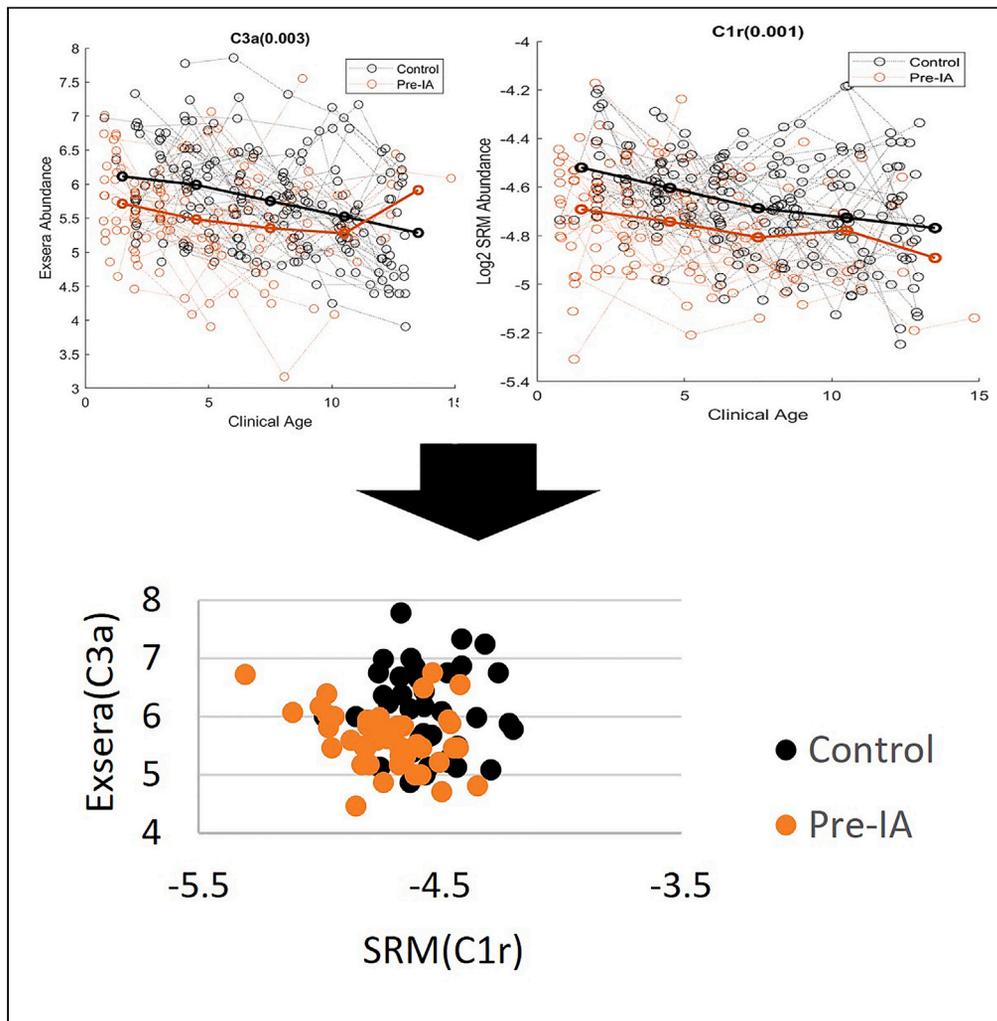


Article

Decrease in multiple complement proteins associated with development of islet autoimmunity and type 1 diabetes



Bobbie-Jo M. Webb-Robertson, Ernesto S. Nakayasu, Fran Dong, ..., V. Michael Holers, Thomas O. Metz, Marian J. Rewers

bj@pnnl.gov (B.-J.M.W.-R.)
marian.rewers@cuanschutz.edu (M.J.R.)

Highlights

The complement pathway is a possible target to predict onset of type 1 diabetes

Complement proteins are lower relative to controls prior to islet autoimmunity

The disruption of the complement system persists through type 1 diabetes onset

Webb-Robertson et al.,
iScience 27, 108769
February 16, 2024 © 2024
Battelle Memorial Institute and
The Author(s). Published by
Elsevier Inc.
<https://doi.org/10.1016/j.isci.2023.108769>



Article

Decrease in multiple complement proteins associated with development of islet autoimmunity and type 1 diabetes

Bobbie-Jo M. Webb-Robertson,^{1,2,3,4,*} Ernesto S. Nakayasu,¹ Fran Dong,⁵ Kathy C. Waugh,⁵ Javier E. Flores,¹ Lisa M. Bramer,¹ Athena A. Schepmoes,¹ Yuqian Gao,¹ Thomas L. Fillmore,¹ Suna Onengut-Gumuscu,⁶ Ashley Frazer-Abel,⁷ Stephen S. Rich,⁶ V. Michael Holers,⁷ Thomas O. Metz,¹ and Marian J. Rewers^{5,8,*}

SUMMARY

Type 1 diabetes (T1D) is a chronic condition caused by autoimmune destruction of the insulin-producing pancreatic β cells. While it is known that gene-environment interactions play a key role in triggering the autoimmune process leading to T1D, the pathogenic mechanism leading to the appearance of islet autoantibodies—biomarkers of autoimmunity—is poorly understood. Here we show that disruption of the complement system precedes the detection of islet autoantibodies and persists through disease onset. Our results suggest that children who exhibit islet autoimmunity and progress to clinical T1D have lower complement protein levels relative to those who do not progress within a similar time frame. Thus, the complement pathway, an understudied mechanistic and therapeutic target in T1D, merits increased attention for use as protein biomarkers of prediction and potentially prevention of T1D.

INTRODUCTION

Clinical onset of type 1 diabetes (T1D) is preceded by a period of islet autoimmunity (IA) that is marked by the appearance of circulating autoantibodies against islet autoantigens.^{1–4} While there is a consensus that chronic β cell autoimmune destruction is triggered by interactions of genetic, genomic, and environmental factors, the pathogenesis of the initiation and progression of the disease is still largely unknown. Identification of biomarkers that predict triggering of IA in at-risk individuals, as well as progression from IA to clinical diabetes, may give clues into the etiology of this complex disease. Complement is part of the innate immune system that once activated can be strongly pro-inflammatory. Activation of the complement system occurs through three interconnect pathways, the classical, the lectin, and the alternative. The classical pathway is primarily activated through interactions with immune complexes; it is important to note that the recognition of immune complexes in circulation by the classical pathway is key to clearance of these complexes.^{5–7} The lectin pathway is primarily activated through recognition of carbohydrates and patterns characteristic of pathogens. The alternative pathway is key as an activation loop, amplifying activation once started. These three pathways converge at C3 and at the start of the terminal pathway. Much of the pro-inflammatory activity of complement resides in the terminal pathway, including formation of the anaphylatoxins C3a and C5a as well as the membrane attack complex (MAC, also known as TCC, C5b-9). The MAC can lyse susceptible bacteria but can be destructive if it attacks self-tissues. Dysregulation of the complement system can affect innate immunity and has been noted in age-related macular degeneration,⁸ cancer,⁹ and autoimmune disorders.^{10,11} Several studies have suggested that the complement system may play a role in the etiology of T1D.^{12–16} For example, local production of complement component C3 is an important survival mechanism in β cells under a pro-inflammatory assault. In response to interleukin-1 β and interferon- γ , C3 expression increases in rodent and human β cells.¹⁷ This increased C3 expression may enhance autophagy—a protective response to β cell stress—and improve β cell function.¹⁸

Proteomics is a core discovery technology utilized to identify biomarkers of disease, as well as gain insight into the molecular processes driving disease progression, which has been employed previously to study the pathogenesis of T1D.^{12,14,15,19,20} These studies have focused on a combination of global and targeted analyses in cohorts evaluating time-based measurements in the progression of T1D. Changes in levels of multiple complement proteins, such as C4, C3, C2, and C1r, have been reported,^{12,14} although the directional changes varied and thus a cohesive pattern associated with the complement system and T1D has not yet emerged.

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

²Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

³Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

⁴Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR, USA

⁵Barbara Davis Center for Diabetes, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁶Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

⁷Division of Rheumatology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁸Lead contact

*Correspondence: bj@pnnl.gov (B.-J.M.W.-R.), marian.rewers@cuanschutz.edu (M.J.R.)

<https://doi.org/10.1016/j.isci.2023.108769>



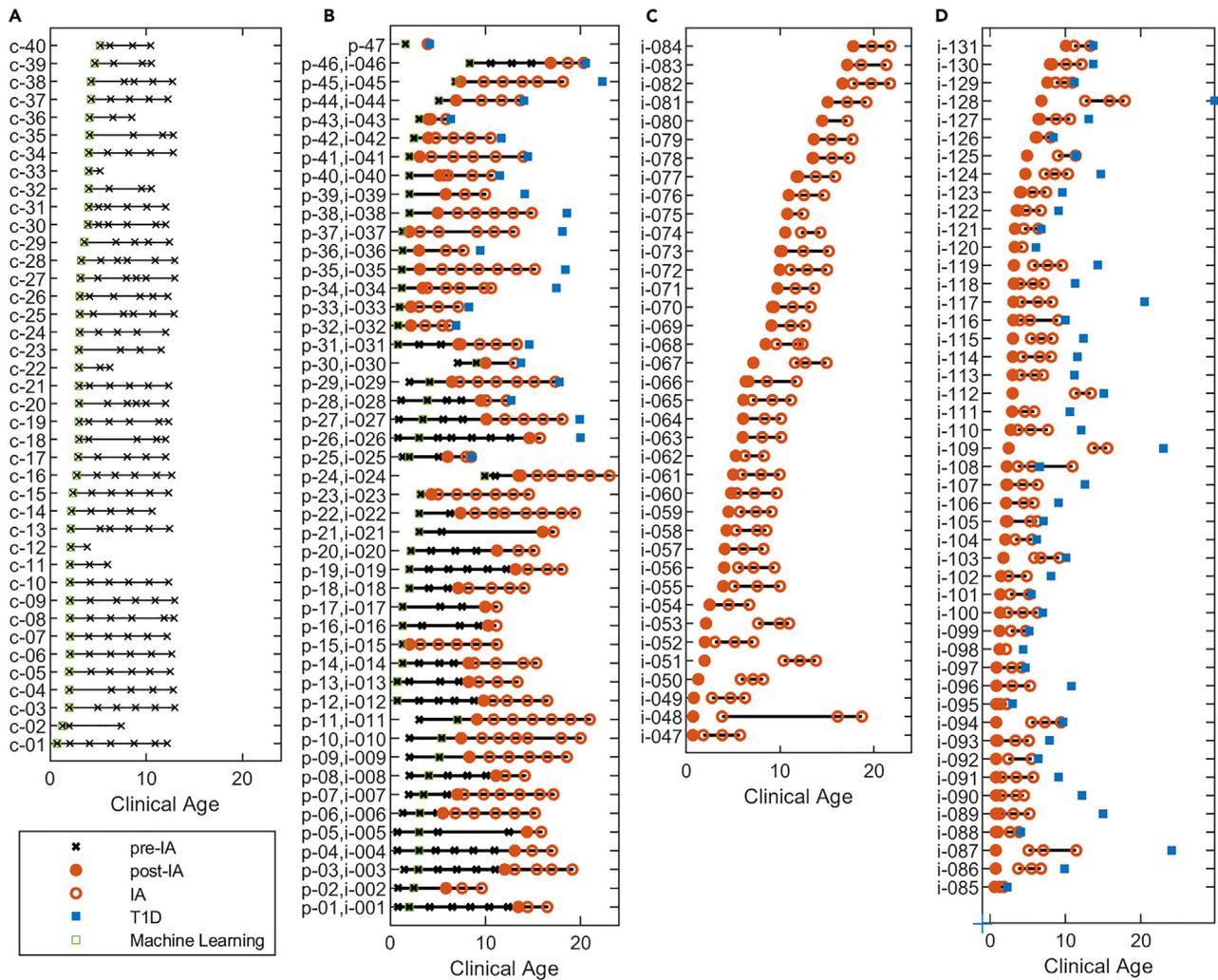


Figure 1. Clinical age of the samples that were analyzed

For each child associated with the study key endpoints highlighted, including if the sample is prior to (x) or after (o) IA, the point of IA (●), the age of T1D diagnosis (■), and the single sample used for the machine learning (ML) analysis (□). Plot is separated into (A) 40 control samples, (B) 47 pre-IA samples, of which 24 were not diagnosed with T1D during the study and 23 were, and 46 of these 47 also had post-IA measurements, (C) 38 post-IA samples who were not diagnosed with T1D during the study, and (D) 47 post-IA samples who were diagnosed with T1D during the study. [Table 1](#) gives specific demographic information on gender, HLA, and first-degree status of the three key groups. Supplemental data provided give specific mapping of individuals with their demographic information and complement measurements: <https://data.pnl.gov/group/nodes/dataset/33588>.

Here we present an investigation of the relationship specifically between the complement system and progression to IA and T1D. To accomplish this task, we analyzed complement proteins in a long-standing birth cohort of high-risk children, the Diabetes Autoimmunity Study in the Young (DAISY). DAISY defines high-risk status as having a first-degree relative with T1D or high-genetic-risk HLA (human leukocyte antigen) genotype.^{21,22} Complement proteins were measured as their constituent peptides using selected reaction monitoring (SRM)-based targeted proteomics.²³ In addition, multiple complement proteins, of which approximately half overlap with the SRM-measured proteins, were also sent for quantitative complement testing using commercial immunoassays in a College of American Pathologists (CAP)/Clinical Laboratory Improvement Amendments (CLIA)-accredited laboratory²⁴ (Exsera BioLabs), referred herein as Exsera. Both approaches are well-established methods for quantitative proteomics and offer confirmatory evidence.

RESULTS

A total of 172 children from the DAISY study with multiple plasma samples collected over time, with up to 23 years of follow-up, were characterized via proteomics analysis, [Figure 1](#). Among the children there were 40 controls ([Figure 1A](#)) and 132 cases ([Figures 1B–1D](#)). All 132 cases had measurements across time relative to IA. Sampling was not consistent for all children. There were 47 of the children who had samples

Table 1. Demographics of the DAISY cohort for the Control, Pre-IA, Post-IA, and Pre-T1D samples

	CTRL (N = 40)	Pre-IA (N = 47)	p value	Post-IA (N = 131)	p value
Female; N(%)	16(40.0)	19(40.4)	0.968	63(48.1)	0.469
HLA; N(%)					
DR3/3 or DR3/X	6(15.0)	5(10.6)		12(17.6)	
DR3/4	11(27.5)	15(31.9)		48(36.6)	
DR4/4 or DR4/X	22(55.0)	21(44.7)		44(33.6)	
DRX/X	1(2.5)	6(12.8)	0.309	16(12.2)	0.048
First-degree relative; N(%)	9(22.5)	27(57.5)	0.001	83(63.4)	<0.001
Age (machine learning)	2.976	3.024	0.107		

The Pre-IA, Post-IA, and Pre-T1D are compared to control where sex, HLA, and first-degree relative are compared to the control group via χ^2 test of independence and age is compared via a two-sample test.

taken and evaluated prior to IA (Pre-IA), represented as p-xx (Figure 1B), and 131 children had measurements at or after IA, but prior to diagnosis of clinical T1D (Post-IA), represented as i-xxx (Figures 1B–1D). The control children were frequency matched on HLA genotypes and age and sex (Table 1) with an observed lower frequency of first-degree relatives within the control group versus the cases. The Pre- and Post-IA proteomics measurements highlighted in Figures 1B–1D were compared to the control samples in Figure 1A using a linear mixed model to evaluate association of complement proteins with these two important events in the progression of T1D.

Association of SRM complement proteins with IA and T1D

The SRM proteomics experiment measured 19 complement proteins, most of which were identified by at least two unique peptides. Protein-level data are presented as the average of the measured peptides. The statistical results are presented visually as an average log₂ fold change at five different age ranges across all subjects. Most complement proteins were distinctly decreased in children with IA relative to controls (Figure 2). The overall pattern of a decrease in complement proteins is persistent both before and after the detection of autoantibodies. Of the 19 proteins measured by SRM, 12 were significantly decreased between IA and T1D post-IA (Figure 2B) at a p value threshold of 0.05. This is consistent with the pattern also observed prior to IA where 5 of the 19 complement proteins significantly decreased (Figure 2A). MBL2 (Mannose-binding lectin 2) unique to the lectin pathway is the only protein in the complement pathways with increased abundance across the entire time course. No proteins specific to the alternative pathway are significant. The time-course plots for each SRM protein with the difference over time by individual subject and data point are available with the data at the identifier listed in the [key resources table](#).²⁵

To validate the SRM proteomics data in Figure 2, we used previously collected data on 16 complement proteins using immunoassays (Exsera)²⁴ of which 10 were in common with the SRM-identified proteins. Figure 3 shows the overall similarity in respect to log₂ fold change between the two measurement types across all of the 10 common complement protein measurements made across all time points relative to control for both Pre- and Post-IA. C3 is the one complement protein whose decrease is significant (p value < 0.05) for both the SRM and Exsera prior to IA and at or after IA. Of the remaining 9 proteins, prior to IA there is one (C4b) significantly decreased for SRM but not Exsera and one (MBL2) significantly increased for Exsera but not SRM. For Post-IA, MBL2 is again significantly increased for Exsera and, although not statistically significant for SRM, shows a similar pattern of increase relative to control. Similarly, C1q, C2, C4b, C5, and CFH (Complement factor H) are significant for SRM and, although not statistically significant for Exsera, show a common pattern of decreased abundance relative to control. Of the six proteins measured by immunoassays that do not overlap with the SRM dataset, three (C3a, C5a, C5b) were significantly decreased for both Pre-IA and Post-IA, again showing the same pattern of decreased abundance relative to controls (Figure 4).

The effect size for each protein based on the statistical model measures the strength of the relationship between the protein abundance and the outcome. As seen in Figure 4, most of the SRM and Exsera quantitatively measured proteins have a negative effect size, meaning that the effect of Pre-IA or Post-IA decreases on average relative to the controls. This matches with the observed log₂ fold changes in Figures 2 and 3. We observe that most proteins, with exception of MBL2, have a negative effect size. For the proteins measured by Exsera we observe that 68.8% and 87.5% of the proteins have a negative effect size for Pre- and Post-IA, respectively, similar to results observed for SRM. To evaluate the likelihood that this could be observed by chance, we simulated effect sizes from a uniform distribution randomly ranging from –1 to 1 for each of the 19 proteins and computed the proportion of negative values. We repeated this process 100 times and as expected the median was near 50%, specifically 47.4%. A Wilcoxon rank-sum test was used to evaluate the null hypothesis that our random distribution could have the same median as our observed data. In all cases, our percentage of negative effect sizes was larger than expected by chance with a p value less than 1.0E-7. We extended this analysis to also include a p value computation based on a random permutation of the class variable to evaluate the likelihood of a significant p value less than 0.05 and a negative effect. The likelihood of observing both the number of significant proteins and a negative directional change by chance is extremely small, less than 2.6E-20, both in the SRM and Exsera datasets. The time-course plots for each Exsera protein with the difference over time by individual subject and data point are available with the data at the identifier listed in the [key resources table](#) for comparison to the SRM protein trends.²⁵

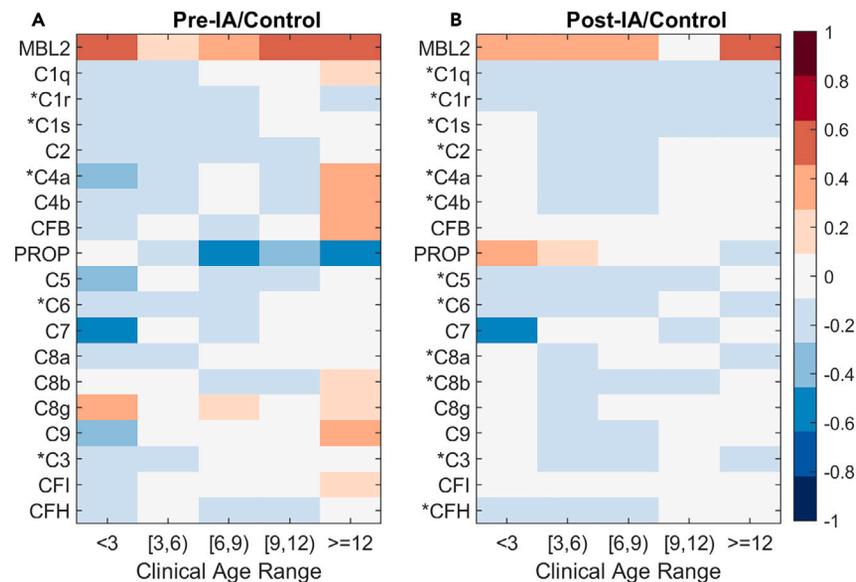


Figure 2. Complement protein log₂ fold change for each statistical comparison of (A) Pre-IA (n=47) and (B) Post-IA (n=131)
Proteins with an asterisk are significant at a p value of 0.05 based on the full linear model.

Prediction of progression based on complement proteins

To further explore the utility of the complement protein quantitation data to screen for children that will develop islet autoantibodies, the 40 control and 47 Pre-IA children were down-selected to a single sample time point for machine learning. For the 40 control children this was the earliest sample collected, and for the 47 Pre-IA children it was a random selection of the first or second time point prior to the detection of autoantibodies to assure the age distributions were not significantly different. Average ages of the two groups are given in Table 1. Of the 47 children in the Pre-IA group, 23 of them are diagnosed with T1D as of the last follow-up. All 25 proteins from both the SRM and Exsera measurements highlighted in Figure 4 were included in the initial dataset. All of the SRM and Exsera data are available at <https://data.pnnl.gov/group/nodes/dataset/33588>, including both the full dataset used for statistics and the down-sampled time point-specific data for machine learning.

To identify the most predictive features, feature importance ranking measure (FIRM)²⁶ was utilized in conjunction with a linear support vector machine (SVM). Analysis was performed using a 75/25 train/test split with repeated 10-fold cross-validation (CV).^{27–29} The accuracy of the model is quantified by a receiver operating characteristic (ROC) curve, specifically the area under this curve (AUC), for which the final average AUC as computed on the test data is 0.82. Figure 5A gives the importance of each of the measured proteins from both the SRM and Exsera technologies ordered from the most important based on the results of FIRM. The most important features are C1r measured by SRM and C3a measured by Exsera, both of which are highly significant and as can be seen in Figure 5B visually separate the two groups fairly well. Evaluation of the data using an alternative approach, recursive feature elimination (RFE), performed repeatedly with 3-fold CV also found that C1r was selected as the last feature to remove from the model in 100% of the RFE iterations and C3a was the second to the last to be removed in 99 of the 100 RFE iterations.

DISCUSSION

There is a consistent pattern of a decrease in complement proteins, except for MBL2, prior to appearance of islet autoantibodies as well as after the seroconversion in children progressing to clinical diabetes in both the SRM and Exsera datasets. The classical and lectin pathways show the strongest decrease in complement proteins, observed by both statistical and machine learning analyses. This is consistent with prior findings of decreased abundance in specific complement proteins, such as C3 and C4, associated with T1D.^{30,31} These findings have two main implications in T1D: (1) complement proteins become strong biomarker candidates and (2) they may play a role in the disease development.

In terms of predictive biomarkers for T1D, The Environmental Determinants of Diabetes in the Young (TEDDY) proteomics study has also identified and validated complement proteins as biomarker candidates.³² This pattern also persists after the onset of the disease.²⁰ The fact that complement proteins have been consistently reported to be reduced during T1D development and even after onset makes these proteins excellent biomarker candidates.³³ A possible challenge of using complement proteins as T1D biomarker is that some other conditions, such as autism, also have complement proteins as potential biomarkers.³⁴ Therefore, multi-panel biomarkers might be needed to properly diagnose specific diseases. Multi-molecule panels are indeed excellent candidates for predictive biomarkers for T1D.^{32,35} With the right panel of molecules it could be possible to develop biomarker assays for simultaneously testing multiple diseases and conditions.

In terms of the mechanism of disease development, an increase in complement activation has been shown to occur in pancreas of individuals with T1D.³⁶ In addition, genome-wide association study combined with glycomics identified associations of glycan structures with

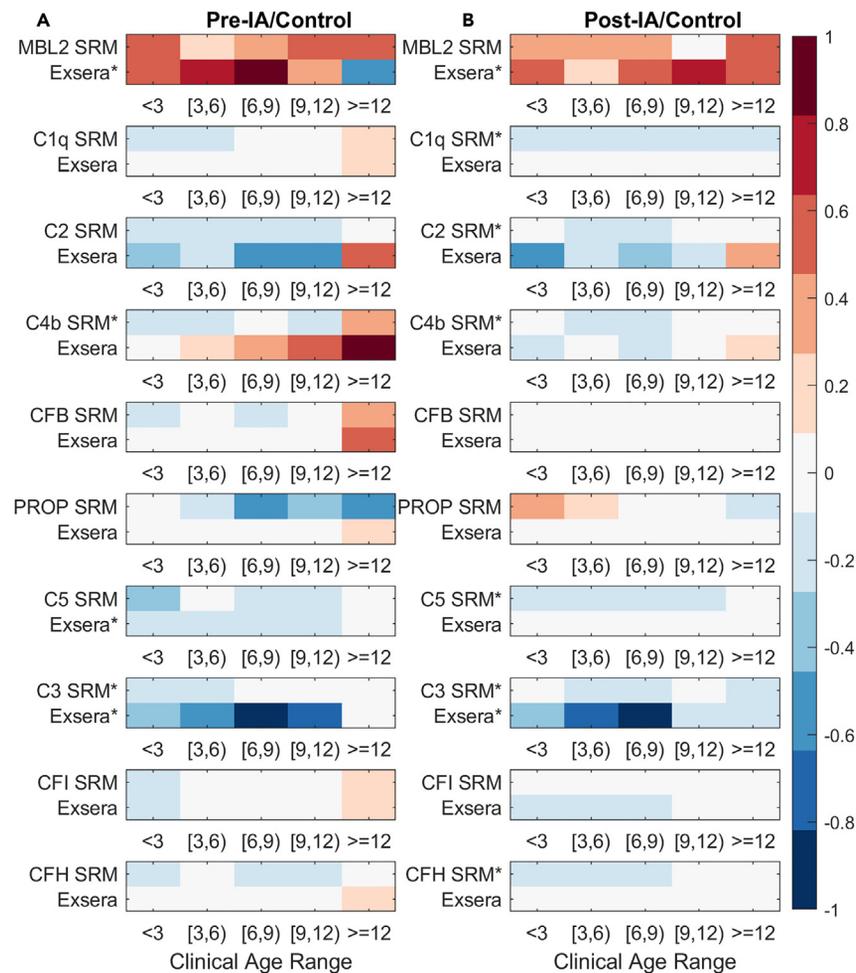


Figure 3. Average Exsera log₂ fold change in comparison to the average log₂ fold change of the SRM measurements across time for the 10 proteins measured by both technologies

(A) Pre-IA comparison to Control.

(B) Post-IA comparison to Control.

The * indicates that the protein measurement was significant based on the full linear model at a p value of 0.05.

possible complement activation in T1D.³⁷ However, it is not known if this deposition has a role in β cell death. Endogenous production of C3 protects β cell apoptosis.³⁸ Similarly, exogenously added C3 protein also has a protective effect on human islets and rodent β cell line INS-1E against cytokine-induced apoptosis.¹⁷ Therefore, it is likely that complement cascade represents protective mechanism for β cells, and that the lower levels of circulating complement proteins fail to provide protection against the autoimmune response. However, understanding how C3-mediated protection relates to systemic levels of complement proteins and whether lower circulating levels are reflective of an inadequate endogenous response to stress will require further study.

In conclusion, our data showed a decrease in complement proteins during T1D development using both targeted proteomics and Exsera assays. This decrease in complement proteins is an excellent biomarker candidate for T1D and might be involved in the disease development.

Limitations of the study

A key limitation of this study is the size of the cohort and the irregularity of the measurements across time. In addition, the statistics and machine learning validation were not performed on an independent cohort.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)

Pathway(s)	Gene	Pre-IA		Post-IA	
		SRM	Exsera	SRM	Exsera
Lectin	MBL2	0.248	0.812*	0.240	0.785*
Classical	C1q	-0.162	-0.016	-0.156*	-0.035
	C1s	-0.107*	nm	-0.092*	nm
	C1r	-0.089*	nm	-0.108*	nm
Classical & Lectin	C2	-0.074	-0.223	-0.081*	-0.186
	C4	nm	-0.121	nm	-0.024
	C4a	-0.268*	nm	-0.177*	nm
	C4b	-0.247*	-0.008	-0.161*	-0.064
Alternative	CFB	-0.062	0.042	-0.070	0.002
	CFD	nm	0.041	nm	-0.015
	PROP	-0.139	-0.045	0.052	-0.030
Terminal	C5	-0.076	-0.166*	-0.132*	-0.101
	C5a	nm	-0.870*	nm	-0.586*
Terminal (MAC)	C5b	nm	-0.306*	nm	-0.189*
	C6	-0.112*	nm	-0.135*	nm
	C7	-0.160	nm	-0.109	nm
	C8a	-0.089	nm	-0.088*	nm
	C8b	-0.069	nm	-0.104*	nm
	C8g	0.140	nm	-0.016	nm
	C9	-0.052	nm	-0.062	nm
Alternative & Terminal	C3	-0.095*	-0.693*	-0.104*	-0.375*
	C3a	nm	-0.370*	nm	-0.203*
	C3b	nm	-0.980*	nm	-0.412
	CFI	-0.060	0.012	-0.052	-0.077
	CFH	-0.051	0.002	-0.090*	-0.013

Effect Size

>0.75

0.5 – 0.75

0.25 – 0.5

0 – 0.25

-0.25 – 0

-0.5 – -0.25

-0.75 – -0.5

<-0.75

nm=not measured

Figure 4. Effect size based on the statistical comparisons either prior to or at or after IA with proteins sorted based on complement pathway

Numbers in the boxes are the effect size from the linear mixed model comparing the difference of protein abundance between Control and the Pre-IA and Post-IA groups. A negative effect size indicates strength of the relationship between the protein abundance and the outcome compared; a positive effect means that Pre- or Post-IA increases the abundance on average relative to control and a negative effect is the contrary.

- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT PARTICIPANT DETAILS**
- **METHOD DETAILS**
 - Targeted proteomic measurements
 - Exsera complement factor measurements
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108769>.

ACKNOWLEDGMENTS

This work was funded by NIH grant R01-DK32493 as well as The Leona M. and Harry B. Helmsley Charitable Trust grants 2018PG-T1D017 and G-2103-05121. Mass spectrometry analyses were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy (DOE) Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle Memorial Institute for the DOE under contract DEAC05-76RLO1830. Identification of SRM targets was based on a large global LC-MS proteomics study supported by Environmental Determinants of Diabetes in the Young (TEDDY) consortium

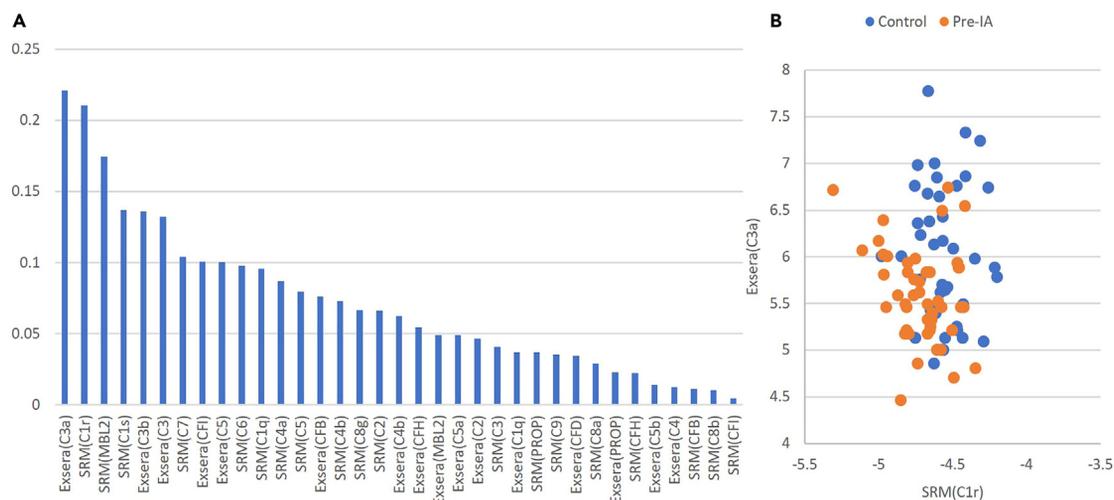


Figure 5. Machine learning results and scatterplot of the two most important features

(A) Feature importance values from most to least.

(B) Scatterplot of the top two features shows a visual separation of control from Pre-IA.

through Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC), and JDRF. TEDDY is supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Colorado (UL1 TR002535).

AUTHOR CONTRIBUTIONS

B.-J.M.W.-R. performed statistical and machine learning analyses and wrote the manuscript. M.J.R. led the study, interpreted results, and wrote the manuscript. E.S.N. and T.O.M. identified the SRM panel and interpreted results. A.A.S., Y.G., and T.L.F. performed SRM proteomics analyses. A.F. and V.M.H. performed the Exsera proteomics analyses and interpreted results. F.D. and K.C.W. designed the study and performed computational analyses. S.O.-G. and S.S.R. conceptualized the study and evaluated results. J.E.F. and L.M.B. generated the machine learning results and data resource. All authors revised and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 13, 2023

Revised: October 16, 2023

Accepted: December 18, 2023

Published: December 20, 2023

REFERENCES

- Ziegler, A.G., Rewers, M., Simell, O., Simell, T., Lempainen, J., Steck, A., Winkler, C., Ilonen, J., Veijola, R., Knip, M., et al. (2013). Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* 309, 2473–2479.
- Kupila, A., Muona, P., Simell, T., Arvilommi, P., Savolainen, H., Hämäläinen, A.M., Korhonen, S., Kimpimäki, T., Sjöroos, M., Ilonen, J., et al. (2001). Feasibility of genetic and immunological prediction of type 1 diabetes in a population-based birth cohort. *Diabetologia* 44, 290–297.
- Liu, X., Vehik, K., Huang, Y., Elding Larsson, H., Toppari, J., Ziegler, A.G., She, J.X., Rewers, M., Hagopian, W.A., Akolkar, B., et al. (2020). Distinct Growth Phases in Early Life Associated With the Risk of Type 1 Diabetes: The TEDDY Study. *Diabetes Care* 43, 556–562.
- Roll, U., Christie, M.R., Fuchtenbusch, M., Payton, M.A., Hawkes, C.J., and Ziegler, A.G. (1996). Perinatal autoimmunity in offspring of diabetic parents. The German Multicenter BABY-DIAB study: detection of humoral immune responses to islet antigens in early childhood. *Diabetes* 45, 967–973.
- Schanzenbacher, J., Köhl, J., and Karsten, C.M. (2022). Anaphylatoxins spark the flame in early autoimmunity. *Front. Immunol.* 13, 958392.
- Tang, A., Zhao, X., Tao, T., Xie, D., Xu, B., Huang, Y., and Li, M. (2023). Unleashing the power of complement activation: unraveling renal damage in human anti-glomerular basement membrane disease. *Front. Immunol.* 14, 1229806.
- Walport, M.J. (2001). Complement. First of two parts. *N. Engl. J. Med.* 344, 1058–1066.
- Armento, A., Ueffing, M., and Clark, S.J. (2021). The complement system in age-related macular degeneration. *Cell. Mol. Life Sci.* 78, 4487–4505.
- Mao, X., Zhou, L., Tey, S.K., Ma, A.P.Y., Yeung, C.L.S., Ng, T.H., Wong, S.W.K., Liu, B.H.M., Fung, Y.M.E., Patz, E.F., Jr., et al. (2020). Tumour extracellular vesicle-derived Complement Factor H promotes tumorigenesis and metastasis by inhibiting complement-dependent cytotoxicity of tumour cells. *J. Extracell. Vesicles* 10, e12031.

10. Giang, J., Seelen, M.A.J., van Doorn, M.B.A., Rissmann, R., Prens, E.P., and Damman, J. (2018). Complement Activation in Inflammatory Skin Diseases. *Front. Immunol.* 9, 639.
11. Sjöwall, C., Mandl, T., Skattum, L., Olsson, M., and Mohammad, A.J. (2018). Epidemiology of hypocomplementaemic urticarial vasculitis (anti-C1q vasculitis). *Rheumatology* 57, 1400–1407.
12. Moulder, R., Bhosale, S.D., Erkkilä, T., Laajala, E., Salmi, J., Nguyen, E.V., Kallionpää, H., Mykkänen, J., Vähä-Mäkilä, M., Hyöty, H., et al. (2015). Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. *Diabetes* 64, 2265–2278.
13. Ajjan, R.A., and Schroeder, V. (2019). Role of complement in diabetes. *Mol. Immunol.* 114, 270–277.
14. von Toerne, C., Laimighofer, M., Achenbach, P., Beyerlein, A., de Las Heras Gala, T., Krumsiek, J., Theis, F.J., Ziegler, A.G., and Hauck, S.M. (2017). Peptide serum markers in islet autoantibody-positive children. *Diabetologia* 60, 287–295.
15. Liu, C.W., Bramer, L., Webb-Robertson, B.J., Waugh, K., Rewers, M.J., and Zhang, Q. (2018). Temporal expression profiling of plasma proteins reveals oxidative stress in early stages of Type 1 Diabetes progression. *J. Proteomics* 172, 100–110.
16. Törn, C., Liu, X., Hagopian, W., Lernmark, Å., Simell, O., Rewers, M., Ziegler, A.G., Schatz, D., Akolkar, B., Onengut-Gumuscu, S., et al. (2016). Complement gene variants in relation to autoantibodies to beta cell specific antigens and type 1 diabetes in the TEDDY Study. *Sci. Rep.* 6, 27887.
17. Dos Santos, R.S., Marroqui, L., Grieco, F.A., Marselli, L., Suleiman, M., Henz, S.R., Marchetti, P., Wernersson, R., and Eizirik, D.L. (2017). Protective Role of Complement C3 Against Cytokine-Mediated β -Cell Apoptosis. *Endocrinology* 158, 2503–2521.
18. Atanes, P., Ruz-Maldonado, I., Pingitore, A., Hawkes, R., Liu, B., Zhao, M., Huang, G.C., Persaud, S.J., and Amisten, S. (2018). C3aR and C5aR1 act as key regulators of human and mouse beta-cell function. *Cell. Mol. Life Sci.* 75, 715–726.
19. Liu, C.W., Bramer, L., Webb-Robertson, B.J., Waugh, K., Rewers, M.J., and Zhang, Q. (2017). Temporal profiles of plasma proteome during childhood development. *J. Proteomics* 152, 321–328.
20. Zhang, Q., Fillmore, T.L., Schepmoes, A.A., Clauss, T.R.W., Gritsenko, M.A., Mueller, P.W., Rewers, M., Atkinson, M.A., Smith, R.D., and Metz, T.O. (2013). Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. *J. Exp. Med.* 210, 191–203.
21. Rewers, M., Bugawan, T.L., Norris, J.M., Blair, A., Beaty, B., Hoffman, M., McDuffie, R.S., Jr., Hamman, R.F., Klingensmith, G., Eisenbarth, G.S., and Erlich, H.A. (1996). Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 39, 807–812.
22. Frohnert, B.I., Ide, L., Dong, F., Barón, A.E., Steck, A.K., Norris, J.M., and Rewers, M.J. (2017). Late-onset islet autoimmunity in childhood: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia* 60, 998–1006.
23. Nakayasu, E.S., Gritsenko, M., Piehowski, P.D., Gao, Y., Orton, D.J., Schepmoes, A.A., Fillmore, T.L., Frohnert, B.I., Rewers, M., Krischer, J.P., et al. (2021). Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* 16, 3737–3760.
24. Prohászka, Z., Nilsson, B., Frazer-Abel, A., and Kirschfink, M. (2016). Complement analysis 2016: Clinical indications, laboratory diagnostics and quality control. *Immunobiology* 221, 1247–1258.
25. Webb-Robertson, B.J., Flores, J., and Rewers, M.J. (2023). DAISY Complement Proteins Dataset.
26. Scholbeck, C.A., Molnar, C., Heumann, C., Bischl, B., and Casalicchio, G. (2020). Sampling, Intervention, Prediction, Aggregation: A Generalized Framework for Model-Agnostic Interpretations. *Commun. Comput. Inf. Sci.* 1167, 205–216.
27. Frohnert, B.I., Webb-Robertson, B.J., Bramer, L.M., Reehl, S.M., Waugh, K., Steck, A.K., Norris, J.M., and Rewers, M. (2020). Predictive Modeling of Type 1 Diabetes Stages Using Disparate Data Sources. *Diabetes* 69, 238–248.
28. Webb-Robertson, B.J.M., Bramer, L.M., Stanfill, B.A., Reehl, S.M., Nakayasu, E.S., Metz, T.O., Frohnert, B.I., Norris, J.M., Johnson, R.K., Rich, S.S., and Rewers, M.J. (2021). Prediction of the development of islet autoantibodies through integration of environmental, genetic, and metabolic markers. *J. Diabetes* 13, 143–153.
29. Webb-Robertson, B.J.M., Nakayasu, E.S., Frohnert, B.I., Bramer, L.M., Akers, S.M., Norris, J.M., Vehik, K., Ziegler, A.G., Metz, T.O., Rich, S.S., and Rewers, M.J. (2022). Integration of Infant Metabolite, Genetic and Islet Autoimmunity Signatures to Predict Type 1 Diabetes by 6 Years of Age. *J. Clin. Endocrinol. Metab.* 107, 2329–2338.
30. Charlesworth, J.A., Timmermans, V., Golding, J., Campbell, L.V., Peake, P.W., Pussell, B.A., Wakefield, D., and Howard, N. (1987). The Complement-System in Type-1 (Insulin-Dependent) Diabetes. *Diabetologia* 30, 372–379.
31. Mason, M.J., Speake, C., Gersuk, V.H., Nguyen, Q.A., O'Brien, K.K., Odegard, J.M., Buckner, J.H., Greenbaum, C.J., Chaussabel, D., and Nepom, G.T. (2014). Low HERV-K (C4) Copy Number Is Associated With Type 1 Diabetes. *Diabetes* 63, 1789–1795.
32. Nakayasu, E.S., Bramer, L.M., Ansong, C., Schepmoes, A.A., Fillmore, T.L., Gritsenko, M.A., Clauss, T.R., Gao, Y., Piehowski, P.D., Stanfill, B.A., et al. (2023). Plasma protein biomarkers predict the development of persistent autoantibodies and type 1 diabetes 6 months prior to the onset of autoimmunity. *Cell Rep. Med.* 4, 101093.
33. Sarkar, S., Elliott, E.C., Henry, H.R., Ludovico, I.D., Melchior, J.T., Frazer-Abel, A., Webb-Robertson, B.J., Davidson, W.S., Holers, V.M., Rewers, M.J., et al. (2023). Systematic review of type 1 diabetes biomarkers reveals regulation in circulating proteins related to complement, lipid metabolism, and immune response. *Clin. Proteomics* 20, 38.
34. Cao, X., Tang, X., Feng, C., Lin, J., Zhang, H., Liu, Q., Zheng, Q., Zhuang, H., Liu, X., Li, H., et al. (2023). A Systematic Investigation of Complement and Coagulation-Related Protein in Autism Spectrum Disorder Using Multiple Reaction Monitoring Technology. *Neurosci. Bull.* 39, 1623–1637.
35. Webb-Robertson, B.J.M., Nakayasu, E.S., Frohnert, B.I., Bramer, L.M., Akers, S.M., Norris, J.M., Vehik, K., Ziegler, A.G., Metz, T.O., Rich, S.S., and Rewers, M.J. (2022). Integration of Infant Metabolite, Genetic, and Islet Autoimmunity Signatures to Predict Type 1 Diabetes by Age 6 Years. *J. Clin. Endocrinol. Metab.* 107, 2329–2338.
36. Rowe, P., Wasserfall, C., Coker, B., Campbell-Thompson, M., Pugliese, A., Atkinson, M., and Schatz, D. (2013). Increased complement activation in human type 1 diabetes pancreata. *Diabetes Care* 36, 3815–3817.
37. Rudman, N., Kaur, S., Simunović, V., Kifer, D., Šoić, D., Keser, T., Stambuk, T., Klarić, L., Pociot, F., Morahan, G., and Gornik, O. (2023). Integrated glycomics and genetics analyses reveal a potential role for N-glycosylation of plasma proteins and IgGs, as well as the complement system, in the development of type 1 diabetes. *Diabetologia* 66, 1071–1083.
38. King, B.C., Kulak, K., Krus, U., Rosberg, R., Golec, E., Wozniak, K., Gomez, M.F., Zhang, E., O'Connell, D.J., Renström, E., and Blom, A.M. (2019). Complement Component C3 Is Highly Expressed in Human Pancreatic Islets and Prevents beta Cell Death via ATG16L1 Interaction and Autophagy Regulation. *Cell Metab.* 29, 202–210.e6.
39. Gianani, R., Rabin, D.U., Verge, C.F., Yu, L., Babu, S.R., Pietropaolo, M., and Eisenbarth, G.S. (1995). ICA512 autoantibody radioassay. *Diabetes* 44, 1340–1344.
40. Wenzlau, J.M., Juhl, K., Yu, L., Moua, O., Sarkar, S.A., Gottlieb, P., Rewers, M., Eisenbarth, G.S., Jensen, J., Davidson, H.W., and Hutton, J.C. (2007). The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc. Natl. Acad. Sci. USA* 104, 17040–17045.
41. American Diabetes Association Professional Practice Committee (2022). 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022. *Diabetes Care* 45, S17–S38.
42. Gibbons, B.C., Fillmore, T.L., Gao, Y., Moore, R.J., Liu, T., Nakayasu, E.S., Metz, T.O., and Payne, S.H. (2019). Rapidly Assessing the Quality of Targeted Proteomics Experiments through Monitoring Stable-Isotope Labeled Standards. *J. Proteome Res.* 18, 694–699.
43. McCulloch, C.E., Searle, S.R., and Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models*, 2nd Edition (Wiley-Interscience).
44. Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *J. Comput. Graph Stat.* 24, 44–65.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human Plasma	Diabetes Autoimmunity Study in the Young (DAISY)	https://medschool.cuanschutz.edu/barbara-davis-center-for-diabetes/research/clinical-epidemiology/daisy-research
Chemicals, peptides, and recombinant proteins		
Custom Synthesized Heavy Isotope-Labeled Peptides	New England Peptides, now Vivitide	N/A
Acetonitrile, HPLC grade	J.T. Baker	9829-03
Acetonitrile anhydrous	Sigma - Aldrich	271004
Ammonium hydroxide solution	Sigma - Aldrich	338818
Buffer A for Multiple Affinity Removal LC Column	Agilent	5185-5987
Buffer B for Multiple Affinity Removal LC Column	Agilent	5185-5988
Chloroform	Sigma - Aldrich	C2432
Dithiothreitol	Thermo Scientific	20291
Ethylenediaminetetraacetic acid	Sigma - Aldrich	E7889
Formic acid	Sigma - Aldrich	33015
Iodoacetamide	Thermo Scientific	90034
HPLC Grade Water	J.T. Baker	4218-03
Hydroxylamine Solution 50%	Sigma - Aldrich	467804
Methanol, HPLC grade	Fluka	34966
Sequencing grade modified trypsin	Promega	V5117
ris (hydroxymethyl)aminomethane hydrochloride pH 8.0	Sigma - Aldrich	T2694
Trifluoroacetic acid	Sigma - Aldrich	91707
Urea	Sigma - Aldrich	U0631
Critical commercial assays		
8-plex iTRAQ kit	Applied Biosystems	4390811
Bb Plus Fragment	Quidel	A027
SC5b-9 Plus EIA	Quidel	A020
C3a Plus EIA	Quidel	A031
Milliplex Human Complement Panel 1	Millipore-Sigma	HCMP1MAG-19K
Milliplex Human Complement Panel 1	Millipore-Sigma	HCMP2MAG-19K
Deposited data		
SRM mass spectrometry data	MassIVE	MSV000090848
SRM and Exsera processed data	DataHub	https://doi.org/10.25584/2229135 https://data.pnnl.gov/group/nodes/dataset/33588
Software and algorithms		
R package (v3.2.3)	The R Project for Statistical Computing	https://www.r-project.org/
MATLAB 2019a	Mathworks Inc.	https://www.mathworks.com/products/matlab.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Q4SRM	Pacific Northwest National Laboratory	https://github.com/PNNL-Comp-Mass-Spec
Skyline	University of Washington	https://skyline.ms/
xPONENT 4.2	Luminex	MAGPX16340721
Gen 5 3.11.19	BioTek	1608102
Other		
Reversed phase tC18 SepPak SPE columns	Waters	WAT054925
3-kDa MWCO Amicon centrifugal filters	Millipore	UFC5003BK
Reversed phase tC18 SepPak SPE columns	Waters	WAT054925

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to Dr. Bobbie-Jo Webb-Robertson (bj@pnnl.gov).

Materials availability

Further information and requests for materials should be directed to and will be fulfilled by the Principal Investigator of the DAISY study, Dr. Marian Rewers (marian.rewers@cuanschutz.edu).

Data and code availability

- The mass spectrometry proteomics data have been deposited in the MassIVE Repository and are publicly available as of the date of publication. The dataset identifier is listed in the [key resources table](#). In addition, individual processed datasets used to generate statistical and machine learning results have been deposited provided in the DataHub system at Pacific Northwest National Laboratory and are publicly available as of the date of publication. The dataset identifier is listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) author upon request.

EXPERIMENTAL MODEL AND SUBJECT PARTICIPANT DETAILS

DAISY follows prospectively 2,547 children at increased risk for T1D. The cohort consists of first-degree relatives of patients with type 1 diabetes and general population children with T1D susceptibility HLA *DR-DQ* genotypes identified by newborn screening,²¹ recruited between 1993 and 2004. Follow-up results are available through April 4, 2022. Written informed consent was obtained from subjects and parents. The Colorado Multiple Institutional Review Board approved all protocols.

Autoantibodies were tested at 9, 15, and 24 months and, if negative, annually thereafter; autoantibody-positive children were retested every 3–6 months. Radiobinding assays for insulin (IAA), GAD (GADA), insulinoma-associated protein 2 (IA-2A), and/or zinc transporter 8 (ZnT8A) autoantibodies were conducted as previously described.^{39,40} Subjects were considered persistently islet autoantibody positive if they had two or more consecutive confirmed positive samples, not due to maternal islet autoantibody transfer, or had one confirmed positive sample and developed diabetes prior to next sample collection. Diabetes was diagnosed using American Diabetes Association guidelines.⁴¹

Analysis included all children with multiple samples across time, which included 172 total children. [Table 1](#) gives an overall summary of the cohort and detailed subject level information is available in the DataHub system at Pacific Northwest National Laboratory and are publicly available as of the date of publication. The dataset identifier is listed in the [key resources table](#). [Figure 1](#) shows the time-based distribution of samples for each child in the context of the experimental groups, which are based on the time a child is autoantibody-positive.

METHOD DETAILS

Targeted proteomic measurements

Protein digestion was carried out in an Eppendorf epMotion 5075 Liquid Handler. Five microliters of plasma from each donor were loaded into 96-well plates and 45 μ L of 8M urea in 50 mM NH_4HCO_3 was added to each sample. Samples were reduced by adding 5 μ L of 100 mM dithiothreitol and shaking at 1200 rpm for 1 h at 37°C. Samples were alkylated by adding 5.5 μ L of 400 mM iodoacetamide and shaking at 1200 rpm in the dark for 1 h at 37°C. Samples were diluted by adding 300 μ L 50 mM NH_4HCO_3 and were supplemented with 1 M CaCl_2 to a final concentration of 1 mM and trypsin (Promega Sequencing Grade Modified Trypsin) to a final ratio of 1/50 (enzyme/protein). Proteins were digested for 6 h at 37°C with shaking at 1200 rpm. Reactions were quenched by adding 10% trifluoroacetic acid to a final concentration of 0.1%. Samples were desalted in C18 solid phase extraction plates (Phenomenex) and dried in a vacuum centrifuge. Samples were dissolved in

100 μL of water and assayed with BCA (Thermo Fisher) to determine peptide concentration. Peptides were spiked with internal standards comprised of synthetic versions of the targeted peptides with heavy-isotope-labeled amino acid residues at their C-termini (Vivitide, previously known as New England Peptide) and diluted to 0.2 $\mu\text{g}/\mu\text{L}$ for mass spectrometry analysis. Assays were tested in different ratios of human and chicken plasma to ensure they were in linear response range.

Two microliters of peptides were loaded into a reverse phase column (Peptide BEH C18, 130 A 1.7 μm 0.1 \times 100mm, Waters) connected to an Acquity M-Class Nano UHPLC system (Waters). The column temperature was set at 45°C and peptides were separated with a gradient of water (solvent A) and acetonitrile (solvent B) both containing 0.1% formic acid. Eluting peptides were analyzed online by selected-reaction monitoring (SRM) in a triple quadrupole mass spectrometry (TSQ Altis, Thermo Fisher). The electrospray voltage was set to 2.1 kV and the source temperature at 350°C. The LC-SRM raw data are available on MassIVE (<https://massive.ucsd.edu>); MSV000090848. Due to limitations on the number of samples that can be run at a single time the subjects were randomized and balanced across a total of 10 plates. Plate was captured for adjustment for this factor in the statistical analysis. Data quality was monitored using an in-lab developed tool name Q4SRM.⁴²

All the LC-SRM data were imported into the Skyline software (MacLean et al., 2010) and the peak boundaries were manually inspected to ensure correct peak assignment and peak boundaries. There were 333 peptides representing 169 proteins measured in the final assay, and each peptide were monitored by 2–3 precursor-fragment ion pairs (i.e., transitions). The information about specific transitions were deposited within the Skyline files and can be accessed at https://panoramaweb.org/DAISY_SRM_PNNL.url. Peak detection and integration were determined based on two criteria: 1) the same LC retention time and 2) approximately the same relative peak intensity ratios across multiple transitions between the endogenous peptides and heavy isotope-labelled internal peptide standards. The total peak areas of endogenous peptides and their ratios to the total peak areas of the corresponding heavy isotope-labelled internal peptide standards were exported directly from the Skyline software. No further data manipulation was performed. This information can also be found with the deposited Skyline files.

Exsera complement factor measurements

Immunological complement analysis at Exsera BioLabs was performed in plasma that had not been previously thawed by a combination of multiplex and single assay methods. For the multiplex analysis the human complement bead-based xMAP technology (Luminex Corp, Northbrook IL) and commercially available kits (EMD Millipore, Milliplex Map, Burlington, MA) were used to measure thirteen complement proteins, spanning all three activation arms and the terminal pathway of complement. Measurements were made on a MagPix Luminex instrument. The Millipore Panel #1 was used to measure C2, C4b, C5a, C9, FD, MBL and Factor I (FI). Panel #2 was used to measure C1q, C3, C3b & iC3b, C4, FB, Factor H (FH), and P. In addition, the complement activation markers Bb, C3a and the soluble terminal complement complex, sC5b-9 were measured by ELISA (Quidel Corp, San Diego CA). All testing methods had been optimized and validated within Exsera BioLabs, a College of American Pathologists (CAP) and Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. The complement measurements were run as blinded samples across consecutive plates as was done in the SRM measurements.

All analysis was performed in duplicate with the resulting mean values report. For the multiplex Luminex data the mean fluorescent intensity was the raw value and for the ELISA analysis the raw values were optical density. Standard curves were utilized with a four-parameter parametric curve fit used to calculate the absolute quality in ng/mL or mg/mL, as appropriate. Three quality controls (QC) were included in each run, including at least one laboratory developed and characterized QC. The QCs were monitored for performance and for all testing in the study the values returned were within required parameter, demonstrating assay performance. No further data manipulation was performed. Human reference ranges for the analytes tested have been determined within Exsera by the measurement of normal individuals.

QUANTIFICATION AND STATISTICAL ANALYSIS

A linear mixed model comparing the difference of SRM protein abundance between CTRL and the Pre-IA and Post-IA groups, adjusting for sex, HLA group, and first-degree relative status with a nested random effect for subject and plate number was performed.⁴³ The final statistics, effect sizes and log₂ fold-changes for the SRM protein data for Pre-IA and Post-IA is available in [Tables S1](#) and [S2](#), respectively. Statistical analysis of the Exsera data was identical except that plate number was not a factor. Statistical analysis was performed in MATLAB using the 'fitlme' function, from which p values and effect size can be extracted as output arguments. The final statistics, effect sizes and log₂ fold-changes for the Exsera protein data for Pre-IA and Post-IA is available in [Tables S3](#) and [S4](#), respectively. The average log₂ abundance values of all Pre-IA, Post-IA and control samples for each subject within the age range was used to compute log₂ fold-changes. All functions are available in Statistics and Machine Learning Toolbox in the MATLAB platform.

A linear support vector machine (SVM) was employed for machine learning with model-agnostic feature importance ranking metrics generated by Feature Importance Ranking Measure (FIRM)²⁶ using the tidymodels R package. The choice of the hyperparameters for the SVM was determined as the combination that yielded the highest average Area Under a Receiver Operator Curve (AUC) across repeated 10-fold cross-validation. The final parameters used for the SVM was a margin of 0.9 and cost of 0.0132. To rank the features, FIRM is based on Individual Conditional Expectation (ICE) curves.⁴⁴ Briefly, ICE curves measure the effect of a predictor (or small subset of predictors) on the estimated prediction surface. Predictors that have no effect on the response correspond to flat ICE curves. Thus, to quantify importance, the FIRM approach averages the computed variances of the ICE values for a continuous predictor. The machine learning analyses used functions available in R using the tidymodels package.