Novel Biomarkers in Type 1 Diabetes

Yulan Jin and Jin-Xiong She

Center for Biotechnology and Genomic Medicine and Department of Pathology, Medical College of Georgia, Georgia Regents University, Augusta, GA, USA. Address correspondence to: Jin-Xiong She, Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Georgia Regents University, 1120 15th Street, Augusta, GA 30912, USA, e-mail: jshe@gru.edu.

Manuscript submitted December 16, 2012; resubmitted December 21, 2012; accepted January 8, 2013

Abstract

Biomarkers are useful tools for research into type 1 diabetes (T1D) for a number of purposes, including elucidation of disease pathogenesis, risk prediction, and therapeutic monitoring. Susceptibility genes and islet autoantibodies are currently the most useful biomarkers for T1D risk prediction. However, these markers do not fully meet the needs of scientists and physicians for several reasons. First, improvement of the specificity and sensitivity is still desirable to achieve better positive predictive values. Second, autoantibodies appear relatively late in the disease process, thus limiting their value in early disease prediction. Third, the currently available biomarkers are not useful for assessing therapeutic outcomes because some are not involved in the disease process (autoantibodies) and others do not change during disease progression (susceptibility genes). Therefore, considerable effort has been devoted to the discovery of novel T1D biomarkers in the last three decades. The advent of high-throughput technologies for genetic, transcriptomic, and proteomic studies has allowed genome-wide examinations of genetic polymorphisms, global gene changes, and protein expression changes in T1D patients and prediabetic subjects. These large-scale studies resulted in the discovery of a large number of susceptibility genes and changes in gene and protein expression. While these studies have provided a number of novel biomarker candidates, their clinical benefits remain to be evaluated in prospective studies, and no new “star biomarker” has been identified until now. Previous studies suggest that significant improvements in study design and analytical methodologies have to be made to identify clinically relevant biomarkers. In this review, we discuss progress, opportunities, challenges, and future directions in the development of T1D biomarkers, mainly by focusing on the genetic, transcriptomic, and proteomic aspects.

Keywords: type 1 diabetes · biomarker · prediction · pathogenesis · risk · prevention · antibody · INS · CTLA4 · PTPN22 · IL2RA

1. Introduction

Type 1 diabetes (T1D) is an autoimmune disease primarily starting in childhood. It results from the destruction of insulin-producing β-cells of the pancreas [1]. Although exogenous insulin can help to maintain the level of blood glucose, there is no cure for this disease, and long-term complications can cause serious disability and shortened lifespan. Furthermore, the pancreatic islet β-cell mass has almost completely been destroyed by the time of disease onset, making prediction and prevention a high priority. Also, the increased incidence of T1D, with an average rate of 3% per year [2-5], commands urgency in developing novel prediction and prevention strategies. Biomarkers play essential roles for both identification of high-risk populations and development of prevention strategies.

The etiology of T1D is caused by poorly understood interactions between genetic and environmental factors. The development of the disease involves a cascade of molecular, cellular, and metabolomic impairments which may provide a vari-
Novel Biomarkers in T1D

2. Prediction of T1D risk using genetic markers

Genetic factors play an important role in T1D pathogenesis [7]. The strong genetic contribution to T1D is illustrated by the increased risk in siblings (5% by age 20) versus the general population (0.3%) [8] and the high concordance rate in identical twins (up to 65% by age 60) [9]. The search for T1D-associated genes started in the 1970s using primarily two approaches:

1. Linkage studies (using pairs of affected relatives, typically siblings)
2. Association studies (using either case-control or family-based designs)

Although linkage studies have revealed the major contribution of major histocompatibility complex (MHC) to T1D, most non-MHC risk loci (such as insulin (INS), cytotoxic T lymphocyte antigen 4 (CTLA4), protein tyrosine phosphatase, non-receptor type 22 (PTPN22), and interleukin 2 receptor alpha (IL2RA)) were identified using candidate-gene association studies. Recently, association studies have evolved from candidate genes to...
The natural history of T1D has four main stages: 1. genetic predisposition, 2. prediabetes or autoimmunity, 3. clinical disease onset, and 4. development of diabetic complications. Primary prevention may be applied to the transition from genetic predisposition to autoimmunity. Primary prevention strategies should be highly effective, low-cost and no side effect and can be applied to large numbers of individuals that have increased genetic risk. Secondary prevention can be applied to prediabetic subjects with islet antibodies and the strategies of secondary prevention should be effective and moderate-cost (for example: antigen-based immunization and immunosuppressive drugs). Tertiary prevention is mainly applied to T1D patients with high risk for diabetic complications and the strategies of this step should be effective and reasonable-cost.

Since 2007, highly dense panels of single nucleotide polymorphisms (SNPs) (>300,000 SNPs) distributed across the human genome have been used in GWAS [32-35]. These high-density GWAS confirmed the previously identified loci such as INS, PTPN22, CTLA4, and IL2RA, but more importantly, they provided evidence for a number of novel loci. Surprisingly, the effect sizes estimated for these novel loci were much lower than those for HLA genes, INS and PTPN22, despite the strong statistical power of the GWAS. In view of the large cohort sizes and genome-wide SNP coverage in these studies, it is extremely unlikely that additional loci with large effect can be identified using similar approaches. Furthermore, few of these GWAS loci have yet been mapped to a specific variant or even to a specific gene. Summary results of the GWAS and meta-analysis are available through http://t1dbase.org/.

The advantages of genetic variants as biomarkers are apparent. As germ-line factors, genetic risk variants can serve as a potential predictive tool at a very early stage, even in uterus. Additionally, these genetic factors are relatively easy, inexpensive, and noninvasive to measure. However, there are many challenges in translating these genetic findings into clinical applications. First, genotyping of HLA loci, combined with family history and autoantibody presence, is a current approach for T1D risk prediction with high speci-
ficity for the high-risk categories of individuals. However, it has low overall sensitivity in the general population, as most of the T1D cases occur in populations with low or moderate risk. Second, most non-HLA loci have only modest or low individual effects on risk, which hardly increase the predictive value, even if using multiple genetic markers. Third, despite the increasing number of potential target genes, a considerable lack of understanding remains regarding the roles of these genes in the pathogenesis of T1D, and the most efficient ways of application for disease prediction and prevention.

Future genetic studies need to be designed to overcome these challenges. One strategy is to perform follow-up studies for GWAS to better understand the involvement of the candidate loci. The first step to determine the responsible gene(s) and allele(s) is a fine mapping of the regions. Once a reliable risk variant has been identified, the next step is to determine the immediate effects of the gene on gene and protein expression, and the assignable phenotypes in patients with T1D. The second strategy is to perform prospective cohort studies to further validate the GWAS findings and to identify novel genetic factors with larger effect size. All susceptibility genes identified so far are derived from cross-sectional studies that include samples from heterogeneous populations. We anticipate that prospective cohorts are much more suitable for gene mapping studies in complex diseases.

3. Transcriptomic biomarkers for T1D

The dynamic state of the transcriptome during prediabetes, disease progression, and clinical treatment can provide potential biomarkers for disease risk prediction, disease subtype classification, and therapeutic monitoring. Microarray technology and bioinformatics allow the analysis of the whole transcriptome (gene expression profile) in a single experiment. This approach has been successfully applied in many studies, especially on cancer. For example, gene expression profiling has been used to distinguish acute myeloid and acute lymphoblastic leukemia cells [36], to predict outcomes in breast and ovarian cancers [37, 38]. It has also been used to classify subtypes of diffuse large B-cell lymphomas for prognostic implications [39, 40].

A number of studies have attempted to discover changes of gene expression profiles during T1D development [41-48]. Our group identified over 100 genes upregulated in peripheral blood mononuclear cells (PBMCs) of T1D subjects. Most of these genes are also upregulated in prediabetic subjects, suggesting that they may be useful predictive markers [46]. Many of the differentially expressed genes are involved in important immunological functions, including antigen processing and presentation, cytotoxicity and apoptosis (e.g. GZMB, GNLY), and immune regulation (e.g. TGFβ1, SEL1L). It was found that several proinflammatory mediators and markers (e.g. S100A8/9, NFκB) are upregulated in diabetic and prediabetic subjects. Kaizer and colleagues found overexpression of IL-1-regulated genes as well as chemotaxis and signaling genes in T1D PBMCs [48]. Transcripts corresponding to genes encoding proteins involved in apoptosis and the cell cycle were downregulated in some studies [48], but upregulated in others [46]. Recently, the expression profile for whole blood has suggested an increase in INF-responsive genes at the prediabetic stage [49], a pathway also altered in other autoimmune diseases such as systemic lupus erythematosus (SLE) [50] and Sjogren’s Syndrome [51]. Gene expression patterns were further analyzed in subgroups of patients and controls. Significant differences in gene sets were observed between healthy first-degree relatives of T1D and healthy controls [52], long-term T1D patients and new-onset T1D [45], and juvenile-onset and adult-onset T1D [41]. In addition to the analysis of fresh PBMCs, expression changes inducible in PBMCs cultured with sera from T1D patients have identified soluble factors associated with T1D [53].

In the last decade, much effort has been devoted to discover gene expression patterns in human T1D. However, these studies had the following serious limitations, which affect the validity of their results:

1. Most studies sampled PBMCs rather than pancreatic islets due to the difficulties in obtaining pancreatic samples from human subjects. Transcriptional regulation in the periphery could not be an accurate reflection of autoimmune response in the islets, given the low percentage of islet-reactive lymphocytes in peripheral blood. Also, changes in gene expression that are confined to a particular cell type (e.g. regulatory T cells, dendritic cells and monocyte) may be difficult to detect in PBMCs.

2. The results of most published studies are largely inconclusive and sometimes contradictory. This may be explained by the
following reasons: (i) Microarray-based gene expression profiling is a powerful discovery platform, but the results need to be validated by an alternative technique such as real-time polymerase chain reaction (RT-PCR). Unfortunately, few of the previous microarray studies on T1D have been further confirmed by a validation study. (ii) Most previous studies had very small sample sizes (less than 100 subjects in each group) which are not adequate for the human population given the large expression variations among individual subjects.

3. Most gene expression profile data were derived from cross-sectional studies, which is a good approach for biomarker discovery, but not good enough for biomarker validation. There were a few studies using longitudinal samples to characterize the gene expression signatures during T1D progression; however, the small sample size limited their study power [44]. Therefore, prospective studies with large cohorts need to be designed for future studies.

Transcriptomic biomarkers for T1D are still not ready for clinical application even though this approach holds great promise. One urgent task is the validation of the previous findings from microarray data using an alternative method such as RT-PCR. To obtain more reliable and reproducible results, the experimental design needs to be improved in the following ways:

1. Use of large sample size, at least a thousand in each group to get consistent results based on our validation data (Table 1), which showed inconsistent data for some genes between two independent medium-sized sample sets.

2. When expression data are generated on multiple 384-well plates for thousands of samples, plate-to-plate variation should be recognized and normalized to ensure consistency across plates.

3. Given the huge intra-group individual variation, only looking at fold changes of gene expression among studied groups, as done in most previous studies, is not

# Table 1. Summary of gene expression validation data

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Microarray (n = 59/35)</th>
<th>Set 1 (n = 155/194)</th>
<th>Set 2 (n = 192/190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBLB</td>
<td>Casitas B-lineage lymphoma b</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>CD74</td>
<td>CD74 antigen</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>PSM B3</td>
<td>Proteasome subunit beta, type 3</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>MF NG</td>
<td>Manic fringe</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>TM 9SF4</td>
<td>Transmembrane 9 superfamily protein member 4</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>PRDX1</td>
<td>Peroxiredoxin 1</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>RSA</td>
<td>Ribosomal protein SA</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>TSPAN 14</td>
<td>Transmembrane 4 superfamily protein member 14</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>M NDA</td>
<td>Myeloid cell nuclear differentiation antigen</td>
<td>U</td>
<td>U</td>
<td>D</td>
</tr>
<tr>
<td>TXNIP</td>
<td>Thioredoxin interacting protein</td>
<td>U</td>
<td>U</td>
<td>D</td>
</tr>
<tr>
<td>RNF10</td>
<td>Ring finger protein 10</td>
<td>U</td>
<td>U</td>
<td>D</td>
</tr>
<tr>
<td>SELL</td>
<td>Selectin L (lymphocyte adhesion molecule 1)</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>TSEN 34</td>
<td>Leukocyte receptor cluster(LRC) member 5</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>P TEN</td>
<td>Phosphatase and tensin homolog</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>RNF 31</td>
<td>Ring finger protein 31</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>CSF 2RB</td>
<td>Colony stimulating factor 2 receptor, beta 2</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>SM A 7</td>
<td>Mothers against decapentaplegic homolog 7</td>
<td>U</td>
<td>U</td>
<td>ns</td>
</tr>
<tr>
<td>INF G</td>
<td>Interferon gamma</td>
<td>U</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>GNLY</td>
<td>Granulysin</td>
<td>U</td>
<td>ns</td>
<td>D</td>
</tr>
<tr>
<td>S100 A 9</td>
<td>S100 calcium binding protein A9</td>
<td>U</td>
<td>ns</td>
<td>D</td>
</tr>
</tbody>
</table>

Legend: U - upregulated in T1D; D - downregulated in T1D; ns - not significant.
enough to identify T1D associated genes. Other statistical approaches, such as conditional logistic regression, should also be used to estimate the relative risk.

4. Selection of appropriate reference genes for normalization of quantitative Real-Time PCR has a major impact on data quality [54-57]. Most of the previous studies have used only a single reference gene for normalization. To avoid biased results, gene transcription studies using RT-PCR should begin with the selection of an appropriate set of reference genes.

Rapid progress is being made in the development of novel technologies for transcriptomics. Notably, the introduction of high-throughput next-generation DNA sequencing (NGS) technologies revolutionized transcriptomics by performing RNA analysis through cDNA sequencing on a massive scale [58]. This development eliminated several challenges associated with microarray technologies, and has provided a better knowledge of both quantitative and qualitative aspects of transcriptomics. This novel technology allows a more comprehensive understanding of transcription initiation sites, the cataloguing of sense and antisense transcripts, previously unknown coding and non-coding RNA species, particularly small RNAs (e.g. micro RNA), improved detection of alternative splicing events, and improved detection of gene fusion transcripts [59]. We believe that this technology will become increasingly important in T1D research, and provide unparalleled opportunities for biomarker discovery.

4. Proteomic biomarkers for T1D

Protein levels execute the aberrant genetic and genomic changes, and therefore are more directly correlated with cellular function and health status. Thus, they have a greater potential to be used as biomarkers for disease prediction or therapeutic monitoring. Today, the most useful proteomic biomarkers for T1D are “islet autoantibodies” present in serum, which are strong predictors of the later development of T1D. Since the 1970s, a series of islet autoantibodies (over 20) involved in T1D has been discovered [60-79]. Autoantibody assays have constantly been improved. The four major autoantibodies of clinical and research interest are islet cell cytoplasmic autoantibodies (ICA), glutamic acid decarboxylase autoantibodies (GADA), insulina 2-associated autoantibodies (IA-2A), and insulin autoantibodies (IAA). ZnT8A, a newly recognized ZnT8 islet autoantibody, may further improve the value of islet autoantibody testing. Other identified autoantibodies are either difficult to measure and/or are not sufficiently sensitive or specific to enable their use as T1D markers [80]. Although the major islet autoantibodies are not considered to be involved in the pathogenesis of T1D, they are the hallmark of autoimmune response to autoantigens, and they are critical for the design of clinical trials for T1D prevention [6].

Unquestionably, the risk of developing T1D will rise with the increase in number of islet autoantibodies present. The risk for T1D in individuals without any autoantibody is only 0.5%, and rise to approximately 3% in individuals with one autoantibody. In subjects with two autoantibodies the risk keeps rising to 16%, and jumps to 40% and 50% in subjects with three and four autoantibodies, respectively [81]. IAA is less predictive of T1D than other autoantibodies. Higher ICA titers and higher concentrations of GADA were more powerful predictors of T1D than lower titers and lower concentrations. Despite the usage of the autoantibodies in T1D prediction, they have several serious limitations:

1. The appearance of islet autoantibodies marks a relatively late stage of the autoimmune process, and therefore is not suitable for early disease intervention. A recent study suggested that some extrapancreatic autoantibodies were present prior to detection of islet autoantibodies; however, their potential for biomarkers need to be further assessed [82].

2. Only a subset of the autoantibody-positive subjects will progress to clinical diabetes. Therefore, it would be desirable to have biomarkers that allow the distinction of the progressors versus non-progressors.

3. Autoantibodies are not useful as biomarkers for therapeutic outcomes.

Other than autoantibodies, several immune molecules, like cytokines and chemokines, have been widely studied for their potential roles in T1D development and for the possibility as T1D biomarkers. The current view is that T1D arises from T cell-mediated islet cell destruction initiated by an imbalance in Th1 and Th2 cells [83]. T1D is believed to be mediated by Th1 cytokines. Several studies have found higher serum levels of Th1 cy-
tokines in diabetic patients and their first-degree relatives compared to healthy controls [84-88]. A similar pattern of Th1/2 cytokine profile between newly diagnosed T1D patients and their healthy siblings was further confirmed in a recent study with a relatively large sample size (500 subjects in each group) [89]. The Th1 cytokines IL-1β, TNF-α, and IFN-γ have been shown to be cytotoxic to β-cells by inducing nitric oxide (NO) production [90]. Some studies suggested a significantly higher IL-1 production but a lower IL-1Ra (IL1 receptor antagonist) in newly diagnosed T1D patients compared with chronic T1D patients [91, 92].

MCP-1 and IP-10 are the two best studied chemokines in T1D being chemo-attractors for monocytes and activated Th1 and NK cells specifically. Studies with animal models have demonstrated that high levels of MCP-1 and IP-10 are released by islets cells during autoimmune attack [93-96]. Serum levels of these two cytokines have been measured in T1D patients and healthy controls in several studies. However, all these studies had extremely small sample sizes, and the conclusions from these studies were inconclusive and inconsistent. A study by our group has measured serum MCP-1 in a large cohort (with 2724 T1D patients and 2654 controls) [97]. Interestingly, serum MCP-1 levels were significantly higher in patients with multiple complications than in patients without any complications. They were also higher in controls than T1D patients, which suggested MCP-1 may have a dual role in T1D and its complications. Currently, assay sensitivity is a major limitation for the accurate measurement of many cytokines and chemokines with low concentration, such as GM-CSF, INF-γ, and interleukins [98].

Quantitative analysis of global protein levels is important for the systematic understanding of the molecular changes associated with disease progression, and may provide insight into disease pathogenesis and management. Recent development of mass spectrometry (MS)-based technology has provided useful platforms for the study of quantitative changes in protein components [99, 100]. Several methods are widely used in proteomic analysis, including two-dimensional (2D) gel electrophoresis followed by MS analysis, MS signal intensity-based quantification, stable isotope labeling-based quantification, and intact protein-based quantification [101-103]. Because of the difficulties in obtaining pathological tissues from T1D patients due to ethical and practical concerns, serum is an excellent alternative resource for biomarker discovery. It is rich in biological information and easy available. However, comprehensive analysis of the serum proteome is a challenging task due to its extraordinary complexity and high dynamic range in concentration. The complexity of serum proteome results from its charge, molecular mass, and hydrophobicity, as well as its expression level and post-translational modifications. Only abundant proteins are analyzable by currently available methods due to their high dynamic range in protein concentration. Unfortunately, low abundance proteins that are promising biomarkers are difficult to detect and quantify. This difficulty is particularly true for serum because more than 99% of the proteins consist of serum albumin and globulins.

Therefore, sample pretreatment is required to enrich the low- to medium-abundance proteins, called protein normalization. Our group compared different approaches for low-abundance enrichment, and showed that random hexapeptide library beads have distinct advantages over the traditional immune-depletion methods due to their higher efficiency, higher binding capacity, and lower costs [104]. We also evaluated in-depth mining of serum/plasma proteome using different separation techniques. Our data suggested that shotgun proteomics—multidimensional separation of digested peptides followed by mass spectrometry analysis—is highly efficient, and therefore should be the desirable method for protein biomarker discovery [105, 106]. Using these preferred strategies, our group, as one of the pioneers, systematically discovered and validated serum proteomic changes in T1D patients [106]. We found that two well-known inflammation mediators, serum amyloid protein A (SAA) and C-reactive protein (CRP), as well as adiponectin and insulin-like growth factor binding protein 2 have significantly higher serum levels in T1D patients. Whereas, there are lower serum levels for two other proteins: transforming growth factor beta induced (TGFβ1) and myeloperoxidase. In particular, the subjects in the top quartile for expression of these markers had the highest risk of T1D (relative risk is up to 10).

In addition to serum, urine represents another excellent specimen for proteome analysis due to its easy availability and higher stability than blood [107]. Maahs and colleagues discovered and further validated urinary proteomic biomarkers for diabetes in general, and for specific type of diabetes [108]. The difference of urinary collagen fragments between T1D and T2D suggested that different mechanisms of extracellular matrix remodeling exist in these two types of diabetes.
In summary, despite considerable progress in proteomic biomarker discovery for T1D, no new “star protein biomarkers” of equal importance to islet autoantibodies have been identified to date, due to both technical and biological issues, as aforementioned. Proteomic technology needs to be improved in several areas, including quantification and identification to low-abundance proteins, assessment of protein distribution among cells and subcellular compartments, and assessment of post-translational modification. In addition to the need for technical improvement, several issues need to be considered for better study design. One is the appropriate selection of biological specimens for each purpose. For example, serum probably is the most suitable specimen for T1D biomarker discovery due to its high richness of proteins, while urinary samples may be more suitable for T1D complication studies, especially for diabetic nephropathy. Furthermore, it should be recognized that no single analytical technique is suited to address the proteomic complexities. For example, the MS platform is suitable for the discovery phase, but not for validation. Finally, reasonable sample size, in both cross-sectional and longitudinal studies, is a key factor for the identification and validation of reliable T1D biomarkers.

5. Conclusions

The long asymptomatic period of T1D provides many opportunities for disease prevention and intervention. Genetic susceptibility and islet autoantibodies are still the most useful biomarkers for T1D risk prediction. However, these currently available markers do not fully meet the need for T1D prediction and prevention due to their low predictive value and relative late appearance.

Increasing efforts have been devoted to novel T1D biomarker discovery in the last three decades. The most noteworthy advances are developments of high throughput “omic” technologies which offer great opportunities for biomarker discovery. Unfortunately, so far “star biomarkers” with the potential to better predict T1D risk or evaluate therapeutic outcomes have not been discovered because of both biological and technical challenges. To overcome these issues, improvements in technologies and study design need to be made in future studies.

Acknowledgments: This work is partly supported by grants from the National Institutes of Health (4R33HD050196, 4R33DK069878, and 2RO1HD37800) and JDRF (1-2004-661) given to JXS.

Disclosure: The authors report no conflict of interests.

References

18. Bottini N, Musumeci I, Alonso A, Rahmouni S,


Novel Biomarkers in T1D


103. Asara JM, Christofk HR, Freimark LM, Cantley LC. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. Proteomics 2008. 8(5):994-999.


