

Yulan Jin,<sup>1,2,3</sup> Ashok Sharma,<sup>2,3</sup> Shan Bai,<sup>2</sup> Colleen Davis,<sup>2</sup> Haitao Liu,<sup>1,2</sup> Diane Hopkins,<sup>2</sup> Kathy Barriga,<sup>4</sup> Marian Rewers,<sup>4</sup> and Jin-Xiong She<sup>1,2,3</sup>



# Risk of Type 1 Diabetes Progression in Islet Autoantibody-Positive Children Can Be Further Stratified Using Expression Patterns of Multiple Genes Implicated in Peripheral Blood Lymphocyte Activation and Function

*Diabetes* 2014;63:2506–2515 | DOI: 10.2337/db13-1716

There is tremendous scientific and clinical value to further improving the predictive power of autoantibodies because autoantibody-positive (AbP) children have heterogeneous rates of progression to clinical diabetes. This study explored the potential of gene expression profiles as biomarkers for risk stratification among 104 AbP subjects from the Diabetes Autoimmunity Study in the Young (DAISY) using a discovery data set based on microarray and a validation data set based on real-time RT-PCR. The microarray data identified 454 candidate genes with expression levels associated with various type 1 diabetes (T1D) progression rates. RT-PCR analyses of the top-27 candidate genes confirmed 5 genes (*BACH2*, *IGLL3*, *EIF3A*, *CDC20*, and *TXNDC5*) associated with differential progression and implicated in lymphocyte activation and function. Multivariate analyses of these five genes in the discovery and validation data sets identified and confirmed four multigene models (BI, ICE, BICE, and BITE, with each letter representing a gene) that consistently stratify high- and low-risk subsets of AbP subjects with hazard ratios >6 ( $P < 0.01$ ).

The results suggest that these genes may be involved in T1D pathogenesis and potentially serve as excellent gene expression biomarkers to predict the risk of progression to clinical diabetes for AbP subjects.

Type 1 diabetes (T1D) is a chronic autoimmune disease resulting from the targeted destruction of insulin-secreting pancreatic islet  $\beta$ -cells. Islet autoantibodies, markers of active islet autoimmunity, can be detected years, and even decades, before the appearance of clinical symptoms (1). The long asymptomatic period between the appearance of islet autoantibodies and disease onset provides a window of opportunity for T1D prevention in subjects who are autoantibody positive (AbP). At least one of four major islet autoantibodies (insulin autoantibody, GAD antibody, IA-2 antibody, and zinc transporter 8 antibody) is detected in >90% of patients with newly diagnosed T1D (2). These autoantibodies have become the gold standard for identifying at-risk subjects from first-degree relatives (FDRs) of T1D patients as well as the general population

<sup>1</sup>Sino-American Institute of Translational Medicine, School of Pharmaceutical Sciences, Nanjing University of Technology, Nanjing, China

<sup>2</sup>Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Georgia Regents University, Augusta, GA

<sup>3</sup>Department of Pathology, Medical College of Georgia, Georgia Regents University, Augusta, GA

<sup>4</sup>Barbara Davis Center for Childhood Diabetes, Aurora, CO

Corresponding authors: Jin-Xiong She, jshe@gru.edu, and Marian Rewers, marian.rewers@ucdenver.edu.

Received 8 November 2013 and accepted 27 February 2014.

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db13-1716/-/DC1>.

© 2014 by the American Diabetes Association. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

(3). Further improvement of risk prediction using autoantibodies has clear scientific and clinical value. Subjects with multiple islet autoantibodies have a high projected risk within 10 years (69.7% [95% CI, 65.1–74.3%]), whereas the presence of a single autoantibody shows a low risk (14.5% [95% CI, 10.3–18.7%]) (4). Furthermore, AbP subjects have variable progression to T1D, with a prediabetes period ranging from 0 to 20 years. Given the variable length of time, further stratifying these individuals for more accurate prediction to clinical disease would be advantageous. Although age at seroconversion and titer of autoantibodies can further improve risk prediction (4,5), additional biomarkers are still needed. Considerable efforts have been devoted to the development of genetic and metabolic biomarkers based on AbP prospective cohorts or T1D prevention trials. HLA (4,6–8), non-HLA (9–12) genetic markers, and metabolic risk scores (13–15) have shown certain levels of improvement for stratifying the risk of AbP subjects. However, the practical potential of these markers is limited by either their intrinsic deficiencies or low predictive values.

Gene expression profiles are expected to dynamically change during disease progression and treatment. Therefore, gene expression patterns may serve as potential biomarkers for risk stratification and therapeutic monitoring. Several studies have examined gene expression changes related to T1D and identified a large number of genes that may differ in expression levels among healthy control subjects, AbP subjects, and T1D patients (16–22). However, these studies have been limited by their cross-sectional design and, hence, hardly suggest biomarker potential. The present study identified five genes that in combination can serve as biomarkers to stratify progression risk in AbP subjects. Our strategy first used microarray data to discover gene expression changes associated with differential progression from AbP to T1D and then validated the top-27 genes using quantitative RT-PCR data from independent AbP subjects from the Diabetes Autoimmunity Study in the Young (DAISY) cohort.

## RESEARCH DESIGN AND METHODS

### Human Subjects and Samples

A total of 104 AbP subjects who were consecutively seen in DAISY and followed until February 2012 were included in the analyses. AbP status was identified based on the presence of at least one of the following three autoantibodies: insulin autoantibody, GAD antibody, and IA-2 antibody. By the cutoff date, diabetes developed in 39 of the 104 AbP subjects, with a median follow-up time (from first AbP) of 5.64 years. The median follow-up time of the 65 nonprogressors was 8.9 years. Diabetes was diagnosed according to American Diabetes Association criteria. Demographic information on the distribution of age and sex, age at the appearance of first antibody, FDR status, genetic risk [classified by HLA genotype (23)], number of antibodies, and follow-up time (after first AbP) is summarized in Table 1. The 104 subjects were split into two

study phases: discovery (microarray) and validation (real-time RT-PCR). Thirty-six subjects were selected for the discovery phase, with progressors ( $n = 21$ ) and non-progressors ( $n = 15$ ) matched for age, sex, age at first AbP, FDR status, genetic risk, and number of autoantibodies (most with two or more AbP). The remaining 68 subjects (18 progressors and 50 nonprogressors) were included in the validation phase. Compared with nonprogressors, the progressors in the validation data set were younger, had an earlier age of first antibody appearance, had a higher frequency of multiple autoantibodies, and had more frequently FDRs with diabetes. Demographic information for the subjects in the two data sets is summarized in Table 1. A dot plot of the distribution of age of first AbP, age of sample collection, and age of T1D/last visit for each subject is shown in Supplementary Fig. 1. For most subjects, gene expression was analyzed at a relatively early time point following first antibody detection, with a mean interval time of  $2.01 \pm 2.17$  years.

### RNA Samples

Peripheral blood (2.5 mL) was immediately preserved in PAXgene RNA tubes. After 2 h at room temperature, the tubes were stored at  $-80^{\circ}\text{C}$ , and total RNA was extracted within a few weeks using a QIAGEN kit specially developed for the PAXgene RNA tubes. RNA quality and quantity was determined using NanoDrop ND-1000. Extracted RNA was stored at  $-80^{\circ}\text{C}$  until use.

### Microarray Analysis

Gene expression profiling was performed using the HumanRef-8 Expression BeadChip (Illumina, San Diego, CA). The BeadChip targets  $\sim 24,526$  probes. Probe preparation and hybridization were carried out according to manufacturer's recommendations. The microarray data are Minimum Information About a Microarray Experiment compliant and have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus repository and are accessible through Gene Expression Omnibus Series accession number GSE51058.

### High-Throughput Real-Time RT-PCR

High-throughput real-time RT-PCR was performed on the BioMark 96.96 Dynamic Array (Fluidigm) with TaqMan Gene Expression Assays (Applied Biosystems). Briefly, an aliquot of total RNA was converted to cDNA through a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), then cDNA was preamplified with pooled gene assays. Preamplified cDNA was loaded into BioMark Dynamic Array chips using the NanoFlex IFC controller. Threshold cycle, as a measurement of relative fluorescence intensity, was extracted by BioMark Real-Time PCR Analysis software. Threshold cycle values were directly used in data analysis after normalization to eight preselected reference genes: *RNF31*, *TM9SF4*, *ABL*, *PP1B*, *ESD*, *FPGS*, *MRPL19*, and *TRAP1* (Supplementary Table 1). Technical reproducibility was examined using a pooled cDNA sample aliquot.

**Table 1—Clinical characteristics of AbP subjects in the whole cohort and separate data sets**

Characteristic	Progressors	Nonprogressors	<i>P</i> value
<b>Total</b>			
Sample size ( <i>n</i> )	39	65	
Age at sampling (years)	7.41 (1.74–15.02)	9.05 (1.55–45.08)	4.07E-04
Age at first AbP (years)	3.48 (0.77–12.56)	7.28 (0.84–40.78)	1.79E-05
Follow-up time (years)	5.64 (0.21–12.12)	8.9 (5.5–16.35)	6.77E-08
Male sex	22 (56.41)	34 (52.31)	0.84
FDR	23 (58.97)	27 (41.54)	0.13
Genetic risk (H/M/L/unknown)	19/7/12/1	16/16/30/3	2.12E-02
≥2 AbP	31 (79.49)	24 (36.92)	6.15E-05
Interval time*	2.60 ± 2.44	1.66 ± 1.94	3.32E-02
<b>Microarray data set</b>			
Sample size ( <i>n</i> )	21	15	
Age at sampling (years)	7.66 (1.74–10.56)	7.86 (2.25–15.18)	0.34
Age at first AbP (years)	3.15 (0.77–9.77)	4.53 (0.84–13.46)	0.12
Follow-up time (years)	6.12 (0.21–11.41)	8.38 (6.23–16.35)	2.09E-03
Male sex	9 (43)	8 (53.33)	0.78
FDR	10 (48)	8 (53.33)	1.00
Genetic risk (H/M/L/unknown)	11/5/5/0	4/5/6/0	0.23
≥2 AbP	20 (95)	14 (93.33)	1.00
Interval time*	2.59 ± 2.48	1.94 ± 2.16	0.42
<b>RT-PCR data set</b>			
Sample size ( <i>n</i> )	18	50	
Age at sampling (years)	7.22 (3.48–15.02)	9.68 (1.55–45.08)	1.82E-03
Age at first AbP (years)	4.02 (0.81–12.56)	8.95 (1.3–40.78)	1.75E-04
Follow-up time (years)	4.6 (2.48–12.12)	9.2 (5.5–15.65)	2.46E-04
Male sex	13 (72.22)	26 (52)	0.23
FDR	13 (72.22)	19 (38)	2.65E-02
Genetic risk (H/M/L/unknown)	8/2/7/1	12/11/24/3	0.18
≥2 AbP	11 (61.11)	10 (20)	3.28E-03
Interval time*	2.60 ± 2.55	1.57 ± 1.88	0.07

Data are median (range), *n* (%), and mean ± SD unless otherwise indicated. H, high; L, low; M, medium. \*Time between first antibody detection and collection of the sample tested.

### Statistical Analysis

All analyses were performed using the R statistics package and environment for statistical computing ([www.r-project.org](http://www.r-project.org)). The lumi package in R was used for pre-processing and normalization of the microarray data using variance stabilizing transformation and robust spline normalization. Probes not detected in >50% of the samples were excluded from further analysis. We used Cox proportional hazard models to evaluate the impact of gene expression levels on diabetes-free survival. Diabetes-free survival was calculated as time from first antibody-positive date to the T1D diagnosis date. Subjects with no evidence of disease were censored at the date of the last follow-up visit. The subjects were divided into two groups (low and high) using the median gene expression value of each individual gene as a cutoff. Kaplan-Meier plots were used to compare differences in diabetes-free survival between these groups. A combination of log-rank test *P* value and absolute value of the hazard ratio (HR) was used for selecting significant genes. Only genes having a significant effect in the univariate analysis were entered into the multivariate analyses to assess the combined effect of various genes on diabetes-free survival. For the models using multiple genes, self-organizing map cluster analysis (24) was used to divide the subjects into two

clusters (C1 and C2) based on the gene expression patterns. Kaplan-Meier plots were used to compare the differences in diabetes-free survival between these groups. The consistency between models was assessed by calculating the percentage of subjects classified in the same group by model. Cox regression analysis was used to evaluate the independent prognostic value of gene expression by including known confounding factors, such as age at first AbP, the number of autoantibodies present, genetic risk, and FDR status, in the regression models. The Akaike information criterion (AIC) was calculated to compare the quality of fit for each model.

### RESULTS

#### Identification of Genes Associated With Progression From Autoimmunity to Diabetes

Microarray analysis was carried out on peripheral blood mononuclear cells from 36 AbP subjects, 21 of whom had progressed to T1D by the end of follow-up. Among the 12,904 genes tested, 454 showed significant associations ( $P < 0.05$ ) with T1D progression, and 112 of them had an HR >3 (or <0.33).

To validate these findings, the top-27 genes were selected for real-time RT-PCR confirmation in an independent cohort that included 68 AbP subjects (18

progressors and 50 nonprogressors). Genes were first selected based on HR and *P* values (HR or 1/HR close to 3, *P* < 0.02) and were further prioritized based on functional relevance to autoimmunity (especially T1D), inflammation, cellular activation, or proliferation. Three randomly selected negative control genes (*XRCC2*, *FKBP11*, and *ORM1*) were also included for validation (Table 2). Among the 27 candidate genes, 3 (*BACH2*, *TXNDC5*, and *CDC20*) had a confirmed high HR (3.9, 3.3, and 2.8, respectively) and significant *P* values (*P* < 0.05), whereas 2 (*IGLL3* and *EIF3A*) had a confirmed good HR (2.6 and 2.3, respectively) and marginal *P* values (*P* = 0.055 and 0.074, respectively). The remaining 22 candidate genes and the 3 negative control genes were excluded from further study because of insignificant *P* values (*P* > 0.1) or inconsistent

HRs in the discovery and validation data sets (Table 2). As shown in Fig. 1, the higher expression of these five genes is associated with a higher risk of T1D progression. The difference in mean gene expression levels of the high-expression subset and low-expression subset are also shown in Fig. 1 (fold change 1.6–3.9 in the validation data set) and are consistent with the microarray data set for four of the five genes.

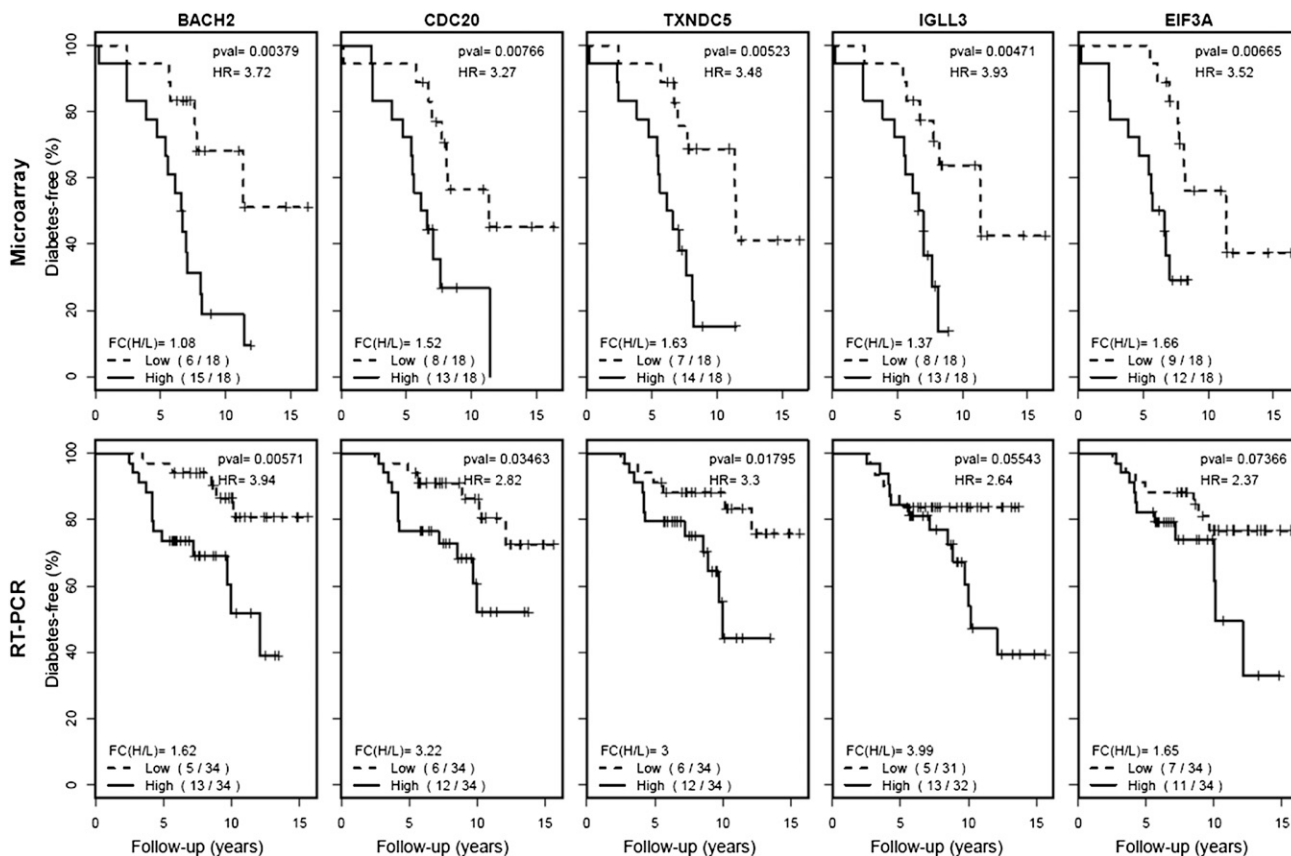
### Multigene Expression Models Are Better Predictors of Progression From AbP to T1D

It is well known that predictive power using multiple genes may be significantly better than that of any single gene. Thus, we evaluated the predictive value of models containing combinations of the five genes selected by the single-gene analysis (*BACH2*, *IGLL3*, *TXNRD5*, *CDC20*,

**Table 2—Genes correlated with T1D progression from AbP**

Gene	Microarray				RT-PCR			
	HR (95% CI)	<i>P</i> value	FC	CV	HR (95% CI)	<i>P</i> value	FC	CV
<i>TXNDC5</i>	3.48 (1.38–8.78)	5.23E-03	1.63	6.86	3.30 (1.17–9.32)	1.80E-02	3.00	4.69
<i>CDC20</i>	3.27 (1.31–8.16)	7.66E-03	1.52	6.50	2.82 (1.04–7.67)	3.46E-02	3.22	5.75
<i>IGLL3</i>	3.93 (1.43–10.80)	4.71E-03	1.37	4.60	2.64 (0.94–7.42)	5.54E-02	3.99	9.27
<i>EIF3A</i>	3.52 (1.35–9.21)	6.65E-03	1.66	5.37	2.37 (0.90–6.28)	7.37E-02	1.65	1.91
<i>BACH2</i>	3.72 (1.43–9.62)	3.79E-03	1.08	1.04	3.94 (1.39–11.21)	5.71E-03	1.62	1.92
<i>LHPP</i>	0.23 (0.09–0.62)	1.83E-03	1.45	3.92	0.84 (0.33–2.15)	7.17E-01	2.13	6.42
<i>ETS2</i>	0.35 (0.14–0.89)	2.12E-02	1.22	2.28	0.54 (0.21–1.41)	2.02E-01	2.26	3.64
<i>TMEM91</i>	0.31 (0.12–0.78)	9.41E-03	1.71	5.12	0.63 (0.24–1.63)	3.34E-01	1.89	2.53
<i>RNF167</i>	0.21 (0.08–0.57)	8.51E-04	1.30	2.87	0.83 (0.33–2.10)	6.95E-01	1.57	1.89
<i>ATP6V1F</i>	0.24 (0.09–0.66)	2.74E-03	1.47	3.55	0.65 (0.25–1.69)	3.77E-01	1.87	2.38
<i>ANKRD9</i>	0.19 (0.07–0.54)	5.71E-04	1.91	6.88	2.56 (0.90–7.31)	6.91E-02	2.41	4.74
<i>CHCHD5</i>	0.31 (0.12–0.83)	1.35E-02	1.36	3.77	1.55 (0.59–4.07)	3.72E-01	2.49	7.17
<i>DOCK8</i>	3.94 (1.43–10.81)	4.62E-03	1.58	4.29	0.61 (0.24–1.59)	3.10E-01	2.07	3.06
<i>ESAM</i>	0.19 (0.07–0.53)	4.57E-04	1.35	3.59	0.85 (0.33–2.14)	7.24E-01	2.37	3.89
<i>F5</i>	0.25 (0.09–0.65)	2.40E-03	1.29	3.59	0.49 (0.18–1.30)	1.43E-01	3.76	5.82
<i>HP</i>	0.30 (0.11–0.81)	1.22E-02	1.39	4.60	1.20 (0.40–3.61)	7.36E-01	6.04	9.28
<i>ITGB5</i>	0.28 (0.11–0.74)	6.31E-03	1.75	6.16	0.64 (0.25–1.65)	3.51E-01	3.16	5.52
<i>KHDRBS1</i>	4.39 (1.61–12.00)	1.87E-03	1.43	3.21	1.61 (0.62–4.15)	3.21E-01	1.61	1.92
<i>LY6G6F</i>	0.25 (0.09–0.68)	3.45E-03	1.36	3.65	1.23 (0.46–3.23)	6.81E-01	2.84	5.41
<i>MAP2K1</i>	2.84 (1.13–7.18)	2.18E-02	1.32	2.83	1.39 (0.54–3.61)	4.91E-01	1.66	2.18
<i>NFE2</i>	0.38 (0.15–0.97)	3.65E-02	1.77	5.11	0.58 (0.23–1.49)	2.53E-01	2.31	3.44
<i>PLD3</i>	0.25 (0.09–0.68)	3.82E-03	1.40	3.35	1.56 (0.60–4.03)	3.60E-01	2.23	3.56
<i>PRDX5</i>	0.30 (0.12–0.75)	6.58E-03	1.81	5.62	1.62 (0.62–4.20)	3.17E-01	1.88	2.49
<i>TRAPPC5</i>	0.14 (0.05–0.43)	7.39E-05	1.46	3.88	1.64 (0.62–4.33)	3.13E-01	2.41	4.98
<i>TSC1</i>	3.76 (1.42–9.98)	4.47E-03	1.45	3.88	1.49 (0.59–3.81)	3.98E-01	1.53	1.79
<i>TSPAN33</i>	0.25 (0.10–0.63)	1.69E-03	1.67	5.22	1.10 (0.43–2.79)	8.44E-01	1.85	2.68
<i>UBP1</i>	3.70 (1.42–9.65)	4.77E-03	1.44	3.45	1.04 (0.41–2.62)	9.33E-01	1.45	1.58
<i>XRCC2</i>	1.36 (0.57–3.24)	4.93E-01	1.52	6.64	1.16 (0.46–2.94)	7.52E-01	2.19	4.37
<i>FKBP11</i>	2.18 (0.91–5.21)	7.43E-02	1.35	3.71	1.30 (0.51–3.31)	5.76E-01	2.93	5.95
<i>ORM1</i>	0.42 (0.17–1.05)	5.51E-02	1.80	6.59	0.86 (0.34–2.18)	7.44E-01	7.22	10.33

CV, coefficient of variance; FC, fold change.



**Figure 1**—Diabetes-free survival in two AbP subgroups (high and low) stratified according to expression of individual genes (*BACH2*, *CDC20*, *TXNDC5*, *IGLL3*, and *EIF3A*). HR, *P* value, fold change between high- and low-expression groups [FC(H/L)] and number of progressors in each subgroup are shown for each gene in both microarray and real-time RT-PCR data sets.

and *EIF3A*). With five genes, there is a total of 25 two-, three-, and four-gene models. In the discovery data set, 17 of the 25 models can distinguish two clusters of subjects with significant differences in T1D progression (Supplementary Table 2). Further assessment of these models in the validation cohort confirmed 14 of the 17 models to have predictive value. It should be noted that not every confirmed combination model had an increased predictive power over the single genes. Four models were preferred for their larger HR values, more significant *P* values, and lower AIC values (Table 3). Model BI (comprising the *BACH2* and *IGLL3* genes) had an HR of 7.34 (adjusted *P* = 0.004) and AIC of 122.02. Model ICE (comprising *IGLL3*, *CDC20*, and *EIF3A*) had an HR of 6.32 (adjusted *P* = 0.007) and AIC of 122.02. Model BICE (comprising *BACH2*, *IGLL3*, *CDC20*, and *EIF3A*) had an HR of 7.76 (adjusted *P* = 0.003) and AIC of 121.12. Finally, model BITE (comprising *BACH2*, *IGLL3*, *TXNDC5*, and *EIF3A*) had an HR of 7.76 (adjusted *P* = 0.003) and AIC of 121.12. As shown in Fig. 2, these four models consistently stratify AbP subjects with differential progression rates in both the discovery and the validation data sets.

### Consistency Among Predictive Models

Having shown the utility of the multigene models in risk stratification of AbP subjects, we next assessed the consistency of risk stratification for individual subjects using the various models. Fig. 3 shows the consistency among all possible combinations of these top-four models. A major conclusion from these results is that 75% of the individuals are consistently classified into the same progression group by all four models. Further examination of the data suggests that model ICE has less consistency than the other three models (BI, BICE, and BITE). The mean consistency among models excluding ICE is >89% for both data sets, whereas the mean consistencies for models including ICE are 82.6% and 84.4% for the discovery and validation data sets, respectively. Indeed, ICE has the lowest HR and highest AIC in both the discovery and the validation data sets compared with the other three models (Fig. 2, Table 3).

### Gene Expression Models Are Independent Predictors of T1D Progression

Several demographic parameters and molecules are well-established risk factors associated with T1D progression. These factors include age at first appearance of

**Table 3—Multigene expression models for T1D risk classification**

Combination	Microarray			RT-PCR		
	HR (95% CI)	Adjusted <i>P</i> value	AIC	HR (95% CI)	Adjusted <i>P</i> value	AIC
BI	3.10 (1.24–7.76)	2.05E-02	122.93	7.34 (2.07–25.99)	3.88E-03	122.02
ICE	2.51 (1.05–6.01)	4.25E-02	125.19	6.32 (1.82–21.96)	6.69E-03	123.62
BICE	3.06 (1.23–7.66)	2.05E-02	123.08	7.76 (2.20–27.39)	2.78E-03	121.12
BITE	5.86 (2.16–15.91)	2.65E-03	115.94	7.76 (2.20–27.39)	2.78E-03	121.12
BC	2.96 (1.21–7.24)	2.09E-02	123.35	3.59 (1.26–10.22)	2.05E-02	132.18
BT	3.35 (1.34–8.34)	1.50E-02	122.01	2.90 (1.08–7.81)	4.25E-02	133.94
BCT	3.86 (1.58–9.47)	7.78E-03	120.23	2.83 (1.06–7.59)	4.25E-02	134.22
BCI	3.99 (1.61–9.88)	7.78E-03	120.04	4.04 (1.43–11.44)	1.20E-02	127.28
BIT	5.86 (2.16–15.91)	2.65E-03	115.94	4.96 (1.62–15.22)	8.99E-03	125.50
BIE	3.11 (1.24–7.76)	2.05E-02	122.93	4.88 (1.58–15.01)	9.24E-03	125.81
CET	2.52 (1.06–5.96)	4.25E-02	125.07	2.83 (1.06–7.59)	4.25E-02	134.22
TIE	3.63 (1.42–9.25)	1.16E-02	121.92	3.63 (1.28–10.34)	2.05E-02	128.57
BICT	3.65 (1.49–8.98)	9.46E-03	120.99	2.84 (1.06–7.63)	4.25E-02	130.72
BCTE	3.86 (1.58–9.46)	7.78E-03	120.23	2.97 (1.09–8.06)	4.23E-02	133.93
TI	3.63 (1.42–9.25)	1.16E-02	121.92	1.46 (0.57–3.75)	0.43	134.57
CIT	3.36 (1.39–7.93)	1.16E-02	122.18	2.51 (0.94–6.72)	0.07	131.67
CITE	3.33 (1.39–7.93)	1.16E-02	122.18	2.53 (0.95–6.75)	0.07	131.63

autoantibodies, the number of autoantibodies present, genetic factors, and FDR status. With the exception of FDR status, significant differences in these factors were observed between the progressors and the nonprogressors (Table 1). To examine whether these known factors were confounding factors for the gene expression models on T1D risk stratification, Cox regression analysis was used to analyze both the discovery and the validation data sets. According to multivariate analyses, three of the four gene expression models (BI, BICE, and BITE) are independent predictors of T1D progression with significant adjusted *P* values ( $P < 0.05$ ) (Table 4).

The study also suggests that using gene expression in conjunction with other known risk markers improves the risk stratification for AbP subjects. Consistent with previous findings, subjects with more than one autoantibody have a significantly higher risk for T1D progression than those with only one autoantibody (HR 3.07,  $P < 0.05$ ) (Supplementary Fig. 2A). Of note, the subjects with more than one autoantibody can be further stratified into a high-risk and a low-risk group by model BI (HR 7.12,  $P < 0.05$ ) (Supplementary Fig. 2B).

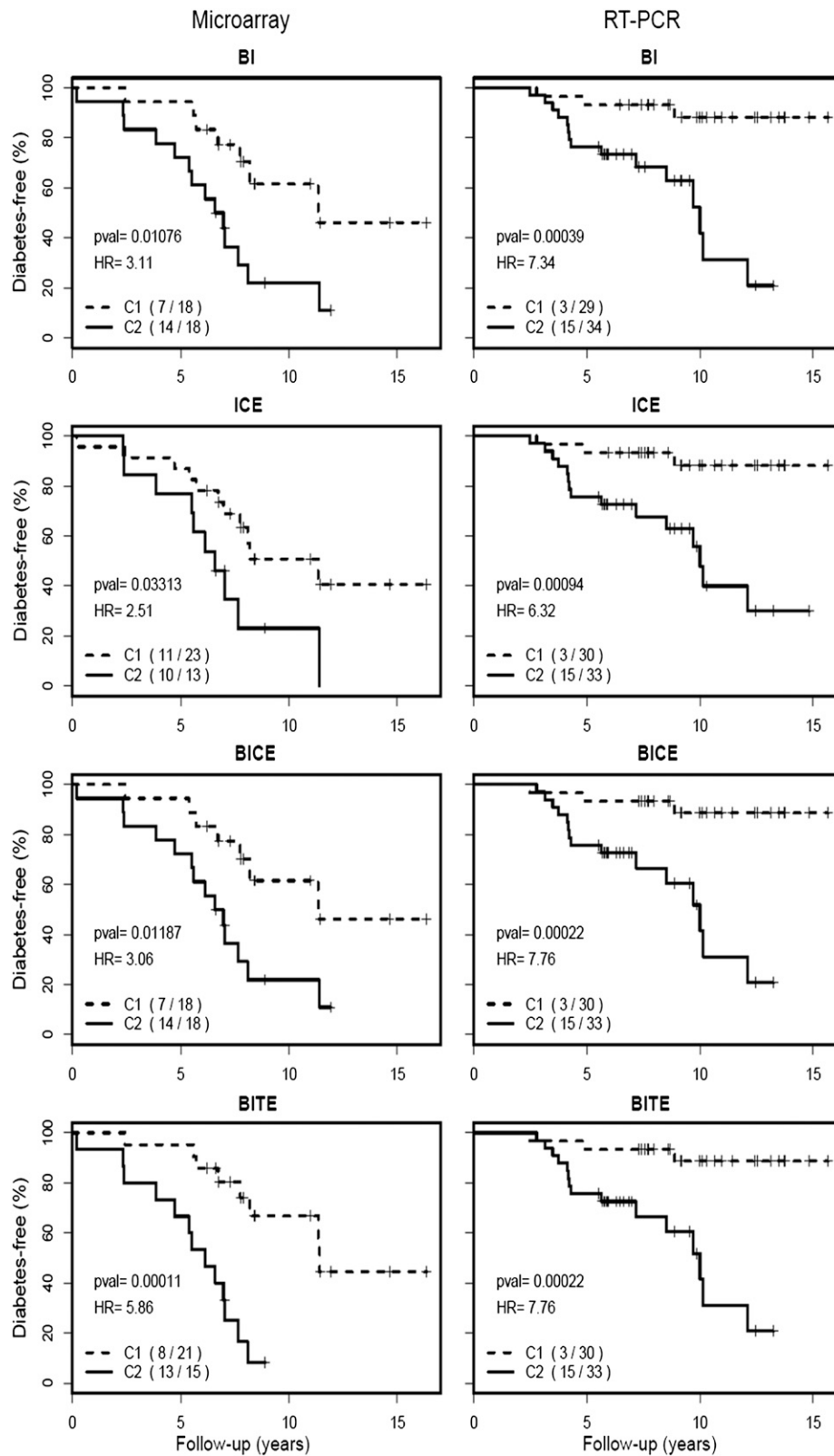
## DISCUSSION

Islet autoantibodies are the current gold standard for risk stratification for the development of T1D and have been used in large population-based studies, including TEDDY (The Environmental Determinants of Diabetes in the Young) and many clinical trials for T1D prevention (25–27). However, testing for islet autoantibodies is not specific enough even when multiple autoantibodies are

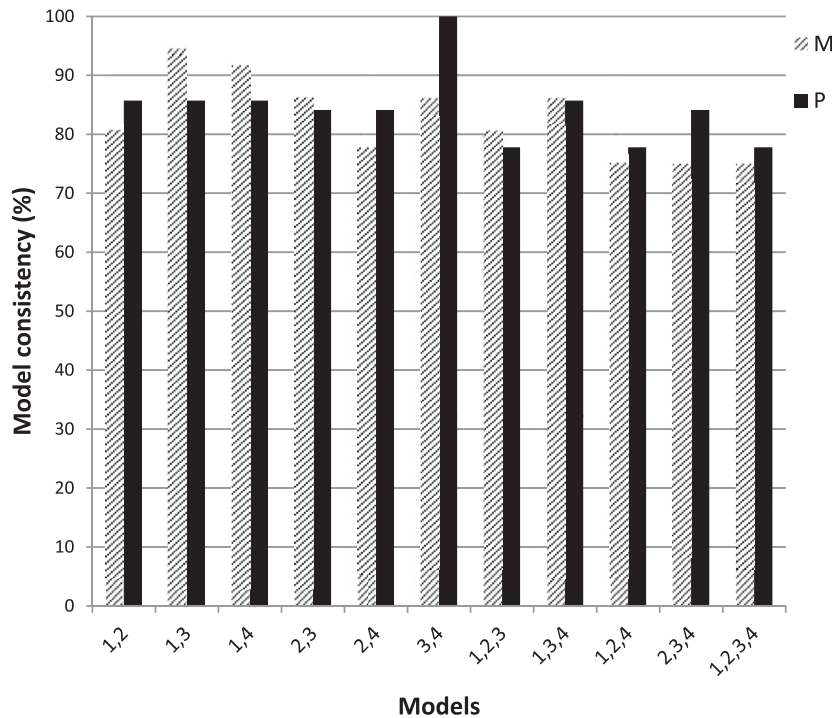
detected. The aim of the current study was to investigate whether gene expression patterns can be used as biomarkers to complement islet autoantibodies for T1D risk stratification. We demonstrate that expression changes of certain genes are associated with T1D progression from autoimmunity. The predictive value of single genes is generally moderate (HRs of 2.4–3.9) and may be further limited because the risk assignment by different genes may not be consistent enough. Several multigene models have shown much better predictive power than individual genes. One two-gene model (*BACH2* and *IGLL3*) identified two AbP subgroups with very different progression rates (HR 7.3). Adding two other genes (*CDC20* and *EIF3A* or *TXNDC5* and *EIF3A*) did not significantly improve the HR values (HR 7.8). However, the two four-gene models are highly consistent in risk assignment (consistency of 100% in the validation data set). Because risk of T1D progression among AbP subjects is influenced by many other factors, such as family history of diabetes, genetic factors, and number of autoantibodies, it was necessary to determine whether the risk associated with differential gene expression is independent of these known factors. Cox regression analyses suggested that risk attributed to gene expression could not be accounted for by the risk factors examined in this study. Indeed, gene expression can be used together with the number of autoantibodies to improve risk stratification.

This study has demonstrated for the first time in our knowledge that gene expression patterns may serve as biomarkers to stratify T1D risk among AbP children. When compared with genetic markers (HLA and non-HLA)





**Figure 2**—Diabetes-free survival in two AbP subgroups (C1 and C2) stratified according to multigene expression models (BI, ICE, BICE, BITE). HR, *P* value, and number of progressors in each subgroup are shown for each model in both microarray and real-time RT-PCR data sets.



**Figure 3**—Risk stratification consistency in individual subjects by model. Model 1, BI; model 2, ICE; model 3, BICE; and model 4, BITE. M, microarray data set; P, real-time RT-PCR data set.

(4,6–12) and metabolic markers (13–15), the present gene expression markers have several strengths. First, the intrinsic deficiencies of genetic and metabolic markers limit their usefulness for T1D prevention. For example, genetic markers cannot be used as surrogate markers for therapeutic outcomes. Whereas metabolic anomalies are indicative of pancreatic islet cell dysfunction, they are only useful markers at a very late stage of disease progression when the opportunity for disease prevention may have passed. In contrast, gene expression markers can be used for both early disease prediction and therapeutic evaluation because of their dynamic change during disease progression. Second, the identified gene expression

markers were shown to further stratify AbP subjects into subgroups with different progression rates to T1D. High HR values (up to 7.76), consistency of their predictive values between discovery and validation data sets, and high model consistency (75–100%) indicate that these gene expression biomarkers have a potentially strong predictive power and reliability. Finally, cohorts in most previous studies were restricted to children with a family history of T1D, such as the genetic studies in BABYDIAB (10–12), the Belgian Diabetes Registry (7), and DiMe (Childhood Diabetes in Finland) (8) and metabolic marker studies in Diabetes Prevention Trial–Type 1 (13–15). The DAISY cohort used in the current study includes children

**Table 4**—Cox regression analysis for multigene expression models

Variable	Univariate			Multivariate		
	P value	HR (95% CI)	AIC	P value	HR (95% CI)	AIC
<b>Microarray</b>						
Model 1: BI	1.08E-02	3.11 (1.24–7.76)	122.93	2.97E-02	3.93 (1.14–13.50)	127.99
Model 2: ICE	3.31E-02	2.51 (1.05–6.00)	125.19	0.14	2.35 (0.76–7.33)	131.27
Model 3: BICE	1.19E-02	3.06 (1.23–7.66)	123.08	3.69E-02	3.56 (1.08–11.75)	128.56
Model 4: BITE	1.06E-04	5.86 (2.17–15.91)	115.94	8.76E-04	12.54 (2.83–55.10)	118.48
<b>RT-PCR</b>						
Model 1: BI	3.88E-04	7.34 (2.1–26.0)	122.02	1.68E-02	5.53 (1.37–22.47)	111.55
Model 2: ICE	9.37E-04	6.32 (1.82–21.96)	123.62	7.75E-03	6.17 (1.62–23.56)	109.50
Model 3: BICE	2.22E-04	7.76 (2.20–27.39)	121.12	8.52E-03	6.48 (1.61–26.07)	110.06
Model 4: BITE	2.22E-04	7.76 (2.20–27.39)	121.12	8.52E-03	6.48 (1.61–26.07)	110.06

In addition to gene expression factor, age at first AbP, the number of autoantibodies present, genetic risk, and FDR status were included in multivariate models.



from the general population and with T1D relatives. The Cox regression results suggest that the gene expression markers are independent of family history.

The limitations of this study include the absence of zinc transporter 8 antibody information and the relatively small number of AbP subjects, especially when split into two sample sets. Despite the potential utility of these gene expression models in risk stratification, it should be noted that as a result of population differences and sampling bias, the best genes and models may not be selected in this study. However, the experimental and analytical approaches advocated in this study should serve as a model for future studies using much larger sample sizes. Also essential is that the candidate genes and models identified in this study are further validated in other cohorts.

From a functional perspective, the genes identified in this study may provide novel insights into the disease mechanism. *BACH2* and *IgLL3* are two B-cell-related genes. *IgLL3* encodes the Ig $\lambda$ -like polypeptide 3, which is expressed mainly in pre-B cells and involved in B-cell development (28). The present study is the first in our knowledge to suggest a potential role of this gene in T1D or other autoimmune diseases and should be investigated further in T1D human and animal models. *BACH2* is a B-cell-specific transcription factor critical for Ig class switching. B cells with lower levels of *BACH2* mature directly to IgM plasma cells, whereas cells with higher levels undergo class switching and differentiate to other cell types, mainly IgG-secreting plasma cells, including autoantibody-secreting cells (29). Higher expression of *BACH2* is associated with a faster and higher progression rate from AbP to T1D, consistent with the concept that antibody switching may be a potential risk factor that can differentiate progression risks from AbP to T1D (30). Genetic polymorphisms within the *BACH2* gene are associated with several autoimmune diseases, including T1D (31), Crohn disease (32), celiac disease (33), multiple sclerosis (34), and vitiligo (35). Although T1D was considered an ultimately T-cell-mediated autoimmune disorder, accumulating evidence in animal models and human studies indicates a crucial role for B cells in T1D development. The present findings further strengthen the connection between B cells and T1D.

*EIF3A* and *CDC20* are involved in cell cycle regulation (36,37). The association of the two cell cycle-related genes with T1D progression is consistent with a suspected increase in the activation and proliferation of immune cells, especially B cells and T cells, in AbP subjects who are progressing to T1D. *TXNDC5* encodes the thioredoxin domain-containing 5 protein, a protein-disulfide isomerase that possesses thioredoxin activity. *TXNDC5* may decrease receptor binding activity of insulin by catalyzing the reduction of insulin disulfide bonds (38). *TXNDC5* may increase T1D progression risk by promoting glucose intolerance through reducing insulin receptor binding

activity, a hypothesis that should be verified in future studies.

In summary, this retrospective study using transcriptomic profiling and RT-PCR validation suggests that five genes may be involved in T1D pathogenesis and may serve as biomarkers for risk stratification among AbP subjects. The findings further indicate that gene expression patterns of multiple genes have much better predictive value and higher consistency in risk assessment than individual genes.

---

**Funding.** This work was supported by grants from the National Institutes of Health (R01-DK-32083, DK-32493, R01-DK-050979) to M.R. and (4R33-DK-069878 and 2R01-HD-37800) to J.-X.S. and from the JDRF (17-2013-535) to M.R.

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Author Contributions.** Y.J. contributed to research design, data acquisition, and manuscript preparation. A.S. and S.B. contributed to the statistical analysis and data interpretation. C.D. and H.L. contributed to data generation. D.H. and K.B. contributed to sample and clinical information collection. M.R. contributed to the clinical data and samples, data interpretation, and manuscript preparation. J.-X.S. supervised the overall study design and data interpretation and edited the manuscript. J.-X.S. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

- Knip M, Korhonen S, Kulmala P, et al. Prediction of type 1 diabetes in the general population. *Diabetes Care* 2010;33:1206–1212
- Wenzlau JM, Hutton JC. Novel diabetes autoantibodies and prediction of type 1 diabetes. *Curr Diab Rep* 2013;13:608–615
- Skyler JS. Prediction and prevention of type 1 diabetes: progress, problems, and prospects. *Clin Pharmacol Ther* 2007;81:768–771
- Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* 2013;309:2473–2479
- Steck AK, Johnson K, Barriga KJ, et al. Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. *Diabetes Care* 2011;34:1397–1399
- Lipponen K, Gombos Z, Kiviniemi M, et al. Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. *Diabetes* 2010;59:3253–3256
- Mbunwe E, Van der Auwera BJ, Vermeulen I, et al.; Belgian Diabetes Registry. HLA-A\*24 is an independent predictor of 5-year progression to diabetes in autoantibody-positive first-degree relatives of type 1 diabetic patients. *Diabetes* 2013;62:1345–1350
- Mrena S, Virtanen SM, Laippala P, et al. Models for predicting type 1 diabetes in siblings of affected children. *Diabetes Care* 2006;29:662–667
- Lempainen J, Hermann R, Veijola R, Simell O, Knip M, Ilonen J. Effect of the PTPN22 and INS risk genotypes on the progression to clinical type 1 diabetes after the initiation of  $\beta$ -cell autoimmunity. *Diabetes* 2012;61:963–966
- Bonifacio E, Krumsiek J, Winkler C, Theis FJ, Ziegler AG. A strategy to find gene combinations that identify children who progress rapidly to type 1 diabetes after islet autoantibody seroconversion. *Acta Diabetol*. 19 November 2013 [Epub ahead of print]
- Achenbach P, Hummel M, Thümer L, Boerschmann H, Höfelmann D, Ziegler AG. Characteristics of rapid vs slow progression to type 1 diabetes in multiple islet autoantibody-positive children. *Diabetologia* 2013;56:1615–1622

12. Winkler C, Krumsiek J, Lempainen J, et al. A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes. *Genes Immun* 2012;13:549–555
13. Sosenko JM, Krischer JP, Palmer JP, et al.; Diabetes Prevention Trial-Type 1 Study Group. A risk score for type 1 diabetes derived from autoantibody-positive participants in the Diabetes Prevention Trial-Type 1. *Diabetes Care* 2008;31:528–533
14. Xu P, Beam CA, Cuthbertson D, Sosenko JM, Skyler JS, Krischer JP; DPT-1 Study Group. Prognostic accuracy of immunologic and metabolic markers for type 1 diabetes in a high-risk population: receiver operating characteristic analysis. *Diabetes Care* 2012;35:1975–1980
15. Xu P, Wu Y, Zhu Y, et al.; Diabetes Prevention Trial-Type 1 (DPT-1) Study Group. Prognostic performance of metabolic indexes in predicting onset of type 1 diabetes. *Diabetes Care* 2010;33:2508–2513
16. Elo LL, Mykkänen J, Nikula T, et al. Early suppression of immune response pathways characterizes children with prediabetes in genome-wide gene expression profiling. *J Autoimmun* 2010;35:70–76
17. Han D, Cai X, Wen J, et al. Innate and adaptive immune gene expression profiles as biomarkers in human type 1 diabetes. *Clin Exp Immunol* 2012;170:131–138
18. Jin Y, She JX. Novel biomarkers in type 1 diabetes. *Rev Diabet Stud* 2012;9:224–235
19. Jin Y, Sharma A, Carey C, et al. The expression of inflammatory genes is upregulated in peripheral blood of patients with type 1 diabetes. *Diabetes Care* 2013;36:2794–2802
20. Kaizer EC, Glaser CL, Chaussabel D, Banchereau J, Pascual V, White PC. Gene expression in peripheral blood mononuclear cells from children with diabetes. *J Clin Endocrinol Metab* 2007;92:3705–3711
21. Padmos RC, Schloot NC, Beyan H, et al.; LADA Consortium. Distinct monocyte gene-expression profiles in autoimmune diabetes. *Diabetes* 2008;57:2768–2773
22. Reynier F, Pachot A, Paye M, et al. Specific gene expression signature associated with development of autoimmune type-1 diabetes using whole-blood microarray analysis. *Genes Immun* 2010;11:269–278
23. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 1996;39:807–812
24. Kohonen T. Essentials of the self-organizing map. *Neural Netw* 2013;37:52–65
25. Hagopian WA, Lernmark A, Rewers MJ, et al. TEDDY—The Environmental Determinants of Diabetes in the Young: an observational clinical trial. *Ann N Y Acad Sci* 2006;1079:320–326
26. Skyler JS, Krischer JP, Wolfsdorf J, et al. Effects of oral insulin in relatives of patients with type 1 diabetes: The Diabetes Prevention Trial—Type 1. *Diabetes Care* 2005;28:1068–1076
27. Skyler JS, Greenbaum CJ, Lachin JM, et al.; Type 1 Diabetes TrialNet Study Group. Type 1 Diabetes TrialNet—an international collaborative clinical trials network. *Ann N Y Acad Sci* 2008;1150:14–24
28. Bauer TR Jr., McDermid HE, Budarf ML, Van Keuren ML, Blomberg BB. Physical location of the human immunoglobulin lambda-like genes, 14.1, 16.1, and 16.2. *Immunogenetics* 1993;38:387–399
29. Muto A, Ochiai K, Kimura Y, et al. Bach2 represses plasma cell gene regulatory network in B cells to promote antibody class switch. *EMBO J* 2010;29:4048–4061
30. Hutchings P, Tonks P, Cooke A. Effect of MHC transgene expression on spontaneous insulin autoantibody class switch in nonobese diabetic mice. *Diabetes* 1997;46:779–784
31. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet* 2008;40:1399–1401
32. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118–1125
33. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295–302
34. Sawcer S, Hellenthal G, Pirinen M, et al.; International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011;476:214–219
35. Jin Y, Birlea SA, Fain PR, et al. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet* 2012;44:676–680
36. Saletta F, Suryo Rahmanto Y, Richardson DR. The translational regulator eIF3a: the tricky eIF3 subunit! *Biochim Biophys Acta* 2010;1806:275–286
37. Yu H. Cdc20: a WD40 activator for a cell cycle degradation machine. *Mol Cell* 2007;27:3–16
38. Holmgren A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J Biol Chem* 1979;254:9627–9632