

# A peripheral blood transcriptomic signature predicts autoantibody development in infants at risk of type 1 diabetes

Ahmed M. Mehdi,<sup>1</sup> Emma E. Hamilton-Williams,<sup>1</sup> Alexandre Cristino,<sup>1</sup> Anette Ziegler,<sup>2</sup> Ezio Bonifacio,<sup>3</sup> Kim-Anh Le Cao,<sup>1</sup> Mark Harris,<sup>1,4</sup> and Ranjeny Thomas<sup>1</sup>

<sup>1</sup>The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Translational Research Institute, Woolloongabba, Australia. <sup>2</sup>Institute of Diabetes Research, Helmholtz Zentrum Munchen, Neuherberg, and Forschergruppe Diabetes, Klinikum rechts der Isar, Institut für Diabetesforschung, Neuherberg, Germany. <sup>3</sup>CRTD-DFG Research Center for Regenerative Therapies Dresden, Medical Faculty Gustav Carus, Technische Universität Dresden, Dresden, Germany. <sup>4</sup>Lady Cilento Children's Hospital, South Brisbane, Australia

**Autoimmune-mediated destruction of pancreatic islet  $\beta$  cells results in type 1 diabetes (T1D). Serum islet autoantibodies usually develop in genetically susceptible individuals in early childhood before T1D onset, with multiple islet autoantibodies predicting diabetes development. However, most at-risk children remain islet-antibody negative, and no test currently identifies those likely to seroconvert. We sought a genomic signature predicting seroconversion risk by integrating longitudinal peripheral blood gene expression profiles collected in high-risk children included in the BABYDIET and DIPP cohorts, of whom 50 seroconverted. Subjects were followed for 10 years to determine time of seroconversion. Any cohort effect and the time of seroconversion were corrected to uncover genes differentially expressed (DE) in seroconverting children. Gene expression signatures associated with seroconversion were evident during the first year of life, with 67 DE genes identified in seroconverting children relative to those remaining antibody negative. These genes contribute to T cell-, DC-, and B cell-related immune responses. Near-birth expression of *ADCY9*, *PTCH1*, *MEX3B*, *IL15RA*, *ZNF714*, *TENM1*, and *PLEKHA5*, along with HLA risk score predicted seroconversion (AUC 0.85). The ubiquitin-proteasome pathway linked DE genes and T1D susceptibility genes. Therefore, a gene expression signature in infancy predicts risk of seroconversion. Ubiquitination may play a mechanistic role in diabetes progression.**

## Introduction

Type 1 diabetes (T1D) results from autoimmune-mediated destruction and dysfunction of insulin-producing pancreatic  $\beta$  cells (1). Prior to the onset of T1D, genetically susceptible individuals develop islet autoantibodies, often during the first few years of life. Almost all individuals who develop multiple islet autoantibodies will progress to diabetes (2). The increasing incidence of T1D, and the fact that only a small proportion of genetically at-risk individuals develop autoantibodies and progress to clinical diabetes (3), highlights the important contribution of environmental factors. Despite intensive research, the gene-environment interactions that trigger progression to T1D are poorly understood.

There is currently no test to predict which genetically susceptible children are likely to become islet antibody positive (ABP). Improved methods to identify which at-risk individuals will progress to diabetes and the tempo of  $\beta$  cell destruction will be crucial to select subjects for future primary and secondary prevention trials (4–6). Data from T cell readouts, metabolomic profiles, serum analytes, and gene expression have the potential to contribute novel insights into mechanisms underlying disease progression (4, 7–11). In particular, gene expression profiling provides a global picture of cellular events under different physiological conditions. However, few studies have measured temporal changes in gene expression in individuals at risk of T1D. The German BABYDIET (12) and the Finnish T1D prediction and prevention (DIPP) (13) studies have collected longitudinal gene expression profiles in children at risk of T1D, making this approach possible.

**Authorship note:** KALC, MH, and RT contributed equally to this work as co-senior author.

**Conflict of interest:** The authors have declared that no conflict of interest exists

**Submitted:** October 19, 2017

**Accepted:** January 25, 2018

**Published:** March 8, 2018

**Reference information:**

*JCI Insight*. 2018;3(5):e98212. <https://doi.org/10.1172/jci.insight.98212>.

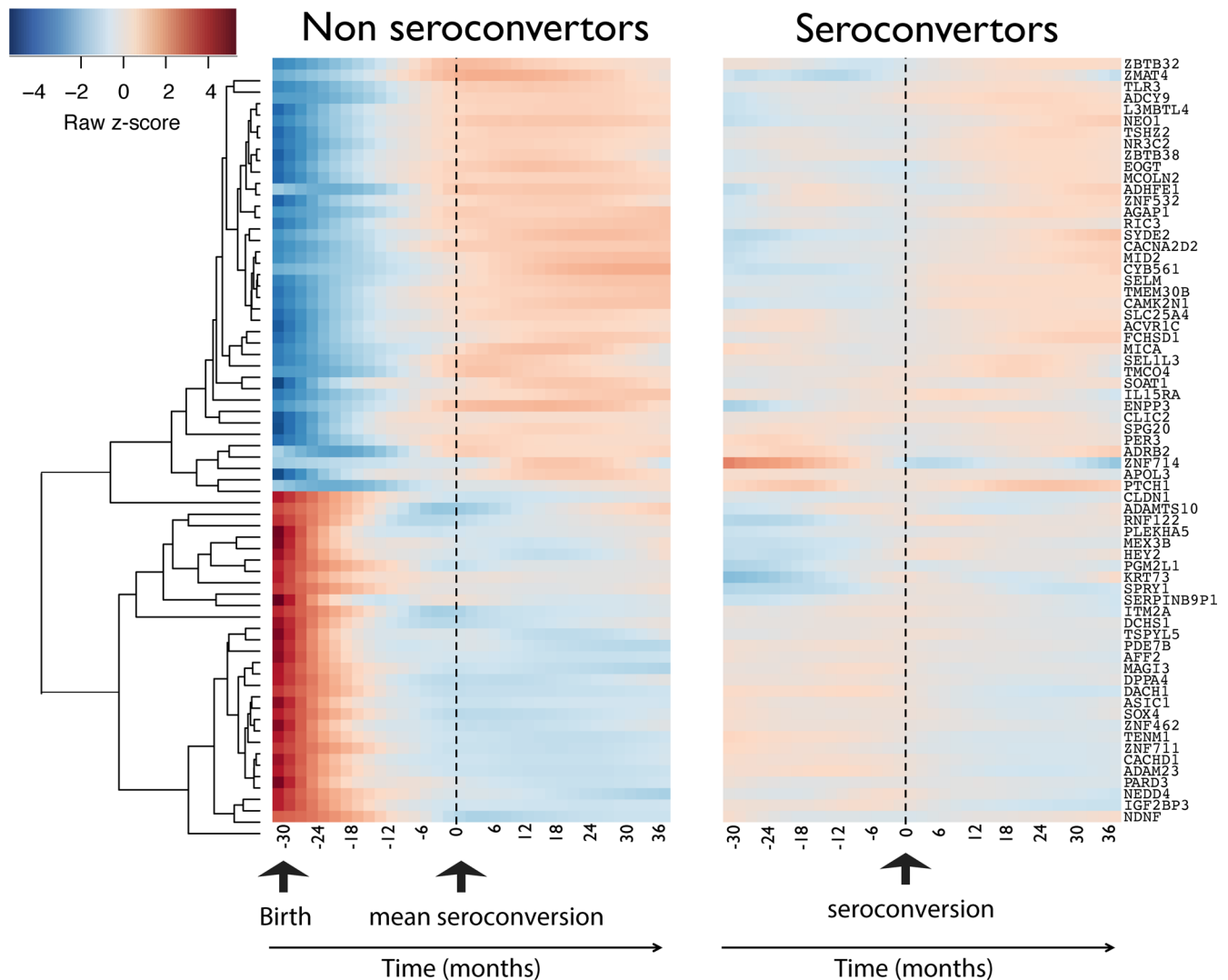
Conventional differential gene expression analysis either categorizes data into different groups or compares gene expression at a specific time point. The former method ignores changes over time within each category, whereas the latter approach assumes that gene expression in subjects of a similar age will be comparable. Because individuals at risk of T1D develop autoimmunity and insulin deficiency at different rates, the timing of seroconversion and progression to T1D varies. Previously, when longitudinal microarray data from BABYDIET were interrogated for differential expression of predefined IFN genes, expression of type 1 IFN genes was transiently increased before seroconversion (12). In the DIPP study, the longitudinal data were aligned relative to seroconversion and then divided into groups for pseudotemporal differential gene expression analyses in progressors to T1D, averaging all data in each group, and thus disregarding most of the time points (13). Nonetheless, in DIPP, differentially expressed (DE) genes associated with type I IFN were again identified in seropositive children progressing to T1D. However, potentially novel mechanistic insights may be gained by also considering the tempo of disease progression after correcting gene expression for an established marker of autoimmunity, such as seroconversion. In the current study, we combined longitudinal gene expression data available from 2 studies, BABYDIET and DIPP, combining children at high genetic risk of seroconversion and corrected the longitudinal gene expression data for time of seroconversion. Gene expression and clinical data were then integrated in a logistic regression model to identify DE genes and predict risk of seroconversion.

## Results

*Combination of longitudinal data and correction for time of seroconversion.* We combined gene expression data from 2 longitudinal cohorts of children at high genetic risk of T1D development. BABYDIET is a German dietary intervention trial that examined the effect of delayed gluten exposure on the development of islet autoimmunity in infants with at least 2 first-degree relatives with T1D, or infants with a high-risk HLA-DR/DQ haplotype. The Finnish DIPP study intensively monitored infants with HLA-DQ susceptibility to T1D from birth until the age of 15 years. In BABYDIET, 68% of data points for seroconvertors were sampled before seroconversion, while in DIPP, 83% of data points were sampled after seroconversion. Gene expression was lower overall in DIPP than in BABYDIET due to methodological differences, therefore leading to systematic variation between the 2 cohorts. Batch effect correction was performed as described in Supplemental Figure 1 (supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.98212DS1>) to combine these data sets, followed by further gene expression data correction relative to the time of seroconversion, with the antibody-negative group corrected relative to the time of mean seroconversion in the antibody-positive group. After data processing, 46% of data points were sampled before and 54% after seroconversion. To assess the impact of changing cell frequencies in whole blood over time, we used a computational deconvolution approach on the gene expression data using CIBERSORT (14). CIBERSORT was trained on whole blood data; therefore, we specifically investigated the cell-specific frequencies of seroconvertors and nonseroconvertors in the DIPP cohort, in which whole blood data were available. We observed that cell-specific frequencies remained stable over time and did not differ between nonseroconvertors and seroconvertors (Supplemental Figure 2).

*Differential gene expression followed by sensitivity analyses uncovered 67 DE genes.* We performed a differential gene expression analysis on all 17,000 genes with a linear model (see Methods) to identify genes with significant differential expression between groups (grp), or between groups over time (t:grp interaction). A total of 110 genes were declared statistically significant, representing either a grp and/or t:grp effect. To remove any bias in our analysis that may result from differences in clinical characteristics between the 2 groups, we performed sensitivity analyses, where we randomly permuted samples containing 110 genes with equal numbers of subjects matching either sex, HLA-risk status, or time points (Supplemental Table 1, using 10,000 permutations) and assessed their differential expression. Genes that remained significant more than 90% of the time after matching were selected (Supplemental Table 1), leading to a filtered list of 67 gene candidates (Supplemental Table 2).

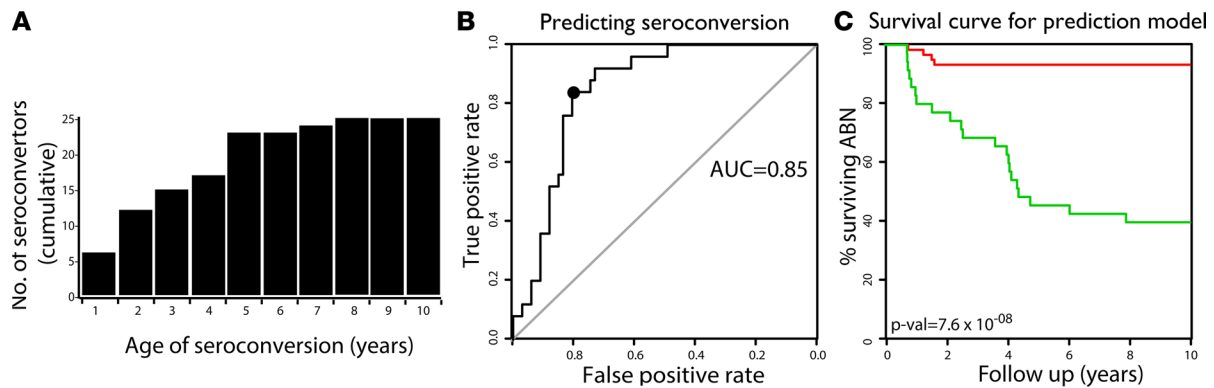
*Disruption of coordinated gene expression in seroconvertors.* Semisupervised hierarchical clustering of the 67 DE genes is shown in Figure 1, in which the longitudinal gene profiles were modeled using local polynomial regression models (see Methods). Patients who remained antibody negative (ABN) showed a coordinated switch in gene expression between approximately 6 and 18 months of age, which was not clearly evident in children who seroconverted. A number of genes had markedly different expression profiles in nonseroconvertors and seroconvertors during the first 2 years after birth. Some genes initially showed increased



**Figure 1. Integrating longitudinal gene expression data from BABYDIET and DIPP cohorts for differential gene expression analyses.** LOESS fitting–based heatmaps of nonseroconvertors (ABN) and seroconvertors (ABP) after differential gene expression analysis. The point “0 months” in nonseroconvertors represents the mean time of ABP seroconversion and, in seroconvertors, represents the actual time of ABP seroconversion.

expression in the nonseroconvertors relative to the seroconvertors (e.g., *CLDN1*, *ADAMTS10*, *RNF122*, *PLEKHA5*, *MEX3B*, *HEY2*, *PGM2L1*, *KRT73*, and *SPRY1*), while other genes were initially downregulated in the nonseroconvertors relative to the seroconvertors (e.g., *PTCH1*, *APOL3*, *ZNF714*, *ADRB2*, *PER3*, *IL15RA*, *MICA*, *ACVR1C*, *SLC25A4*). For other genes, although the pattern of gene expression over time was qualitatively similar in the 2 groups, the transition occurred later and sometimes after the development of autoantibodies in the seroconvertors (e.g., *NEO1*, *EOGT*, *MCOLN2*, *AGAP1*, *SYDE2*, *ZMAT4*, *DACHI*, *TENM1*, *PDE7B*). Overall, the data revealed a coordinated switch in gene expression before the average age of seroconversion in children who remained ABN, which was delayed or disrupted in children who developed autoantibodies. While 59 of 67 genes significantly differed with age in nonseroconvertors, only the neonatal developmental gene *NEO1* was significantly affected by age if seroconvertors were aligned relative to the time of mean seroconversion (Supplemental Table 3).

*Seven DE genes in the first year of life and HLA score predict seroconversion.* Given that the mean age of seroconversion was 2.59 years, we reasoned that DE genes close to birth might identify subjects most likely to develop autoimmunity. We included all subjects with gene expression data available in the first year and considered the time point closest to birth. Gene expression data were available during the first year for 25 children who did seroconvert and 67 children who did not seroconvert by the age of 10 (Figure 2A). A power analysis (15) for a survival analysis determined that such a sample size of 92 individuals in this data



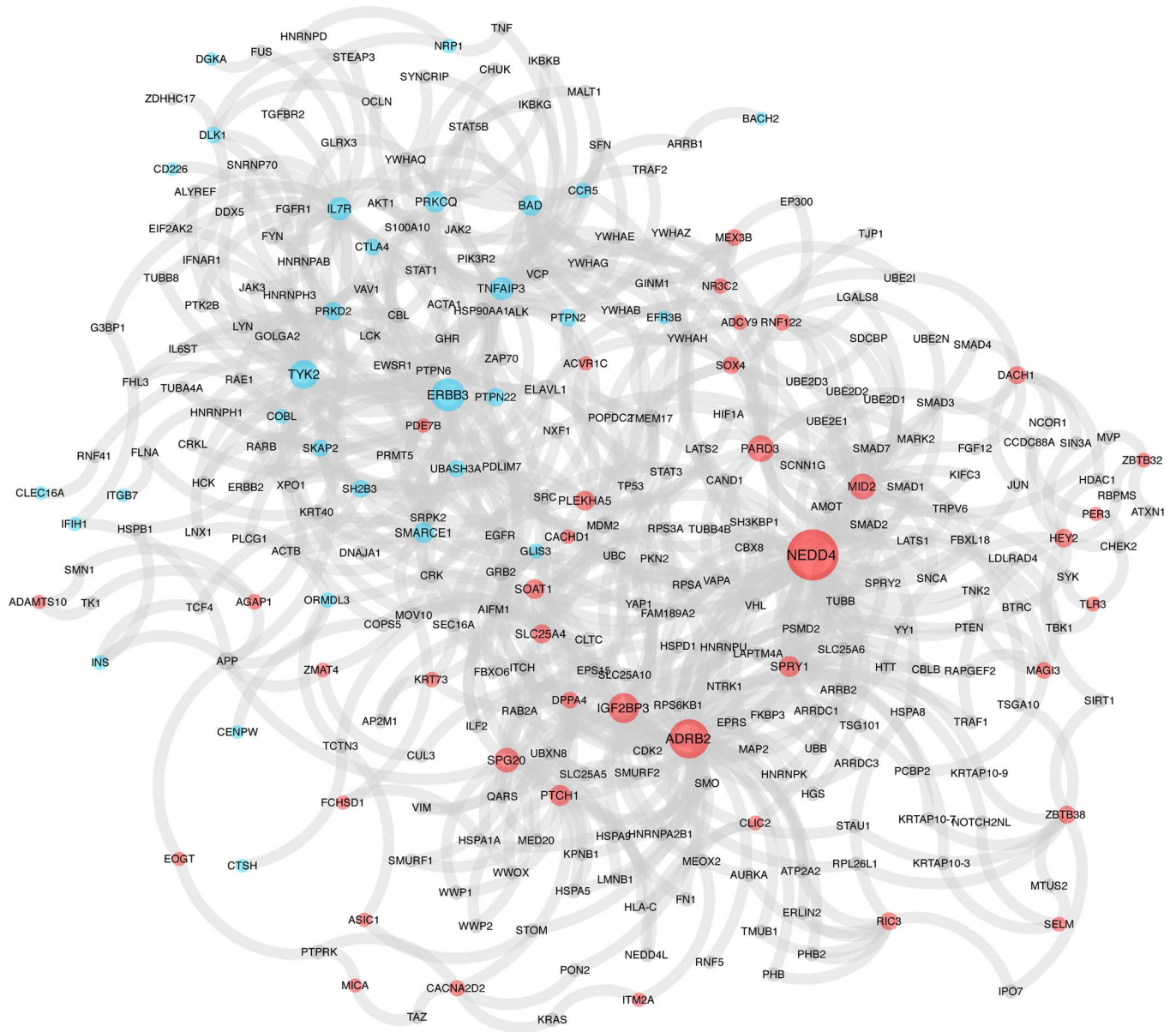
**Figure 2. Predicting seroconversion using at/near birth gene expression and clinical data.** (A) Number of seroconvertors with at least a single clinic visit by age 1 ( $n = 25$ ). A cumulative distribution of seroconversion is shown. Step-wise (backward) logistic regression models uncover (B) seroconversion features. Variables that predicted seroconversion were *ADCY9*, *PTCH1*, *MEX3B*, *IL15RA*, *ZNF714*, *TENM1*, *PLEKHA5*, and HLA score. Receiver operator characteristic curve (ROC), area under ROC (AUC), and Akaike information criterion (AIC) are shown for the prediction model. The shaded circle represents the probability threshold from the logistic regression model (0.233) for a sensitivity and specificity above 80%. (C) Survival curve for the prediction model with 34 subjects stratified as predicted seroconvertors (green) and 58 as predicted nonseroconvertors (red) using the same probability threshold.

set would reach a statistical power of 95% to detect DE genes at/near birth (significance level set to  $P < 0.05$ , Supplemental Figure 3). To determine whether near-birth DE genes could help predict future seroconversion, we applied a random forest algorithm to uncover the 20 most informative DE genes related to seroconversion prediction (Supplemental Figure 4). We manually added 3 clinical variables (age, sex, and HLA score), as these are known to influence the risk of seroconversion (2), and then applied a step-wise logistic regression model containing these 23 variables to identify features associated with the risk of seroconversion. A final logistic regression model with 8 features (*ADCY9*, *PTCH1*, *MEX3B*, *IL15RA*, *ZNF714*, *TENM1*, *PLEKHA5*, and HLA score) accurately predicted seroconversion (Figure 2B; AUC = 0.85, using leave-one-out cross validation). The logistic regression threshold (probability = 0.233, shaded circle in Figure 2B) yielded sensitivity and specificity of  $>80\%$ . We used this threshold to stratify subjects into predicted seroconvertors and predicted nonseroconvertors. A Kaplan-Meier curve (Figure 2C) demonstrated that the logistic regression model stratified risk of seroconversion ( $P = 7.6 \times 10^{-8}$ ). These results indicate that relative gene expression during the first year may be used to predict seroconversion by age 10.

Carriage of high risk HLA (DR3/4) and sharing of identical haplotypes with siblings has been shown to confer extreme risk of islet antibody development during childhood (16). Furthermore, testing children with high-risk HLA (17) multiple times during childhood identified individuals with multiple islet antibodies and high rate of progression to T1D (18). To determine the influence of HLA score on risk of seroconversion, we first assessed the contribution of the 7 genes predicting seroconversion at birth without HLA score information in a new logistic regression model. The 7 genes alone predicted seroconversion with an AUC of 0.81 (Supplemental Figure 5). In contrast, the contribution of HLA score only to predict seroconversion led to an AUC of 0.43. These results indicate that these 7 genes contribute significantly to the risk of seroconversion, while HLA score had a reduced but independent effect, thus improving the model's classification accuracy.

*The ubiquitin-proteasome pathway links DE genes in seroconvertors and T1D susceptibility genes.* To further understand the functional pathways associated with these genes, we built a protein-protein interaction (PPI) network using the BioGRID database (<https://thebiogrid.org/>) to identify seed proteins (DE genes) and their first-degree interacting proteins, which are connected to more than one seed protein. Forty of the 67 DE genes (60%) encoded proteins that shared at least 1 common neighbor, forming an interconnected PPI network (DE-PPI network, Figure 3). The functional enrichment analysis indicated that ubiquitin-dependent protein degradation was the most significantly enriched functional category in the DE-PPI network (Fisher's exact test,  $P$  value corrected with Bonferroni step down =  $4 \times 10^{-9}$ ; Supplemental Table 4). The main local hub in the DE-PPI network was NEDD4, which is an important enzyme in the ubiquitin-proteasome pathway. We then investigated whether the DE-PPI network is associated with T1D susceptibility genes previously identified in a genome-wide association study (T1D-GWAS) and publicly available at ImmunoBase (<http://www.immunobase.org>). We found that 31 T1D-GWAS

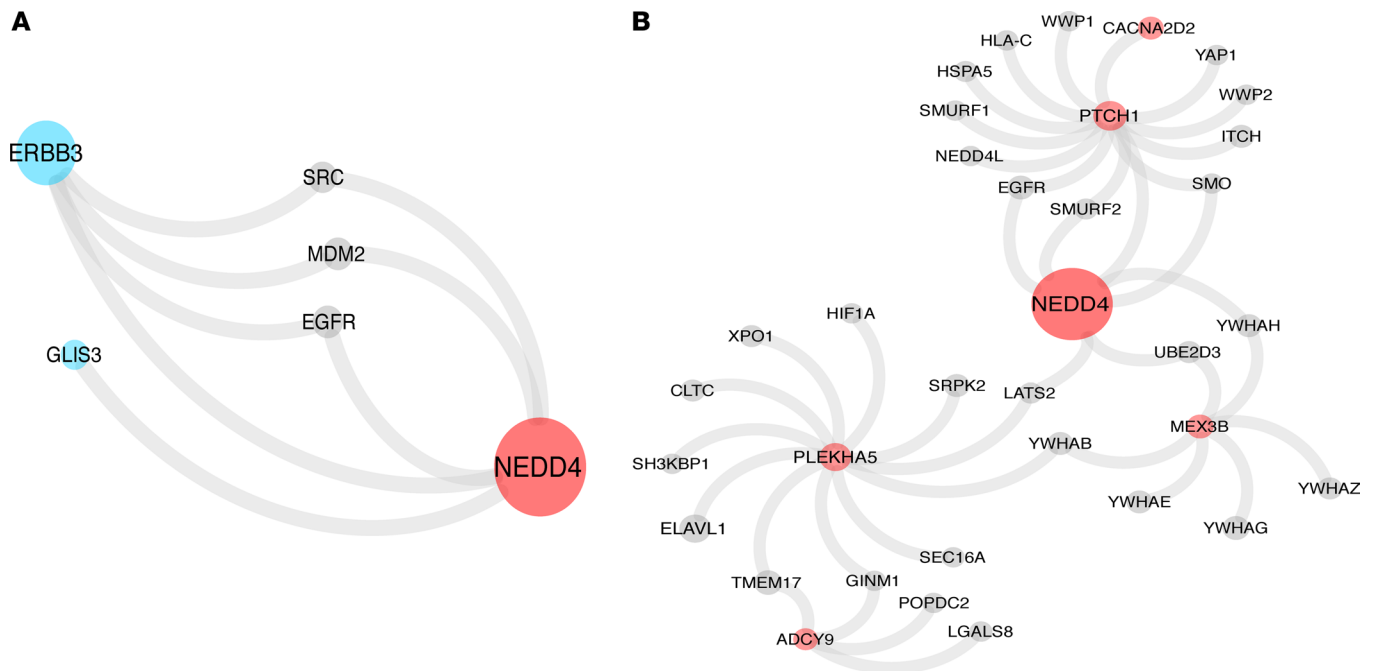




**Figure 3. Protein-protein interaction network linking the products of differentially expressed genes in blood with T1D susceptibility region candidate genes.** Interactions between T1D susceptibility region gene products obtained through GWAS studies (blue) and differentially expressed gene products (red) are shown.

genes formed another PPI network (T1D-GWAS PPI network; Figure 3). Remarkably, NEDD4 appeared to be a central hub linking T1D susceptibility genes and DE genes, directly interacting with DE genes *PTCH1*, *PLEKHA5*, *MEX3B*, and *ADCY9* (Figure 4B), and 2 T1D susceptibility genes, *ERBB3* and *GLIS3* (Figure 4A). These data suggest an important role for NEDD4 regulation of the ubiquitin-proteasome pathway during T1D development (19).

*Immune enrichment analysis.* We undertook gene set enrichment analysis (GSEA) of the 67 DE genes, using the Broad Institute’s collection of immunological signatures (20, 21). Each gene set module was manually annotated with the pathway in which these genes are involved. Then, pathway enrichment sorting based on the number of gene set modules was performed (Table 1 and Supplemental Table 5). More than half of these modules were related to DC and T cell function. The top hits included genes involved in T cell activation (containing 30 gene set modules and involving 43 DE genes), naive CD8 T cell development and function (17 gene set modules and 26 DE genes), and thymocyte development (16 gene set modules and 31 DE genes). Genes involved in TLR signaling in DCs and naive B cell genes (containing 16 gene set modules



**Figure 4. Protein-protein interaction network linking DE genes with T1D susceptibility region candidate genes identifies NEDD4 as a central hub. (A)** NEDD4 directly interacts with 2 T1D susceptibility genes, *ERBB3* and *GLIS3*. **(B)** NEDD4 serves as a central hub for 4 (*PTCH1*, *PLEKHA5*, *MEX3B*, *ADCY9*) of the 7 genes that predict seroconversion.

involving 26 genes and 9 gene set modules involving 14 DE genes, respectively) — and several classical inflammatory mediator pathways, including the type I and type II IFN responses, the viral RNA response, and the pathogen response — were also enriched.

*Connectivity maps to discover drugs with the potential to prevent seroconversion.* Connectivity maps can establish functional relationships between genes, drugs, and diseases, using reference gene-expression profiles from human cancer cell lines treated with US Food and Drug Administration–approved (FDA-approved) drugs (22, 23). We examined drugs that could potentially promote gene expression signatures similar to those seen in children who remained seronegative. We therefore used the LOESS-fitted, near-birth expression of DE genes of nonseroconvertors as the ideal reference profile. Drugs shown in Table 2 induced an expression signature of DE genes correlated with nonseroconversion, with an FDR  $\leq 0.01$ . Some drugs induced expression changes in multiple genes; for example, the COX-2 inhibitor Methyl [5-methylsulfonyl-1-(4-chlorobenzyl)-1H-2-indolyl]carboxylate altered the expression of 10 DE genes. Notably, a number of drugs (marked with an asterisk) have previously been predicted to prevent or delay T1D in NOD mice and human subjects (22). Others may represent novel potential therapeutics in subjects at risk of T1D.

## Discussion

Inherited factors are important determinants of an individual's risk of developing T1D. However, the fact that >90% of children with an increased genetic risk remain ABN highlights the crucial role of gene-environment interactions. Very few studies have measured temporal changes in gene expression in individuals at risk of T1D. In general, differential gene expression analyses have not corrected for the timing of seroconversion and progression to T1D for each subject individually (24). Additionally, no attempts have been made to integrate large-scale gene expression data from cohorts of children at risk for T1D, which may help to uncover broadly applicable genetic signatures. By comparing longitudinal gene expression during the first 2 years in at-risk children, we identified genetic signatures that predict seroconversion and may provide important mechanistic insights. In particular ubiquitin-dependent protein degradation, during a critical period of development, appears to influence the onset of islet autoimmunity.

The mean age of seroconversion for the 2 combined cohorts included in this study was 2.59 years. Gene expression signatures associated with seroconversion were evident during the first 2 years of life, supporting the hypothesis that factors contributing to the onset of autoimmunity in the course of T1D development

**Table 1. GSEA to uncover pathways**

GSEA pathway	Number of gene set modules	Number of genes involved	$P_{adj}$ (false discovery rate)	Fold enrichment
T cell activation	30	43	3.317e-09	2.379
Naive CD8 T cell genes	17	26	$1.08 \times 10^{-6}$	2.912
Thymocyte development	16	31	$4.853 \times 10^{-8}$	2.906
TLR signaling in DCs	16	26	$1.318 \times 10^{-5}$	2.543
Pathogen response	15	32	$7.284 \times 10^{-9}$	3.089
Genes down regulated in Foxp3 <sup>+</sup> T cells	15	28	$1.001 \times 10^{-7}$	3.091
Macrophage differentiation	10	24	$4.72 \times 10^{-7}$	3.314
Memory CD4 T cell genes	10	23	$5.823 \times 10^{-7}$	3.394
Naive B cell genes	9	14	0.001965	2.657
Type I IFN response	9	25	$3.317 \times 10^{-9}$	4.29
B cell lineage	8	23	$8.672 \times 10^{-9}$	4.379
Naive CD4 T cell genes	7	13	0.001038	3.033
Memory CD8 T cell genes	7	15	0.0006548	2.901
Viral response	7	16	0.0002756	2.997
TLR signaling in macrophages	6	15	0.0001772	3.301
CD4 T cell lineage	6	12	0.002486	2.854
IFN- $\gamma$ response	5	14	$4.024 \times 10^{-5}$	4.028
IL-4-induced genes	5	15	$1.63 \times 10^{-5}$	4.076

GSEA pathways were manually annotated to perform enrichment analysis. Pathways with  $\geq 5$  gene set modules are shown.

are operating during infancy. Indeed, the coordinated transition of DE genes during the first year of life in children who remained seronegative appeared to be disrupted in children who seroconverted. Not surprisingly, many of the DE genes were associated with innate or adaptive immune function. Zinc finger and BTB domain containing 32 (*ZBTB32*), which is highly expressed in lymph nodes and regulates T helper cell activation, was upregulated in the blood of seroconvertors during the first year. *TLR3* and IL-15 receptor subunit  $\alpha$  (*IL15RA*), which increase NF- $\kappa$ B signaling-mediated immune activation and promote memory T cell differentiation, respectively, were downregulated in blood of nonseroconvertors during the first year. A recent temporal study of antigen-responsive CD4<sup>+</sup> T cells in children from the BABYDIET cohort also identified the development of T cell memory concomitant with autoantibody development (25). Patched 1 (*PTCH1*), which was upregulated in blood of seroconverters over time, is required for development of the T and B lymphoid lineages (26). Some of the DE genes are known to be important coordinators of diverse developmental processes (e.g., Sprouty RTK signaling antagonist 1 [*SPRY1*]), and although likely to have important functions in circulating immune cell populations, their specific functions are still to be elucidated (25). Adrenoreceptor  $\beta$  2 (*ADRB2*) and adenylate cyclase 9 (*ADCY9*) gene expression data further suggest that  $\beta$  adrenergic signaling pathways may also differentiate children who are likely to seroconvert.

Ubiquitin and ubiquitin-like modifications are increasingly recognized as key regulatory events in health and disease (27). The DE gene signature and PPI network suggests that altered ubiquitination may play an important role in determining the development of autoimmunity. It has previously been reported that ubiquitination modulates NF- $\kappa$ B signaling, which in turn determines DC differentiation and maturation (28, 29). Indeed, increased levels of NF- $\kappa$ B DNA binding in DCs has been shown in both subjects with T1D and nonobese diabetic mice, highlighting the importance of NF- $\kappa$ B signaling (30, 31). NEDD4 is an E3 ubiquitin ligase that was found to promote activation of naive T cells in mice by degrading Cbl-b (32). In muscle cells, NEDD4 suppressed Notch1 (33), which facilitates TGF- $\beta$ -mediated Treg function (34, 35). In this regard, it is of interest that *NEDD4* was differentially overexpressed in blood around the time of seroconversion in children developing islet autoantibodies. The transmembrane protein neural precursor cell expressed NEDD4 family interacting protein 1 (NDFIP1) binds and activates NEDD4 family E3 ubiquitin ligases (36). In mice, silencing Ndfip1 inhibited cytokine-mediated  $\beta$  cell apoptosis by reducing ER stress (37), again highlighting the importance of NEDD4. The specific importance of altered ubiquitination in T1D is highlighted by a recent comparison of gene expression signatures in 6 different autoimmune diseases.

**Table 2. Connectivity map of DE genes predicted to correlate with expression in non seroconvertors after exposure to drugs**

Set	$P_{adj}$	Compound	Dose (M)	Genes
83	$2.4 \times 10^{-12}$	monorden	$1.5 \times 10^{-7}$	<i>TMEM30B, ZBTB38, NEO1, CAMK2N1, SOAT1</i>
125	$2.3 \times 10^{-9}$	<sup>A</sup> troglitazone	$1.0 \times 10^{-5}$	<i>SPG20, NEO1, PER3</i>
102	$5.5 \times 10^{-8}$	(R)-(+)-trans-N-(4-Pyridyl)-4-(1-aminoethyl)-cyclohexanecarboxamide	$3.0 \times 10^{-6}$	<i>SPG20, PER3</i>
233	$7.0 \times 10^{-6}$	<sup>A</sup> Methyl [5-methylsulfonyl-1-(4-chlorobenzyl)-1H-2-indolyl]carboxylate	$1.0 \times 10^{-5}$	<i>IGF2BP3, TSHZ2, SPG20, AFF2, ASIC1, NDNF, AMAT4, RNF122, DCHS1, ITM2A</i>
98	$1.2 \times 10^{-5}$	prazosin	$1.0 \times 10^{-5}$	<i>NEO1, PER3</i>
100	$1.2 \times 10^{-6}$	pyrvinium	$1.3 \times 10^{-6}$	<i>NEO1, PER3</i>
130	$1.2 \times 10^{-5}$	tyrphostin AG-825	$2.5 \times 10^{-5}$	<i>NEO1, PER3</i>
53	$5.4 \times 10^{-5}$	demecolcine	$1.2 \times 10^{-5}$	<i>PDE7B, NEO1</i>
147	$9.6 \times 10^{-5}$	<sup>A</sup> wortmannin	$1.0 \times 10^{-6}$	<i>SPG20, NEO1, ENPP3, PER3</i>
103	$5.0 \times 10^{-4}$	resveratrol	$1.0 \times 10^{-5}$	<i>PER3</i>
131	$5.0 \times 10^{-4}$	<sup>A</sup> valproic acid	$5.0 \times 10^{-5}$	<i>PER3</i>
1	$6.5 \times 10^{-4}$	(-)-catechin	$1.1 \times 10^{-5}$	<i>PDE7B, NEO1, ENPP3, PER3</i>
22	$6.5 \times 10^{-4}$	2-benzoyl-1,4-phenylene bis(4-bromo-3-nitrobenzoate)	$7.0 \times 10^{-6}$	<i>PDE7B, NEO1, ENPP3, PER3</i>
115	$9.5 \times 10^{-4}$	topiramate	$3.0 \times 10^{-6}$	<i>ZNF532</i>
117	$9.7 \times 10^{-4}$	tretinoin	$1.0 \times 10^{-6}$	<i>NR3C2</i>
166	$1.4 \times 10^{-3}$	<sup>A</sup> sodium phenylbutyrate	$1.0 \times 10^{-4}$	<i>SOX4</i>
119	$1.9 \times 10^{-3}$	<sup>A</sup> trichostatin A	$1.0 \times 10^{-7}$	<i>APOL2, PDE7B, PER3, ITM2A</i>
23	$2.0 \times 10^{-3}$	2-chloro-N-(2-nitrophenyl)-5-(1-piperidinylsulfonyl)benzamide	$1.2 \times 10^{-5}$	<i>SPRY1, SPG20, PDE7B, NEO1, ENPP3, PER3</i>
124	$2.2 \times 10^{-3}$	trifluoperazine	$1.0 \times 10^{-5}$	<i>ITM2A</i>
24	$2.5 \times 10^{-3}$	2-hydroxy-3-methoxybenzaldehyde 2-pyridinylhydrazone	$2.5 \times 10^{-5}$	<i>PDE7B</i>
47	$2.5 \times 10^{-3}$	clotrimazole	$5.0 \times 10^{-5}$	<i>PDE7B</i>
30	$5.8 \times 10^{-3}$	4-nitrophenyl (1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)acetate	$1.5 \times 10^{-5}$	<i>PDE7B, NEO1, ENPP3, RIC3, CLDN1, CACNA2D2</i>
29	$1.2 \times 10^{-2}$	4-methyl-2-oxo-2H-chromen-7-yl ([3-(2,3-dihydro-1,4-benzodioxin-6-yl)-2-methyl-4-oxo4H-chromen-7-yl]oxy)acetate	$1.0 \times 10^{-5}$	<i>NEO1</i>

Compounds marked <sup>A</sup> were previously predicted to be associated with T1D.

Although some gene expression signatures were common to all diseases, alterations in the ubiquitination pathway were only evident in subjects with T1D (38). Such observations would require extensive further experimental validation in children at risk of progressing to T1D.

An analysis of the immune gene set modules associated with the DE genes showed that some are involved in B cell-related activities. This is not surprising, considering that our analysis focused on seroconversion to islet antibody positivity. We found that expression of the SOX4 transcription factor, which was associated with early onset of autoimmunity, was enriched in 50 immune gene sets. SOX4 belongs to a group of important transcriptional regulators involved in a number of developmental processes, including formation of the endocrine pancreas (39). Mutations in SOX4 resulted in reduced insulin secretion and impaired glucose tolerance (40). Furthermore, expression of SOX4 in pro-B cells was linked to their survival (41).

The ability to identify which at-risk children are unlikely to develop autoimmunity would be of significant benefit to families, obviating the need for regular screening. In addition, information related to gene expression in infancy would allow improved selection of at-risk subjects for future intervention trials and provide a useful endpoint to assess therapeutic responses. A linear regression model based on near-birth expression of the 67 DE genes identified 7 genes (*ADCY9, PTCH1, MEX38, IL15RA, ZNF714, TENM1*, and *PLEKHA5*) and HLA score as predictive of seroconversion within the first 10 years, with an AUC of 0.85 and AIC of 79.6. In practice, T1D-associated HLA high-risk genes could be used in a prospective study to preselect children for further screening based on a combination of the DE genes identified.

Preventing autoimmunity in high-risk infants remains a considerable challenge, and improved prediction models would represent an important step forward. The connectivity map provides a strategy



to select approved therapeutic agents predicted to promote gene expression signatures resembling those in infants who remain seronegative. Such agents could be repurposed or used to develop new therapeutic strategies for future prevention of diabetes in high-risk children. Twenty-three potentially novel drugs were identified, and the capacity of drugs such as trichostatin A and sodium phenylbutyrate to modulate autoimmunity, diabetes onset, or insulin metabolism has previously been suggested (22, 42, 43). The top hit was monorden (radicol), which inhibits heat shock protein 90 (Hsp90), a molecular chaperone responsible for polypeptide maturation, rematuration, and folding of denatured proteins (44). Hsp90 is upregulated in response to pancreatic  $\beta$  cell stress in T1D (45, 46). Inhibitors of Hsp90 promote ubiquitin-proteasome-mediated degradation of their protein substrates, block NF- $\kappa$ B activity, and interfere with MHC class II-mediated islet antigen presentation (44, 47, 48). Sodium phenylbutyrate has also been shown to reduce ER stress and to alleviate insulin resistance and  $\beta$  cell dysfunction in humans (43). The current data provide a framework for further validation and therapeutic mapping toward primary prevention trials.

The sophisticated statistical analyses that we performed to combine these longitudinal gene expression studies from 2 separate cohorts may raise some caveats. Batch effect correction adjusted for differences in sample collection, processing, gene expression assessment, and cohort variability, in addition to time correction. Our sensitivity analyses suggest that those corrections were effective; however, further study will be needed to fully assess some potential unknown confounders that may remain in our analyses. Furthermore, we focused on seroconversion, as defined by the appearance of a single autoantibody without information on antibody specificity or progression to multiple antibody positivity. Notably, 34 of the 50 children who developed a single serum autoantibody were later diagnosed with T1D; thus, it would be of interest to study whether a similar DE gene profile and HLA risk score near birth also predicts seropositivity for multiple autoantibodies. There are also limitations associated with using a bioinformatics approach to assess gene expression data. For genes with similar functions, posttranscriptional regulation may affect protein expression and function, particularly since the ubiquitin-proteasome pathway is implicated (49–52). Conversely, genes with similar expression profiles may not have similar functions (53).

Previous studies on cord and peripheral blood of infants, together with the current results, suggest that the autoimmune trajectory to T1D is set very early in life, associated with fundamental alterations in immune function and metabolism. This observation has profound implications. The gene expression signature in early infancy predicts risk of seroconversion and could be used to stratify subjects for surveillance or clinical trials. Moreover, the signature suggests that ubiquitin-dependent protein degradation may play an important role in the development of islet autoimmunity. Finally, the identified gene expression pattern is potentially correctable, and the gene expression signature could be used to monitor the impact of environmental influences on mothers and babies or the efficacy of therapeutic interventions in clinical trials.

## Methods

### Study design and patient characteristics

The longitudinal microarray data of peripheral blood mononuclear cells (PBMC) from 109 children in the BABYDIET cohort (454 longitudinal microarray experiments) were taken from ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1724/>). The clinical features used were age, sex, HLA status, whether a sibling had T1D, and time of seroconversion. The samples were divided into ABP ( $n = 22$ ) and ABN ( $n = 87$ ) groups. Twelve of the 22 ABP infants progressed to T1D. Further details of this study including PBMC preparation, RNA isolation, cDNA synthesis, and microarray hybridization are available as published (12).

Longitudinal whole blood microarray data are available for 58 children in the DIPP cohort (356 longitudinal microarray experiments) accessible through Gene Expression Omnibus (GEO) SuperSeries access number GSE30211 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30211>). The clinical features used in this research were age, sex, HLA status, whether a sibling had T1D, and time of seroconversion. Twenty-eight subjects who seroconverted (22 progressed to T1D) were matched (date and place of birth, sex, HLA-DQ genotype) with 30 ABN subjects. Further experimental details including whole blood preparation and microarray hybridization are available as published (13). We performed computational deconvolution using CIBERSORT (<https://cibersort.stanford.edu/>) to access the impact of changing cell frequencies in whole blood from the DIPP cohort.

Combined BABYDIET and DIPP cohorts included 92 subjects (25 seroconvertors and 67 nonseroconvertors) with data available from at least a single microarray data during the first year of life. The power analysis on the first year microarray data was performed using *samr* package in R (15) to establish the statistical power required to uncover DE genes at or near birth.

We classified subjects as high risk (HLA score = 3) if 1 allele was HLA-DR3 and the other DR4, or — in the absence of HLA-DR typing — carriage of both HLA-DQB1\*0201 and \*0302. Subjects were classified medium risk (HLA score = 2) if he or she were HLA-DR3 or DR4 homozygous, 1 allele was HLA-DR-3/4 and the other HLA-DRX, or a single allele was HLA-DQB1\*0201 or \*0302. All remaining subjects were classified low risk (HLA score = 1).

## Statistics

*Microarray batch correction and longitudinal alignment relative to seroconversion.* Affymetrix microarray data from DIPP and BABYDIET cohorts were combined to increase the statistical power for data analyses. The cohort effect was corrected using the *removebatchEffect* function in Limma package in R. Longitudinal microarray data were aligned relative to seroconversion. Specifically, for each sample, age at seroconversion was subtracted from age at RNA isolation. For the ABN group, age of mean seroconversion (in ABP) was subtracted from the age of RNA isolation.

## Computational analysis of gene clusters, functional pathways, and PPI networks

*Data mining.* The PPI network was constructed using DE genes as “seeds” and their first-degree-interacting neighbors. The PPI data were retrieved from the BioGRID database (version 3.4.136) (54). For statistical analysis of PPI networks, we used a published computational method (55) to measure 3 structural properties of PPI networks: average shortest path length, clustering coefficient, and density (normalized average degree). We tested whether the distributions of these structural properties were similar between DE-PPI and random PPI networks. Nonparametric Kolmogorov-Smirnov test was used to determine statistical significance of structural differences between DE-PPI networks and random PPI networks (null hypothesis). Network visualization and annotation were performed using Cytoscape (56). ClueGo (57) was used to find functional pathways enriched in the DE-PPI network based on the Gene Ontology database (58).

We also built a PPI network for T1D susceptibility genes previously identified in a GWAS and taken from the immunobase database (<https://www.immunobase.org/>). For DE analysis during the first 5 years of life, we fitted a linear model as a function of cohort, time, group, and time/group interactions as shown in the equation below. A window of 10 years was used to perform differential gene expression analysis after time correction. A total of 147 subjects fell into this window:  $DE_i = t + grp + t:grp + cohort$ , where  $DE_i$  represents differential expression of  $i^{th}$  gene,  $t$  represents the corrected time,  $grp$  is the data groups seroconvertors (ABP) and nonseroconvertors (ABN),  $t:grp$  represents time and group interactions, and  $cohort$  represents the cohort effect. We calculated a  $P$  value using F-statistics and further examined the FDR using the Benjamini-Hochberg correction. For our analyses, we filtered all genes with corrected  $P$  values for  $grp$  or  $t:grp \leq 0.05$  and  $P$  values for  $cohort > 0.05$  to exclude a cohort effect. The linear model, including significance tests, was performed using R (<http://cran.r-project.org/>).

The number and timing of microarray samples were not consistent for each individual subject. Some subjects in the cohort had a single time point measurement, whereas others had up to 12 time point measurements. To visualize gene expression profiles over time during the first 5 years, we used local polynomial regression models (LOESS) to obtain fitted values of DE genes for subjects over time. Such models help guard against deviant points due to oversampling and summarize and extrapolate time points across all individuals (3). We then generated heatmaps using the *heatmap.2* function available in *gplots* library of R. The heatmap shows dendrograms by performing semisupervised hierarchical clustering, whereby gene expression on the  $y$  axis was clustered according to the Euclidian distance and Ward linkage, and time on the  $x$  axis was ordered. The heatmap visualizes the change of gene expression data relative to time of seroconversion.

## GSEA of immunological signatures

To interpret gene expression data, the Broad Institute’s GSEA was used to perform GSEA of DE genes (20, 21). We used molecular signature database (MSigDB), collection C7 (immunologic gene sets) to find gene set modules expressing our DE genes. We downloaded the MSigDB-C7 dataset from

Broad Institute GSEA ([http://software.broadinstitute.org/gsea/msigdb/download\\_file.jsp?filePath=/resources/msigdb/6.0/c7.all.v6.0.entrez.gmt](http://software.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/msigdb/6.0/c7.all.v6.0.entrez.gmt)) and wrote R script to perform gene set enrichment analysis and uncovered 270 gene set modules. The analyses to calculate statistical significance of gene set modules used in R script was described (20, 21) and detailed online ([http://software.broadinstitute.org/gsea/msigdb/help\\_annotations.jsp#overlap](http://software.broadinstitute.org/gsea/msigdb/help_annotations.jsp#overlap))

To further understand the most prevalent immunological pathways expressed by these gene set modules, we manually annotated each gene set module to its immunological pathway by linking the collective overrepresented function(s) of such genes and experimental perturbation that initiates such functions. For statistical analyses, we compared the number of genes that had a specific gene module annotation against those that did not. The hypergeometric test was used to establish a *P* value, which was further corrected using FDR. We calculated fold enrichment by taking the ratio of proportion of genes containing specific gene module annotation (GSEA pathway) in our DE gene list and the proportion of genes involved in GSEA pathway in a universal dataset. The universal dataset was established by taking the union of genes in all gene set modules expressing DE genes.

### Logistic regression model and survival analysis

We filtered subjects with microarray experiments performed before 1 year of age, as we were specifically interested in using single time point gene expression data close to birth. Therefore, if there were multiple measurements available for a single subject, we chose the microarray data closest to birth to obtain single time point data.

We used the random forest algorithm (randomForest package in R) and the mean decrease in the Gini Index as a metric to prefilter the 20 most informative features predicting seroconversion. We applied step-wise backward elimination logistic regression modeling to predict seroconversion from DE genes and clinical variables (age, sex, and HLA score). The Akaike information criterion (AIC) was calculated to assess the quality of fit of the logistic regression model. We also reported the area under receiver operator characteristic curve for the model using a leave-one-out cross validation.

All subjects were followed up for 10 years to determine time of seroconversion. A logistic regression model outputs a probability as an outcome for each individual. We determined an optimal logistic regression probability threshold that gives sensitivity and specificity above 80%. By using the optimal logistic regression threshold, we stratified the subjects into 2 survival subgroups: predicted seroconvertors and predicted nonseroconvertors. Comparisons of time of seroconversion and associations with predicted seroconvertors and predicted nonseroconvertors were analyzed using a Kaplan–Meier log-rank test and a  $\chi^2$  test using the survival R package. All survival analyses were performed in R.

### Connectivity mapping

To develop connectivity maps, the reference gene expression data from cancer cell lines treated with FDA-approved drugs were taken from the ArrayExpress database with ID: GEOD5258. Using the gCMAPWeb package in R and polynomial regression fitted gene expression values near birth, we found compounds producing similar up/downregulation of DE genes in the cell lines to that observed in the ABN group.

### Data and materials availability

BABYDIET cohort microarray data is available at ArrayExpress database (accession number E-MTAB-1724). DIPP cohort microarray data is available at GEO database (accession number GSE30211).

### Study approval

Approval for analysis of data from these studies was obtained from the University of Queensland and Metro South Human Research Ethics Committees. Subjects provided informed consent prior to their participation in the original studies from which the data were derived.

### Author contributions

Concept and design of the study contributed by AMM, MH, and RT. Data collection, data analysis, and interpretation contributed by AMM, KALC, AC, EEHW, AZ, EB, and RT. Manuscript preparation contributed by AMM, MH, and RT.

## Acknowledgments

Supported by NHMRC grant 1071822. AM was supported by a Postdoctoral Fellowship from JDRF (3-PDF-2016-198-A-N), K-ALC by a Career Development Fellowship from NHMRC, and RT by Arthritis Queensland and a Fellowship from NHMRC.

Address correspondence to: Ranjeny Thomas, The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Translational Research Institute, 37 Kent Street, Woolloongabba, QLD, 4171, Australia. Phone: 617.34436960; Email: ranjeny.thomas@uq.edu.au.

KALC's present address is: Melbourne Integrative Genomics, University of Melbourne, Parkville, Australia.

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383(9911):69–82.
2. Ziegler AG, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*. 2013;309(23):2473–2479.
3. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*. 1979;74(368):829–836.
4. Lipman TH, et al. Increasing incidence of type 1 diabetes in youth: twenty years of the Philadelphia Pediatric Diabetes Registry. *Diabetes Care*. 2013;36(6):1597–1603.
5. Imperatore G, et al. Projections of type 1 and type 2 diabetes burden in the U.S. population aged <20 years through 2050: dynamic modeling of incidence, mortality, and population growth. *Diabetes Care*. 2012;35(12):2515–2520.
6. Todd JA. Genetic analysis of type 1 diabetes using whole genome approaches. *Proc Natl Acad Sci USA*. 1995;92(19):8560–8565.
7. Mäkinen VP, et al. Diagnosing diabetic nephropathy by 1H NMR metabolomics of serum. *MAGMA*. 2006;19(6):281–296.
8. Mäkinen VP, et al. 1H NMR metabolomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol*. 2008;4:167.
9. Martinuzzi E, Lemonnier FA, Boitard C, Mallone R. Measurement of CD8 T cell responses in human type 1 diabetes. *Ann N Y Acad Sci*. 2008;1150:61–67.
10. Tsai S, Shameli A, Santamaria P. CD8+ T cells in type 1 diabetes. *Adv Immunol*. 2008;100:79–124.
11. van Belle TL, Coppieters KT, von Herrath MG. Type 1 diabetes: etiology, immunology, and therapeutic strategies. *Physiol Rev*. 2011;91(1):79–118.
12. Ferreira RC, et al. A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes*. 2014;63(7):2538–2550.
13. Kallionpää H, et al. Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. *Diabetes*. 2014;63(7):2402–2414.
14. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457.
15. Tibshirani R. A simple method for assessing sample sizes in microarray experiments. *BMC Bioinformatics*. 2006;7:106.
16. Baschal EE, Eisenbarth GS. Extreme genetic risk for type 1A diabetes in the post-genome era. *J Autoimmun*. 2008;31(1):1–6.
17. Hagopian WA, et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. *Pediatr Diabetes*. 2011;12(8):733–743.
18. Insel RA, Dunne JL, Ziegler AG. General population screening for type 1 diabetes: has its time come? *Curr Opin Endocrinol Diabetes Obes*. 2015;22(4):270–276.
19. Wertz IE, Dixit VM. Signaling to NF-kappaB: regulation by ubiquitination. *Cold Spring Harb Perspect Biol*. 2010;2(3):a003350.
20. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–15550.
21. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267–273.
22. Pang D, et al. Expression profiling pre-diabetic mice to uncover drugs with clinical application to type 1 diabetes. *Clin Transl Immunology*. 2015;4(8):e41.
23. Lamb J, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929–1935.
24. Jin Y, et al. 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(7):2506–2515.
25. Heninger AK, et al. A divergent population of autoantigen-responsive CD4+T cells in infants prior to  $\beta$  cell autoimmunity. *Sci Transl Med*. 2017;9(378):eaaf8848.
26. Uhmman A, et al. The Hedgehog receptor Patched controls lymphoid lineage commitment. *Blood*. 2007;110(6):1814–1823.
27. Balasubramanyam M, Sampathkumar R, Mohan V. Is insulin signaling molecules misguided in diabetes for ubiquitin-proteasome mediated degradation? *Mol Cell Biochem*. 2005;275(1-2):117–125.
28. O'Sullivan BJ, Thomas R. CD40 ligation conditions dendritic cell antigen-presenting function through sustained activation of NF-kappaB. *J Immunol*. 2002;168(11):5491–5498.
29. Chen J, Chen ZJ. Regulation of NF-kB by ubiquitination. *Curr Opin Immunol*. 2013;25(1):4–12.
30. Mollah ZU, et al. Abnormal NF-kappa B function characterizes human type 1 diabetes dendritic cells and monocytes. *J Immunol*. 2008;180(5):3166–3175.
31. Poligone B, Weaver DJ, Sen P, Baldwin AS, Tisch R. Elevated NF-kappaB activation in nonobese diabetic mouse dendritic cells results in enhanced APC function. *J Immunol*. 2002;168(1):188–196.
32. Yang B, et al. Nedd4 augments the adaptive immune response by promoting ubiquitin-mediated degradation of Cbl-b in activated T cells. *Nat Immunol*. 2008;9(12):1356–1363.

33. Koncarevic A, Jackman RW, Kandarian SC. The ubiquitin-protein ligase Nedd4 targets Notch1 in skeletal muscle and distinguishes the subset of atrophies caused by reduced muscle tension. *FASEB J*. 2007;21(2):427–437.
34. Marcel N, Sarin A. Notch1 regulated autophagy controls survival and suppressor activity of activated murine T-regulatory cells. *Elife*. 2016;e14023.
35. Ostroukhova M, Qi Z, Oriss TB, Dixon-McCarthy B, Ray P, Ray A. Treg-mediated immunosuppression involves activation of the Notch-HES1 axis by membrane-bound TGF-beta. *J Clin Invest*. 2006;116(4):996–1004.
36. Harvey KF, Shearwin-Whyatt LM, Fotia A, Parton RG, Kumar S. N4WBP5, a potential target for ubiquitination by the Nedd4 family of proteins, is a novel Golgi-associated protein. *J Biol Chem*. 2002;277(11):9307–9317.
37. Beck A, Shatz-Azoulay H, Vinik Y, Isaac R, Boura-Halfon S, Zick Y. Nedd4 family interacting protein 1 (Ndfip1) promotes death of pancreatic beta cells. *Biochem Biophys Res Commun*. 2015;465(4):851–856.
38. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes Immun*. 2013;14(2):67–82.
39. Wilson ME, et al. The HMG box transcription factor Sox4 contributes to the development of the endocrine pancreas. *Diabetes*. 2005;54(12):3402–3409.
40. Goldsworthy M, et al. Role of the transcription factor sox4 in insulin secretion and impaired glucose tolerance. *Diabetes*. 2008;57(8):2234–2244.
41. Sun B, Mallampati S, Gong Y, Wang D, Lefebvre V, Sun X. Sox4 is required for the survival of pro-B cells. *J Immunol*. 2013;190(5):2080–2089.
42. Tiernan AR, Champion JA, Sambanis A. Trichostatin A affects the secretion pathways of beta and intestinal endocrine cells. *Exp Cell Res*. 2015;330(1):212–221.
43. Xiao C, Giacca A, Lewis GF. Sodium phenylbutyrate, a drug with known capacity to reduce endoplasmic reticulum stress, partially alleviates lipid-induced insulin resistance and beta-cell dysfunction in humans. *Diabetes*. 2011;60(3):918–924.
44. Khandelwal A, Crowley VM, Blagg BSJ. Natural Product Inspired N-Terminal Hsp90 Inhibitors: From Bench to Bedside? *Med Res Rev*. 2016;36(1):92–118.
45. Watkins RA, et al. Proinsulin and heat shock protein 90 as biomarkers of beta-cell stress in the early period after onset of type 1 diabetes. *Transl Res*. 2016;168:96–106.e1.
46. Ocaña GJ, et al. Inflammatory stress of pancreatic beta cells drives release of extracellular heat-shock protein 90 $\alpha$ . *Immunology*. 2017;151(2):198–210.
47. Houlihan JL, Metzler JJ, Blum JS. HSP90 $\alpha$  and HSP90 $\beta$  isoforms selectively modulate MHC class II antigen presentation in B cells. *J Immunol*. 2009;182(12):7451–7458.
48. Crèvecoeur J, Merville MP, Piette J, Gloire G. Geldanamycin inhibits tyrosine phosphorylation-dependent NF-kappaB activation. *Biochem Pharmacol*. 2008;75(11):2183–2191.
49. Lelli KM, Slattery M, Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet*. 2012;46:43–68.
50. Barth TK, Imhof A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci*. 2010;35(11):618–626.
51. Zou C, Mallampalli RK. Regulation of histone modifying enzymes by the ubiquitin-proteasome system. *Biochim Biophys Acta*. 2014;1843(4):694–702.
52. Wertz IE. TNFR1-activated NF- $\kappa$ B signal transduction: regulation by the ubiquitin/proteasome system. *Curr Opin Chem Biol*. 2014;23:71–77.
53. Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*. 2004;2(1):E9.
54. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 2006;34(Database issue):D535–D539.
55. Cristino AS, et al. Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol Psychiatry*. 2014;19(3):294–301.
56. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
57. Bindea G, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091–1093.
58. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–29.