 864-866.
[23] Field LL. Genetic linkage and association studies of Type I diabetes: challenges and rewards. Diabetologia 2002; 45: 21-35.
[24] Bell GI, Murray JC, Nakamura Y, Kayano T, Eddy RL, Fan YS, Byers MG and Shows TB. Polymorphic human insulin-responsive glucose-transporter gene on chromosome 17p13. Diabetes 1989; 38: 1072-1075.
[25] Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon-Thomsen A, Deschamps I, Rotter JI, Djoulah S, James MR and . Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 1994; 371: 161-164.
[26] Ronningen KS, Spurkland A, Iwe T, Sollid LM, Vartdal F and Thorsby E. Novel HLA-DR2 and -DR3 haplotypes among Norwegian Caucasians. Tissue Antigens 1991; 37: 165-167.
[27] Ronningen KS, Spurkland A, Iwe T, Vartdal F and Thorsby E. Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus. Tissue Antigens 1991; 37: 105-111.
[28] Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R and Tuomilehto J. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. Diabetes Care 2000; 23: 1516-1526.
[29] Veijola R, Reijonen H, Vahasalo P, Sabbah E, Kulmala P, Ilonen J, Akerblom HK and Knip M. HLA-DQB1-defined genetic susceptibility, beta cell autoimmunity, and metabolic characteristics in familial and nonfamilial insulin-dependent diabetes mellitus. Childhood Diabetes in Finland (DiMe) Study Group. J Clin Invest 1996; 98: 2489-2495.
[30] Veijola R, Vahasalo P, Tuomilehto-Wolf E, Reijonen H, Kulmala P, Ilonen J, Akerblom HK and Knip M. Human leukocyte antigen identity and DQ risk alleles in autoantibody- positive siblings of children with IDDM are associated with reduced early insulin response. Childhood Diabetes in Finland (DiMe) Study Group. Diabetes 1995; 44: 1021-1028.
[31] Kulmala P, Savola K, Reijonen H, Veijola R, Vahasalo P, Karjalainen J, Tuomilehto-Wolf E, Ilonen J, Tuomilehto J, Akerblom HK and Knip M. Genetic markers, humoral autoimmunity, and prediction of type 1 diabetes in siblings of affected children. Childhood Diabetes in Finland Study Group. Diabetes 2000; 49: 48-58.

[32] Noble JA, Valdes AM, Cook M, Klitz W, Thomson G and Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. Am J Hum Genet 1996; 59: 1134-1148.

[33] Erlich HA, Zeidler A, Chang J, Shaw S, Raffel LJ, Klitz W, Beshkov Y, Costin G, Pressman S, Bugawan T and . HLA class II alleles and susceptibility and resistance to insulin dependent diabetes mellitus in Mexican-American families. Nat Genet 1993; 3: 358-364.

[34] Tait BD, Drummond BP, Varney MD and Harrison LC. HLA-DRB1*0401 is associated with susceptibility to insulin-dependent diabetes mellitus independently of the DQB1 locus. Eur J Immunogenet 1995; 22: 289-297.

[35] Cucca F, Muntoni F, Lampis R, Frau F, Argiolas L, Silvetti M, Angius E, Cao A, De VS and Congia M. Combinations of specific DRB1, DQA1, DQB1 haplotypes are associated with insulin-dependent diabetes mellitus in Sardinia. Hum Immunol 1993; 37: 85-94.

[36] Ikegami H, Fujisawa T, Kawabata Y, Noso S and Ogihara T. Genetics of type 1 diabetes: similarities and differences between Asian and Caucasian populations. Ann N Y Acad Sci 2006; 1079: 51-59.

[37] 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 and Spielman RS. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat Genet 1998; 19: 292-296.

[38] Kristiansen OP, Larsen ZM and Pociot F. CTLA-4 in autoimmune diseases--a general susceptibility gene to autoimmunity? Genes Immun 2000; 1: 170-184.

[39] Marron MP, Raffel LJ, Garchon HJ, Jacob CO, Serrano-Rios M, Martinez Larrad MT, Teng WP, Park Y, Zhang ZX, Goldstein DR, Tao YW, Beaurain G, Bach JF, Huang HS, Luo DF, Zeidler A, Rotter JI, Yang MC, Modilevsky T, Maclaren NK and She JX. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. Hum Mol Genet 1997; 6: 1275-1282.

[40] Nistico L, Buzzetti R, Pritchard LE, Van der AB, Giovannini C, Bosi E, Larrad MT, Rios MS, Chow CC, Cockram CS, Jacobs K, Mijovic C, Bain SC, Barnett AH, Vandewalle CL, Schuit F, Gorus FK, Tosi R, Pozzilli P and Todd JA. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum Mol Genet 1996; 5: 1075-1080.

[41] Marron MP, Zeidler A, Raffel LJ, Eckenrode SE, Yang JJ, Hopkins DI, Garchon HJ, Jacob CO, Serrano-Rios M, Martinez Larrad MT, Park Y, Bach JF, Rotter JI, Yang MC and She JX. Genetic and physical mapping of a type 1 diabetes susceptibility gene (IDDM12) to a 100-kb phagemid artificial chromosome clone containing D2S72-CTLA4-D2S105 on chromosome 2q33. Diabetes 2000; 49: 492-499.

[42] Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA and Gough SC. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003; 423: 506-511.

[43] Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D and Mustelin T. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004; 36: 337-338.

[44] Zheng W and She JX. Genetic association between a lymphoid tyrosine phosphatase (PTPN22) and type 1 diabetes. Diabetes 2005; 54: 906-908.

[45] Smyth D, Cooper JD, Collins JE, Heward JM, Franklyn JA, Howson JM, Vella A, Nutland S, Rance HE, Maier L, Barratt BJ, Guja C, Ionescu-Tirgoviste C, Savage DA, Dunger DB, Widmer B, Strachan DP, Ring SM, Walker N, Clayton DG, Twells RC, Gough SC and Todd JA. 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 2004; 53: 3020-3023.

[46] Steck AK, Liu SY, McFann K, Barriga KJ, Babu SR, Eisenbarth GS, Rewers MJ and She JX. Association of the PTPN22/LYP gene with type 1 diabetes. Pediatr Diabetes 2006; 7: 274-278.

[47] Mirel DB, Valdes AM, Lazzeroni LC, Reynolds RL, Erlich HA and Noble JA. Association of IL4R haplotypes with type 1 diabetes. Diabetes 2002; 51: 3336-3341.

[48] 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 RC, Clayton DG and Todd JA. Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. Am J Hum Genet 2005; 76: 773-779.

[49] Guo D, Li M, Zhang Y, Yang P, Eckenrode S, Hopkins D, Zheng W, Purohit S, Podolsky RH,

Muir A, Wang J, Dong Z, Brusko T, Atkinson M, Pozzilli P, Zeidler A, Raffel LJ, Jacob CO, Park Y, Serrano-Rios M, Larrad MT, Zhang Z, Garchon HJ, Bach JF, Rotter JI, She JX and Wang CY. A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. Nat Genet 2004; 36: 837-841.
[50] Park Y. Why is type 1 diabetes uncommon in Asia? Ann N Y Acad Sci 2006; 1079: 31-40.
[51] Wang CY and She JX. SUMO4 and its role in type 1 diabetes pathogenesis. Diabetes Metab Res Rev 2007.
[52] Wang CY, Podolsky R and She JX. Genetic and functional evidence supporting SUMO4 as a type 1 diabetes susceptibility gene. Ann N Y Acad Sci 2006; 1079: 257-267.
[53] Caputo M, Cerrone GE, Mazza C, Cedola N, Targovnik HM and Gustavo DF. No evidence of association of CTLA-4 -318 C/T, 159 C/T, 3' STR and SUMO4 163 AG polymorphism with autoimmune diabetes. Immunol Invest 2007; 36: 259-270.
[54] Noso S, Ikegami H, Fujisawa T, Kawabata Y, Asano K, Hiromine Y, Sugihara S, Lee I, Kawasaki E, Awata T and Ogihara T. Association of SUMO4, as a candidate gene for IDDM5, with susceptibility to type 1 diabetes in Asian populations. Ann N Y Acad Sci 2006; 1079: 41-46.
[55] Rich SS, Concannon P, Erlich H, Julier C, Morahan G, Nerup J, Pociot F and Todd JA. The Type 1 Diabetes Genetics Consortium. Ann N Y Acad Sci 2006; 1079: 1-8.
[56] Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS and Clayton DG. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007; 39: 857-864.
[57] Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS and Todd JA. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet 2007; 39: 1074-1082.
[58] 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 and Todd JA. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nat Genet 2006; 38: 617-619.
[59] Smyth DJ, Cooper JD, Lowe CE, Nutland S, Walker NM, Clayton DG and Todd JA. No evidence for association of OAS1 with type 1 diabetes in unaffected siblings or type 1 diabetic cases. Diabetes 2006; 55: 1525-1528.
[60] Mayans S, Lackovic K, Nyholm C, Lindgren P, Ruikka K, Eliasson M, Cilio CM and Holmberg D. CT60 genotype does not affect CTLA-4 isoform expression despite association to T1D and AITD in northern Sweden. BMC Med Genet 2007; 8:3.: 3.
[61] Park Y. Functional evaluation of the type 1 diabetes (T1D) susceptibility candidate genes. Diabetes Res Clin Pract 2007; 77 Suppl 1:S110-5. Epub;%2007 Apr;%19.: S110-S115.
[62] Bugawan TL, Mirel DB, Valdes AM, Panelo A, Pozzilli P and Erlich HA. Association and interaction of the IL4R, IL4, and IL13 loci with type 1 diabetes among Filipinos. Am J Hum Genet 2003; 72: 1505-1514.
[63] Maier LM, Chapman J, Howson JM, Clayton DG, Pask R, Strachan DP, McArdle WL, Twells RC and Todd JA. No evidence of association or interaction between the IL4RA, IL4, and IL13 genes in type 1 diabetes. Am J Hum Genet 2005; 76: 517-521.
[64] Bergholdt R, Taxvig C, Eising S, Nerup J and Pociot F. CBLB variants in type 1 diabetes and their genetic interaction with CTLA4. J Leukoc Biol 2005; 77: 579-585.
[65] Motzo C, Contu D, Cordell HJ, Lampis R, Congia M, Marrosu MG, Todd JA, Devoto M and Cucca F. Heterogeneity in the magnitude of the insulin gene effect on HLA risk in type 1 diabetes. Diabetes 2004; 53: 3286-3291.
[66] Hagopian WA, Lernmark A, Rewers MJ, Simell OG, She JX, Ziegler AG, Krischer JP and Akolkar B. TEDDY--The Environmental Determinants of Diabetes in the Young: an observational clinical trial. Ann N Y Acad Sci 2006; 1079:320-6.: 320-326.
[67] The TEDDY Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. Pediatr Diabetes 2007; 8: 286-298.
[68] Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, McDuffie RS, Jr., Hamman RF, Klingensmith G, Eisenbarth GS and Erlich HA. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). Diabetologia 1996; 39: 807-812.
[69] Huang HS, Peng JT, She JY, Zhang LP, Chao CC, Liu KH and She JX. HLA-encoded susceptibility to insulin-dependent diabetes mellitus is determined by DR and DQ genes as well as their linkage disequilibria in a Chinese population. Hum Immunol 1995; 44: 210-219.
[70] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. Molecular classification of cancer: class discovery and class prediction by gene

expression monitoring. Science 1999; 286: 531-537.
[71] Berchuck A, Iversen ES, Lancaster JM, Pittman J, Luo J, Lee P, Murphy S, Dressman HK, Febbo PG, West M, Nevins JR and Marks JR. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. Clin Cancer Res 2005; 11: 3686-3696.
[72] Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR and Huang AT. Gene expression predictors of breast cancer outcomes. Lancet 2003; 361: 1590-1596.
[73] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000; 403: 503-511.
[74] Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS and Via CS. Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. Clin Immunol 2002; 102: 283-290.
[75] Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, Shark KB, Grande WJ, Hughes KM, Kapur V, Gregersen PK and Behrens TW. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci 2003; 100: 2610-2615.
[76] Chabas D, Baranzini SE, Mitchell D, Bernard CC, Rittling SR, Denhardt DT, Sobel RA, Lock C, Karpuj M, Pedotti R, Heller R, Oksenberg JR and Steinman L. The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease. Science 2001; 294: 1731-1735.
[77] Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, Langer-Gould A, Strober S, Cannella B, Allard J, Klonowski P, Austin A, Lad N, Kaminski N, Galli SJ, Oksenberg JR, Raine CS, Heller R and Steinman L. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. Nat Med 2002; 8: 500-508.
[78] Eckenrode SE, Ruan Q, Yang P, Zheng W, McIndoe RA and She JX. Gene expression profiles define a key checkpoint for type 1 diabetes in NOD mice. Diabetes 2004; 53: 366-375.
[79] Collins CD, Purohit S, Podolsky RH, Zhao HS, Schatz D, Eckenrode SE, Yang P, Hopkins D, Muir A, Hoffman M, McIndoe RA, Rewers M and She JX. The application of genomic and proteomic technologies in predictive, preventive and personalized medicine. Vascul Pharmacol 2006; 45: 258-267.
[80] Jarvis JN, Dozmorov I, Jiang K, Frank MB, Szodoray P, Alex P and Centola M. Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis. Arthritis Res Ther 2004; 6: R15-R32.
[81] Winter WE, Harris N and Schatz D. Type 1 diabetes islet autoantibody markers. Diabetes Technol Ther 2002; 4: 817-839.
[82] Baekkeskov S, Aanstoot HJ, Christgau S, Reetz A, Solimena M, Cascalho M, Folli F, Richter-Olesen H, De Camilli P and Camilli PD. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. Nature 1990; 347: 151-156.
[83] Bottazzo GF, Florin-Christensen A and Doniach D. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. Lancet 1974; 2: 1279-1283.
[84] Lu J, Li Q, Xie H, Chen ZJ, Borovitskaya AE, Maclaren NK, Notkins AL and Lan MS. Identification of a second transmembrane protein tyrosine phosphatase, IA-2beta, as an autoantigen in insulin-dependent diabetes mellitus: precursor of the 37-kDa tryptic fragment. Proc Natl Acad Sci U S A 1996; 93: 2307-2311.
[85] Myers MA, Rabin DU and Rowley MJ. Pancreatic islet cell cytoplasmic antibody in diabetes is represented by antibodies to islet cell antigen 512 and glutamic acid decarboxylase. Diabetes 1995; 44: 1290-1295.
[86] Palmer JP, Asplin CM, Clemons P, Lyen K, Tatpati O, Raghu PK and Paquette TL. Insulin antibodies in insulin-dependent diabetics before insulin treatment. Science 1983; 222: 1337-1339.
[87] Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, Rewers M, Eisenbarth GS, Jensen J, Davidson HW and Hutton JC. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci U S A 2007; 104: 17040-17045.
[88] Wasserfall C, Atkinson M, Jodoin E, Schatz D, She JX, Henderson LO and Ellis T. Glutamic acid decarboxylase and IA-2 autoantibodies in type 1 diabetes: comparing sample substrates for autoantibody determinations. Pediatr Diabetes 2000; 1: 10-16.
[89] Eisenbarth GS, Gianani R, Yu L, Pietropaolo M, Verge CF, Chase HP, Redondo MJ, Colman P, Harrison L and Jackson R. Dual-parameter model for prediction of type I diabetes mellitus. Proc Assoc Am Physicians 1998; 110: 126-135.
[90] Ziegler R and Eisenbarth GS. Multiple target antigens in pre-type I diabetes: implications for prediction. Horm Res 1990; 33: 144-150.
[91] Rulli M and Simell O. Avidity of islet cell

autoantibodies in non-diabetic children and children with insulin-dependent diabetes. Autoimmunity 1999; 31: 187-193.

[92] Carmichael SK, Johnson SB, Baughcum A, North K, Hopkins D, Dukes MG, She JX and Schatz DA. Prospective assessment in newborns of diabetes autoimmunity (PANDA): maternal understanding of infant diabetes risk. Genet Med 2003; 5: 77-83.

[93] Notkins AL. Type 1 diabetes as a model for autoantibodies as predictors of autoimmune diseases. Autoimmun Rev 2004; 3 Suppl 1:S7-9.: S7-S9.

[94] Notkins AL and Lernmark A. Autoimmune type 1 diabetes: resolved and unresolved issues. J Clin Invest 2001; 108: 1247-1252.

[95] Bingley PJ, Bonifacio E and Mueller PW. Diabetes Antibody Standardization Program: first assay proficiency evaluation. Diabetes 2003; 52: 1128-1136.

[96] Achenbach P, Schlosser M, Williams AJ, Yu L, Mueller PW, Bingley PJ and Bonifacio E. Combined testing of antibody titer and affinity improves insulin autoantibody measurement: Diabetes Antibody Standardization Program. Clin Immunol 2007; 122: 85-90.

[97] Mueller PW, Bingley PJ, Bonifacio E, Steinberg KK and Sampson EJ. Predicting type 1 diabetes using autoantibodies: the latest results from the diabetes autoantibody standardization program. Diabetes Technol Ther 2002; 4: 397-400.

[98] Trudeau JD, Kelly-Smith C, Verchere CB, Elliott JF, Dutz JP, Finegood DT, Santamaria P and Tan R. Prediction of spontaneous autoimmune diabetes in NOD mice by quantification of autoreactive T cells in peripheral blood. J Clin Invest 2003; 111: 217-223.

[99] Eisenbarth GS and Kotzin BL. Enumerating autoreactive T cells in peripheral blood: a big step in diabetes prediction. J Clin Invest 2003; 111: 179-181.

[100] Schloot NC, Meierhoff G, Karlsson FM, Ott P, Putnam A, Lehmann P, Gottlieb P, Roep BO, Peakman M and Tree T. Comparison of cytokine ELISpot assay formats for the detection of islet antigen autoreactive T cells. Report of the third immunology of diabetes society T-cell workshop. J Autoimmun 2003; 21: 365-376.

[101] Roep BO, Atkinson MA, van Endert PM, Gottlieb PA, Wilson SB and Sachs JA. Autoreactive T cell responses in insulin-dependent (Type 1) diabetes mellitus. Report of the first international workshop for standardization of T cell assays. J Autoimmun 1999; 13: 267-282.

[102] Roep BO. Standardization of T-cell assays in Type I diabetes. Immunology of Diabetes Society T-cell Committee. Diabetologia 1999; 42: 636-637.

[103] Nagata M, Kotani R, Moriyama H, Yokono K, Roep BO and Peakman M. Detection of autoreactive T cells in type 1 diabetes using coded autoantigens and an immunoglobulin-free cytokine ELISPOT assay: report from the fourth immunology of diabetes society T cell workshop. Ann N Y Acad Sci 2004; 1037: 10-15.

[104] Corbett JA and McDaniel ML. Intraislet release of interleukin 1 inhibits beta cell function by inducing beta cell expression of inducible nitric oxide synthase. J Exp Med 1995; 181: 559-568.

[105] Reimers JI, Andersen HU, Mauricio D, Pociot F, Karlsen AE, Petersen JS, Mandrup-Poulsen T and Nerup J. Strain-dependent differences in sensitivity of rat beta-cells to interleukin 1 beta in vitro and in vivo: association with islet nitric oxide synthesis. Diabetes 1996; 45: 771-778.

[106] Pukel C, Baquerizo H and Rabinovitch A. Destruction of rat islet cell monolayers by cytokines. Synergistic interactions of interferon-gamma, tumor necrosis factor, lymphotoxin, and interleukin 1. Diabetes 1988; 37: 133-136.

[107] Rabinovitch A and Suarez-Pinzon WL. Role of cytokines in the pathogenesis of autoimmune diabetes mellitus. Rev Endocr Metab Disord 2003; 4: 291-299.

[108] Rabinovitch A and Suarez-Pinzon WL. Roles of cytokines in the pathogenesis and therapy of type 1 diabetes. Cell Biochem Biophys 2007; 48: 159-163.

[109] al-Nawawy A, Soliman T, el-Azzouni O, Abassy AA, Massound MN, Marzouk S, IBrahim F and Helal L. Interleuking 1-beta, tumor necrosis factor-alpha, insulin secretion and oral glucose tolerance in non-diabetic siblings of children with IDDM. Indian J Pediatr 1998; 65: 455-460.

[110] Aribi M, Moulessehoul S, Kendouci-Tani M, Benabadji AB, Hichami A and Khan NA. Relationship between interleukin-1beta and lipids in type 1 diabetic patients. Med Sci Monit 2007; 13: CR372-CR378.

[111] Netea MG, Hancu N, Blok WL, Grigorescu-Sido P, Popa L, Popa V and van der Meer JW. Interleukin 1 beta, tumour necrosis factor-alpha and interleukin 1 receptor antagonist in newly diagnosed insulin-dependent diabetes mellitus: comparison to long-standing diabetes and healthy individuals.

[112] Lechleitner M, Koch T, Herold M, Dzien A and Hoppichler F. Tumour necrosis factor-alpha plasma level in patients with type 1 diabetes mellitus and its association with glycaemic control and cardiovascular risk factors. J Intern Med 2000; 248: 67-76.

[113] Reddy S, Bai Y, Robinson E and Ross J. Immunolocalization of monocyte chemoattractant protein-1 in islets of NOD mice during cyclophosphamide administration. Ann N Y Acad Sci 2006; 1079:103-8.: 103-108.

[114] Lee I, Wang L, Wells AD, Ye Q, Han R, Dorf ME,

Kuziel WA, Rollins BJ, Chen L and Hancock WW. Blocking the monocyte chemoattractant protein-1/CCR2 chemokine pathway induces permanent survival of islet allografts through a programmed death-1 ligand-1-dependent mechanism. J Immunol 2003; 171: 6929-6935.
[115]Chipitsyna G, Gong Q, Gray CF, Haroon Y, Kamer E and Arafat HA. Induction of monocyte chemoattractant protein-1 expression by angiotensin II in the pancreatic islets and beta-cells. Endocrinology 2007; 148: 2198-2208.
[116]Christen U and Von Herrath MG. IP-10 and type 1 diabetes: a question of time and location. Autoimmunity 2004; 37: 273-282.
[117]Shimada A, Morimoto J, Kodama K, Suzuki R, Oikawa Y, Funae O, Kasuga A, Saruta T and Narumi S. Elevated serum IP-10 levels observed in type 1 diabetes. Diabetes Care 2001; 24: 510-515.
[118]Rotondi M, Falorni A, De BA, Laureti S, Ferruzzi P, Romagnani P, Buonamano A, Lazzeri E, Crescioli C, Mannelli M, Santeusanio F, Bellastella A and Serio M. Elevated serum interferon-gamma-inducible chemokine-10/CXC chemokine ligand-10 in autoimmune primary adrenal insufficiency and in vitro expression in human adrenal cells primary cultures after stimulation with proinflammatory cytokines. J Clin Endocrinol Metab 2005; 90: 2357-2363.
[119]Glowinska B, Urban M, Peczynska J and Florys B. Soluble adhesion molecules (sICAM-1, sVCAM-1) and selectins (sE selectin, sP selectin, sL selectin) levels in children and adolescents with obesity, hypertension, and diabetes. Metabolism 2005; 54: 1020-1026.
[120]Toivonen A, Kulmala P, Rahko J, Ilonen J and Knip M. Soluble adhesion molecules in Finnish schoolchildren with signs of preclinical type 1 diabetes. Diabetes Metab Res Rev 2004; 20: 48-54.
[121]Toivonen AM, Kimpimaki T, Kupila A, Korhonen S, Hyoty H, Virtanen SM, Ilonen J, Simell O and Knip M. Soluble adhesion molecules in young children with signs of beta-cell autoimmunity--prospective follow-up from birth. Diabetes Metab Res Rev 2006; 22: 176-183.
[122]Salmi M, Stolen C, Jousilahti P, Yegutkin GG, Tapanainen P, Janatuinen T, Knip M, Jalkanen S and Salomaa V. Insulin-regulated increase of soluble vascular adhesion protein-1 in diabetes. Am J Pathol 2002; 161: 2255-2262.
[123]Mohamed-Ali V, Armstrong L, Clarke D, Bolton CH and Pinkney JH. Evidence for the regulation of levels of plasma adhesion molecules by proinflammatory cytokines and their soluble receptors in type 1 diabetes. J Intern Med 2001; 250: 415-421.
[124]Soedamah-Muthu SS, Chaturvedi N, Schalkwijk CG, Stehouwer CD, Ebeling P and Fuller JH. Soluble vascular cell adhesion molecule-1 and soluble E-selectin are associated with micro- and macrovascular complications in Type 1 diabetic patients. J Diabetes Complications 2006; 20: 188-195.
[125]Dogruel N, Kirel B, Akgun Y and Us T. Serum soluble endothelial-cell specific adhesion molecules in children with insulin-dependent diabetes mellitus. J Pediatr Endocrinol Metab 2001; 14: 287-293.
[126]Yan W and Chen SS. Mass spectrometry-based quantitative proteomic profiling. Brief Funct Genomic Proteomic 2005; 4: 27-38.
[127]von Eggeling F, Junker K, Fiedle W, Wollscheid V, Durst M, Claussen U and Ernst G. Mass spectrometry meets chip technology: a new proteomic tool in cancer research? Electrophoresis 2001; 22: 2898-2902.
[128]Petricoin EF, III, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, Velassco A, Trucco C, Wiegand L, Wood K, Simone CB, Levine PJ, Linehan WM, Emmert-Buck MR, Steinberg SM, Kohn EC and Liotta LA. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst 2002; 94: 1576-1578.
[129]Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC and Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002; 359: 572-577.
[130]Purohit S, Podolsky R, Schatz D, Muir A, Hopkins D, Huang YH and She JX. Assessing the utility of SELDI-TOF and model averaging for serum proteomic biomarker discovery. Proteomics 2006; 6: 6405-6415.
[131]Anderson NL and Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002; 1: 845-867.
[132]Yang Q, Khoury MJ, Botto L, Friedman JM and Flanders WD. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. Am J Hum Genet 2003; 72: 636-649.
[133]Hirji KF, Mehta CR and Patel NR. Computing distributions for exact logistic regression. J Am Stat Assoc 1987; 82: 1110-1117.
[134]Huang E, West M and Nevins JR. Gene expression profiling for prediction of clinical characteristics of breast cancer. Recent Prog Horm Res 2003; 58: 55-73.
[135]Breiman L: Classification and regression trees. Belmont, CA Wadsworth 2004.
[136]Breiman L. Bagging predictors. Machine Learning 1996; 24: 123-140.
[137]Breiman L. Arcing classifiers. Ann Stat 2004; 26: 801-824.
[138]Dudoit S, Fridlyand J and Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association 2002; 97: 77-87.
[139]Breiman L. Random Forests - Random

Features. Technical Report 1999; 567.
[140]Kleinberg EM. Stochastic Discrimination. Ann Math Artificial, Intell 1990; 1: 207-239.
[141]Kleinberg EM. An Overtraining-Resistant Stochastic Modeling Method for Pattern Recognition. Ann Stat 1996; 24: 2319-2349.