COMIRB Protocol

COLORADO MULTIPLE INSTITUTIONAL REVIEW BOARD
CAMPUS BOX F-490 TELEPHONE: 303-724-1055 Fax: 303-724-0990

Project Title: DIABETES AUTOIMMUNITY STUDY IN THE YOUNG
Principal Investigator: DR. MARIAN REWERS
Version date: 5/6/21

I. Hypotheses and Specific Aims:

The incidence of type 1 diabetes (T1D) has quadrupled over the past 40 years. T1D now affects 1.4 million people in the U.S.; at least as many asymptomatic persons express multiple islet autoantibodies and will develop T1D in the next 10 years. The Diabetes Autoimmunity in the Young (DAISY) prospective cohort has already provided a wealth of clues concerning the natural history and etiology of childhood T1D. With the cohort entering adulthood we are proposing to explore the natural history of autoimmune diabetes phenotypes in young adults. While it is generally accepted that the chronic autoimmune destruction of pancreatic β-cells leading to T1D is triggered by an interaction of environmental factor(s) with a relatively common genetic background, the specific cause remains elusive. We also poorly understand the determinants of prolonged relapsing-remitting course of islet autoimmunity (IA) leading to diabetes. In this application, we propose to shift the paradigm from evaluating a handful of candidate triggers or promoters at a time, to predictive modeling across an array of biomedical domains and serial measurements spanning up to 25 years of life. A major focus of this application is to expand our understanding of changes in the human proteome associated with triggering of IA and progression to diabetes. Previously collected environmental exposure and immune markers data will be combined with genetic, epigenetic, transcriptomic, and metabolomic profiles of DAISY participants obtained through independent funding. We will fully interrogate the complex interplay of IA/T1D biomarkers by combining of Bayesian integration with the power of individualized algorithms to inform new approaches to prevention.

- **Specific Aim 1. Determine the natural history of islet autoimmunity and to explore the heterogeneity of diabetes and other autoimmune phenotypes in young adults through follow-up of the existing cohort for 5 more years.**
- We will follow already established cohort of youth at high risk of T1D and other autoimmune diseases (current n=1149, median age 17.2 y, IQR 13.5-20.2 y). We will estimate overall burden of pre-clinical and clinical T1D, celiac, thyroid, adrenal, rheumatic and parietal autoimmune disease in Colorado by age 25. This will inform future screening and prevention programs. We will further explore the apparent heterogeneity of IA and its implications for adult-onset diabetes. Selected hypotheses include:
- Hypothesis 1.1: Persistent IA triggered after the age of 6 y results in T1D diagnosis in early adulthood
- Hypothesis 1.2: Transient childhood IA does not increase the risk of diabetes in early adulthood
- Hypothesis 1.3: Sedentary life-style and insulin resistance predict faster progression from persistent IA to T1D

- **Specific Aim 2. Validate candidate proteomic biomarkers of IA and T1D, in a nested-case study of 213 youth with persistent IA and 213 controls.**

- So far, none of the 'omics biomarkers reported to predict progression from IA to T1D have been confirmed by independent studies. Using targeted proteomics we will validate candidate peptides/proteins reported from several discovery studies. Selected hypotheses:
- Hypothesis 2.1: A reproducible set of proteomic biomarkers will predict development of IA independently of demographic factors and HLA class II genotypes.
- Hypothesis 2.2: A reproducible set of proteomic biomarkers will predict the rate of progression from IA to T1D independently of demographic factors, HLA class II genotypes.

- **Specific Aim 3. Develop an integrated comprehensive model of the relapsing-remitting process leading to T1D in youth using prospectively collected DAISY 'omics data.**
- Integrative Bayesian modeling, based on a small set of disparate features e.g., gene variants, proteins, or metabolites, will be used to generate individualized prediction algorithms in IA progressors vs. non-progressors to identify potential pathways. Selected hypotheses:
- Hypothesis 3.1: A limited number of features derived from high-resolution 'omics data can reliably predict development of IA on an individual basis
- Hypothesis 3.2: A limited number of features and time-points derived from high-resolution 'omics data can reliably predict the rate of progression from IA to T1D on an individual basis
-
- The proposed studies are important to reach our overarching goals: to find the environmental causes of T1D, develop primary prevention, and inform public health screening for several autoimmune disorders in children and adolescents. We will continue to make every effort to share DAISY resources with multiple investigators studying T1D and other autoimmune diseases through an open-source database/ repository.

II. **Background and Significance**:

The incidence of type 1 diabetes (T1D) has quadrupled over the past 40 years, now affecting1.4 million people in the U.S. At least as many asymptomatic persons express multiple islet autoantibodies and will develop T1D in the next 10 years. The Diabetes Autoimmunity in the Young (DAISY) prospective cohort is now entering adulthood so we propose to explore the natural history of autoimmune diabetes phenotypes in young adults. While it is generally accepted that the chronic autoimmune destruction of pancreatic β-cells leading to T1D is triggered by an interaction of environmental factor(s) with a relatively common genetic background, the specific cause remains elusive. We also poorly understand the determinants of prolonged relapsing-remitting course of islet autoimmunity (IA) leading to T1D. We propose to shift the paradigm from evaluating a handful of candidate triggers at a time, to predictive modeling across an array of biomedical domains and serial measurements spanning up to 25 years of life. Our focus is to expand our understanding of changes in human proteome associated with development of IA and T1D. Combining environmental exposures and immune marker data with genetic, epigenetic, transcriptomic, and metabolomic profiles will allow us to interrogate the complex interplay of IA/T1D biomarkers utilizing Bayesian integration with the power of individualized algorithms to inform new approaches to prevention. 1) The established cohort of youth at high risk of T1D and other autoimmune diseases (n=1149, median age 17.2 y, IQR 13.5-20.2 y) will be followed to estimate overall burden of T1D, celiac, thyroid and other autoimmune disease, by age 25. We will further explore the apparent heterogeneity of islet autoimmunity and its implications for adult-onset diabetes. 2) Targeted proteomics will validate candidate protein biomarkers of IA and T1D reported from discovery studies. 3) Integrative Bayesian modeling, based on a small set of disparate features (gene variants, proteins, or metabolites) will be used to generate individualized prediction algorithms in IA progressors vs. non-progressors and to identify potential pathways. The proposed studies are important to reach our overarching goals: to find the environmental causes of T1D, develop primary prevention, and inform public health screening for several autoimmune disorders

in children and adolescents. We will continue to make every effort to share DAISY resources with multiple investigators studying T1D and other autoimmune diseases through an open-source database/ repository. 6

DAISY Study History

The DAISY Study has received initial COMIRB approval on 02/17/1994 and has received COMIRB approval annually, thereafter. The DAISY study has submitted occasional Amendments and updates to the original protocol as the study subjects progressed in the main study and sub-studies were established. The DAISY study has secured NIH-NIDDK funding (R01 DK032493) since 1994 through 2020.

Autoimmunity as the intermediate end point in studies of T1D etiology and prevention

Islet autoimmunity (IA), defined structurally as insulitis (difficult to study in humans) or functionally as the presence of autoantibodies and T-cell clones specific to pancreatic beta-cell autoantigens, has become an alternative endpoint in T1D research. Advantages of this approach include the opportunity to study the pathological process underlying T1D and to verify that a candidate risk factor is not only related to diabetes, but also to the alteration of structure and function preceding clinical onset. Studies using as end points both IA and T1D allow differentiation of the risk factors which initiate humoral autoimmunity from those which promote progression from subclinical autoimmunity to T1D. For instance, of 15 genetic loci linked to diabetes in a mouse model (NOD mouse), three affect insulitis (J.A.Todd personal communication). Even within one locus, e.g., the human HLA-DQB1 gene, some alleles (*0302 or *0201) appear to be associated with IA, while others (*0602) suppress progression from IA to T1D[16,17]. It is plausible that the environmental factors causing T1D (yet to be identified) interact with initiating and promoting/protecting genetic factors at various stages of the pathogenic process leading to T1D. Some will be found to trigger IA, while others may be found to promote progression to T1D or to protect from it. This is a key hypothesis for this prospective study. Unlike the binary diabetes outcome, progression (and theoretically regression) of the severity of autoimmunity can be temporally related to serial measurements of risk factors. An increasing number of epidemiological studies into the etiology of T1D have been using autoantibody-defined endpoints[18]. In the near future, accurate serial measurement of autoimmunity progression or regression scores may become a useful end point for intervention trials, rendering them vastly smaller, shorter and less costly than trials using diabetes as the end point. T1D research follows, in this respect, the road traveled by research in AIDS[19], cancer[20] and cardiovascular disease[21,22].

The methods to detect islet autoantibodies have improved dramatically over the years. The islet cell antibody (ICA) assays, based on immunohistochemistry with pancreatic tissue, have been extremely useful in defining persons with IA and at risk of T1D, but their results have been markedly variable among laboratories. DAISY has contributed to the “harmonization” of assays across laboratories and which has improved the sensitivity, specificity and predictive values of screening programs to identify persons at risk for T1D (refer to Progress Report-New Assays and Technologies). The DAISY project is unique in the field of T1D research in that it combines "cutting edge" molecular immunology and genetics techniques with population-based epidemiological methods and immunology.

Gene-environment interaction as the cause of T1D

The cause of T1D is multifactorial and includes the effects of genes, at perhaps as many as eleven loci[12], interacting with several plausible environmental agents[9-11,17]. The optimal prevention of T1D will hardly be possible without full identification of these environmental factors, assessment of the relative risks they confer and the understanding of how they interact with the susceptibility or protective genotypes. Large family studies can yield relative risk estimates for occurrence of T1D associated with a putative environmental exposure, but inference from such data to the general population, where 90% of T1D cases occur, may not be straightforward[23].

While T1D is polygenic in nature, the *IDDM1* locus, including the HLA-DR and DQ genes, is the only major genetic determinant, accounting for up to 50% of the familiar clustering of the disease[24,

[25]. Approximately 95% of all T1D cases have either the DRB1*03,DQB1*0201 or the DRB1*04,DQB1*0302 haplotype. While only 2% of the general population are DRB1*0301,DQB1*0201/ DRB1*04,DQB1*0302 heterozygotes, this genotype is present in 30-40% of T1D patients[26-28] (in up to 52% of those who develop diabetes in the first 10 years of age)[29]. Thus, a great deal is to be learned about the causes of T1D by studying the interactions between plausible environmental causes and the HLA-DR,DQ genotypes. Such studies, in families and especially in the general population have been hampered by the cost and logistics of a large-scale HLA typing. We have developed a relatively simple and highly reliable PCR-based test for detection of the HLA-DRB1*03,DQB1*0201/ *04,DQB1*0302 genotype and other T1D susceptibility and protection genotypes, and we tested the procedure in over 20,000 general population newborns[30]. This protocol is focused on optimization of the screening methods for future use in routine general population screening and prevention programs. This includes development of reliable and cost-efficient sample collection and HLA typing methods as well as evaluation of the usefulness of additional (non-HLA) genetic markers for possible addition to the screening. Among non-HLA loci linked with T1D, the *IDDM2* locus near the insulin gene on chromosome 11p is currently best characterized and is potentially applicable to the epidemiological studies. It can be assessed by assaying the -23 Hph diallelic polymorphism[31].

Candidate environmental causes of T1D

The major unresolved questions concern the environmental causes of islet autoimmunity (IA) and the exact age and sequence in which they operate. It is also unknown whether environmental factors that initiate IA are different from those that promote progression from IA to T1D. The DAISY Study was designed to answer these questions in the general population of children followed prospectively from birth with frequent measurements of candidate environmental exposures and the outcome defined as the presence of islet autoantibodies.

Viruses There has been a recent resurgence of evidence that both T1D and IA may be associated with enteroviral infections[11,32-34]. Presence of antibodies against enteroviruses at the time of T1D onset or in persons with IA does not necessarily prove a causal relationship. People with T1D or IA may be more prone to enteroviral infection, may have a stronger humoral response to infection due to their particular HLA genotypes or, as suggested by our data, they may be in a non-specific hyper-immune state marked by elevation of antibody levels to a variety of exogenous antigens. However, if the enteroviral infections are indeed associated with triggering of both IA and T1D, one of the following mechanisms could operate: (i) several distinct infections are necessary, the first, early in ontogenesis, leads to a low-grade persistent infection while the subsequent infections with antigenetically similar virus initiate and sustain IA until the final 'hit' results in beta-cell loss sufficient to cause T1D; (ii) enteroviral infection does not trigger IA, but recurrent acute lytic infections of beta-cells promote beta-cell loss and T1D in persons with IA; (iii) persistent enteroviral infection of beta-cells impairs insulin secretion without cell lysis and promotes T1D in persons with IA; or (iv) an acute or chronic enteroviral infection of periinsular tissue leads to beta-cell destruction from abundance of free radicals - the 'innocent bystander theory'. Finally, it is plausible that early infection with a non-diabetogenic strain of the virus can induce immunity to antigenetically similar diabetogenic strains and protects from IA and T1D. These hypotheses are based on extensive studies in animal models[11,35-37], but have been extremely difficult to test in human populations because of the need to follow prospectively large groups of young children at risk for T1D. DAISY offers a unique opportunity to test at least some of these hypotheses, because it has been collecting at relatively frequent intervals appropriate biologic samples in such a population. In addition we have unique experience in using state-of-the-art techniques for sensitive and specific detection of both enteroviral RNA (to demonstrate current acute or chronic infection) and antibodies (to document past infection).

Dietary factors A proposed protective effect of breast-feeding on the incidence of T1D[38] has attracted enormous interest. Subsequent studies have shown varying results: (i) diabetic children

were significantly less likely to have been breast-fed than non-diabetic children[39-42], (ii) there was no significant difference in the frequency of breast feeding between diabetic and non-diabetic children[38.42-51], or (iii) diabetic children were significantly more likely to have been breast-fed than non-diabetic children[52]. Certain studies suggested that the longer children had been breast-fed, the lower was their risk for developing diabetes[38,40,41,53,54], while other studies did not show increased protection with longer duration of breast-feeding[48-50,52,55]. A meta-analysis of selected studies suggested that children with diabetes are 60% more likely to have had an early exposure to cow's milk than non-diabetic children[56]. A major limitation of most of the aforementioned studies is that the infant diet data were based on long-term maternal recall, which is subject to error[57]. In fact, our meta-analysis found that the studies using prospectively-collected records to assess infant diet[48,50,58] did not find the associations between T1D and infant diet exposures found in studies that relied on recalled data. This suggests that there may be bias in the retrospective assessment of infant diet. In fact, DAISY has found little evidence, so far in the sibling/offspring cohort, for a role of exposure to cow's milk in the origin of IA. Despite these limitations, dietary intervention trials to prevent T1D are underway or have been proposed.

Environmental toxins Toxic doses of nitrosamine compounds can cause diabetes[59-61], due to the generation of free radicals. The effect of dietary nitrate, nitrite or nitrosamine exposure on human T1D risk is less clear[55,62-64]. Breast milk and vitamin supplements are important sources of anti-oxidants and may theoretically reduce the concentration of free radicals and thus the risk of beta-cell damage. A number of additional environmental toxins have been implicated in T1D etiology[65,66], but never explored in a cohort study.

Importance of further follow-up of DAISY cohort

The DAISY cohort, including very young relatives and children without diabetic relatives, represent over 90% of T1D cases diagnosed in childhood and young adulthood[1,67,68]. Individuals without a family history of T1D or who are younger than 4 years old have been excluded from the Diabetes Prevention Trial-1, because of insufficient data concerning the natural history of pre-T1D autoimmunity in these groups. Our study is filling an important gap in the understanding of the events leading to most cases of T1D. Studies in these populations are difficult because of the relatively low incidence of IA and T1D and participation of young children and persons who lack familiarity with diabetes. Follow-up of a sufficiently large study population through the high-risk age categories will likely bring more complete understanding of T1D causes and prevention opportunities.

III. Preliminary Studies/Progress Report:

DAISY has made major contributions to the understanding of the natural history of islet autoimmunity (IA) and the etiology of type 1 diabetes (T1D). It was the first study in the world to make the leap from studying relatives of affected patients (<10% of T1D cases[1]) to including the general population[2]. While absolute risk differs, DAISY has demonstrated that the natural history of pre-T1D and associated risk factors are essentially the same in first-degrees relatives and the general population. In collaboration with German and Finnish investigators, DAISY has shown that 70% of children with multiple islet autoantibodies progressed to diabetes over the next 10 years, regardless of family history of T1D or country of residence[3]. DAISY findings had an immediate and reassuring impact on public perception of potential causes of T1D, e.g., that routine immunizations[4], shorter breast-feeding duration or cow's milk formulas do not cause IA or T1D[5;6]. DAISY has discovered both environmental[7;8] and genetic[9-11] factors associated with initiation of IA[12-14] and progression from IA to T1D and that can be translated into prevention trials. And finally, this cohort has contributed significantly to the "discovery" of celiac disease as a common entity in the U.S.[15;16].

The 2001 competitive DAISY application was awarded Merit Status (R37). The 2011 renewal received score 10 at 1%. A brief summary of the work completed to fulfill 2011specific aims is provided below. For additional preliminary results see Approach sections.

2011 Specific Aim 1. To determine the natural history of islet autoimmunity (IA) and other autoimmune phenotypes and to explore the heterogeneity of IA and T1D in children and adolescents

Cumulative incidence of IA by age 15 in Denver general population children by HLA-DR,DQ genotypes is summarized in Figure 1.

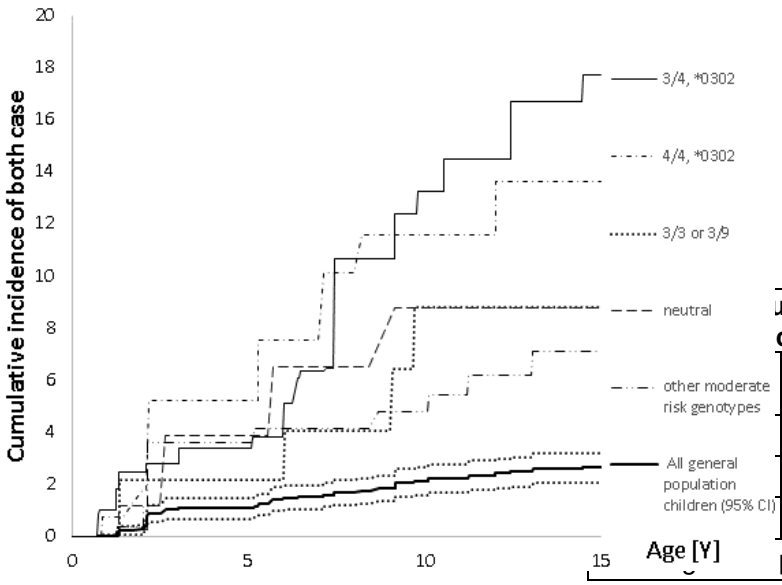


Figure 1: Cumulative incidence of persistent IA in general population by HLA-DR,DQ genotype and a weighted estimate for all children and 95% C.I.

Other moderate risk genotypes: DR1,8,9,10,11,12,13 or 14 and 4,DQB1*03:02; DR 3/1, 3/2, or 3/4,DQB1*03:01

ılative incidence % by age confidence interval)	
Persistent IA[#]	T1D or multiple autoantibodies
1.1 (0.7-1.5)	0.5 (0.3-0.7)
2.1 (1.6-2.7)	1.0 (0.7-1.4)
2.7 (2.1-3.2)	1.4 (1.0-1.8)
population children, Fig.1 (left)	

For efficiency, DAISY follows primarily children with increased-risk HLA genotypes present in 84% of T1D patients diagnosed below age 30y, but only few with protective HLA genotypes present in 71% of the general population[17]. To estimate cumulative incidence of persistent IA in all Denver children, we used the observed HLA genotype-specific rates (**Figure 1**) weighted for the genotype population frequencies[17]. We conservatively assumed no IA cases in children with protective HLA genotypes. By age 15 y, an estimated 2.7% (2.1-3.2%) of Denver children develop IA persisting for at least 6-12 months (**Table 1**). Nearly half - 1.4% (1.0-1.8%) develop T1D or multiple autoantibodies conferring an 85% risk of T1D in the next 15-years[3]. The observation that only ~0.4% of the general population develops T1D by age 20[18] with the remaining 1.0% diagnosed as adults has informed Specific Aim 1 of this 2015 proposal.

The incidence of IA is higher in relatives of T1D patients particularly in siblings and offspring of men with T1D. Among those with the HLA-DR3/4,DQB1*0302 genotype, 51% (95%CI 33-66%) of siblings develop IA by age 15 y, compared to 18% (95% CI: 13-24%) of children without a T1D first-degree relative HR=3.8 (2.2-6.3).

Cumulative incidence of celiac autoimmunity and disease in Denver general population children

In children homozygous for HLA-DR3-DQB1*02, cumulative incidence of persistent transglutaminase autoantibodies (TGA) by 5, 10 and 15 y was 14.8%, 22.1%, and 26.4%, respectively; celiac disease (CD) was confirmed in, respectively 9.8%, 14.2%, and 14.2%.

In HLA-DR3-DQB1*02 heterozygotes, the cumulative incidence of TGA was 8.6%, 15.1%, and 16.3% and that of CD was 6.1%, 11.1%, and 11.6%, respectively. In contrast to IA and T1D, few new cases of TGA or CD occurred after age 10 y. The cumulative incidence of persistent TGA and CD in all Denver children (**Table 2**) was estimated similarly as for IA and T1D, but using CD-specific risk genotypes: DR3-DQ2/DR3-DQ2, DR3-DQ2/X, DR3-DQ2/DR4-DQ8, DR4-DQ8/DR4-DQ8 and DR4-DQ8/X. A comparison of Table 2 and Table 1

Table 2: Cumulative incidence % by age (95% confidence interval)

Age, y	Persistent TGA	Celiac disease[#]
5	2.4 (1.8-3.2)	1.7 (1.2-2.4)
10	4.3 (3.5-5.3)	3.1 (2.4-4.0)
15	5.1 (4.2-6.1)	3.3 (2.6-4.2)

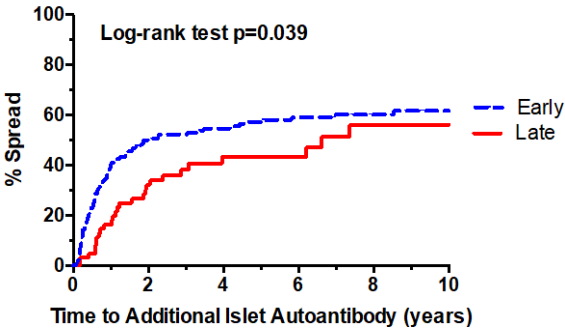
[#]positive biopsy or high-titer TGA

shows that, in the general population, the cumulative incidence of TGA and CD is approximately twice that of IA and T1D.

Multiple autoimmune phenotypes DAISY has previously reported that 23.8% (20.7-27.2%) of participants developed at least one autoimmune marker: IA, celiac, thyroid, adrenal, and rheumatoid autoimmunity by age 12 and hypothesized that at least 7% of Denver children express one or more of the autoimmune phenotypes by age 18 y. To balance administrative reduction of the budget, measurements of thyroid, adrenal, rheumatic and parietal autoantibodies on stored samples, was postponed until most of the cohort reaches the age of 20 y. T1D and CD are of particular interest as their detection at a pre-clinical stage meets the WHO criteria for large-scale screening[2;19]. We have piloted a screening in general pediatric care clinics[20] and have been invited by the JDRF and the Helmsley Charitable Trust to develop a demonstration project to screen 70,000 Denver children for IA and TGA. DAISY has shown that early detection of IA and education on diabetes symptoms can prevent diabetic ketoacidosis (DKA) and ~80% of hospitalizations at onset of T1D[21]. This is important as the proportion of children presenting with DKA is on the rise[22].

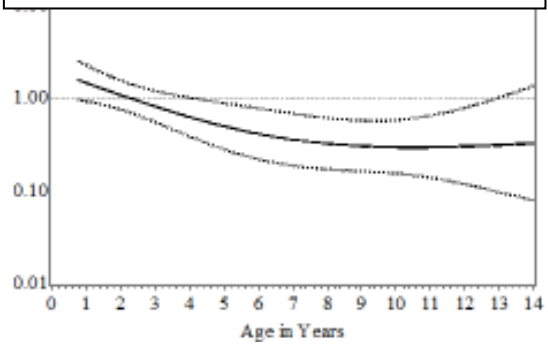
Persistent IA developing after age 7 progresses more slowly to diabetes than early-onset IA and is associated with different dietary exposures. Early-onset IA developed in 143 subjects; late-onset IA developed in 65. Late-onset IA was less likely to initially present with multiple autoantibodies (15% vs. 32%, *p*=0.011), spread to additional autoantibodies (**Figure 2**) or progress to T1D (p=0.01) than early IA. Children with late-onset IA were less likely to have a sibling with T1D (12% vs. 27%, p=0.004) or be non-Hispanic white (14% vs. 32% vs. 14%, *p*=0.002). However, the frequencies of the HLA-DR3/4,DQB1*03:02 genotype or T1D-associated SNPs in BACH2, CTLA4, ERBB3, GLIS3, GSDM/ORMDL3, IL27, IL2RA, INS, MICA, PTPN22, UBASH3A, IFIH1 and CLEC16A did not differ between early and late IA.

Figure 2: Slower spread to additional autoantibodies in late vs. early IA



Higher levels of Ω-3 fatty acids (**Figure 3**)[23] and vitamin 25[OH]D decrease IA risk after the age 4 and 8 y, respectively, in restricted cubic splines analysis[24] controlling for HLA-DR3/4, DQB1*03:02, family history of T1D and ethnicity. These findings suggest environmental, rather than genetic reasons for heterogeneity of IA but the effect of GxE interactions has not been fully explored.

Figure 3: Erythrocyte membrane Ω-3 fatty



2011 Specific Aim 2. To evaluate selected environmental risk factors as potential initiators of IA and/or promoters of progression. During the past four years, DAISY has examined environmental factors associated with initiation of IA[12-14;25] and progression from IA to T1D[7;8]. We have focused on vitamin D[23;26-28] and omega-3 fatty acids[14;29] as potential protective factors, and other components of childhood diet as potentially increasing risk[7;12;30]. Gastrointestinal illnesses in DAISY subjects were associated with an increased risk of IA, but only in interaction with exposure to gluten-containing grains, while upper respiratory illnesses or fever was not associated with IA[25] . Despite priority score at 1st percentile, the direct cost budget for 2011-16 was administratively reduced by $308,000. To balance the budget, we had to curtail the molecular microbiology substudy to explore this association.

Insulin resistance, increased by physical inactivity and central obesity, promotes progression from IA to diabetes, independently of puberty. Please see progress report in Specific Aim 1, Hypothesis 1.3. Approach.

2011 Specific Aim 3. Define the outcomes of early diagnosis and treatment on preservation of endogenous insulin, glycemic control and variability, and quality of life in children with screening-identified T1D Screening-identified T1D patients will continue to have higher C-peptide

levels, lower A1c, lower glycemic variability by CGM, lower incidence of severe hypoglycemia and will report greater health-related quality of life than community controls. Metabolic characteristics of 21 DAISY participants vs. 21 community controls (matched by age, time of diagnosis) were compared at diagnosis of T1D, 6 and 12-month later[31]. Within 2 months of diagnosis, DAISY children, compared to controls, had lower baseline hemoglobin A1c (6.5+/-1.4% vs. 9.2+/-2.9%; P=0.0007) and higher mixed-meal stimulated C-peptide (2.5+/-0.5 vs. 1.6+/-0.2 ng/mL; P=0.02). They also had lower insulin dose-adjusted hemoglobin A1c at 6 months (7.4+/-2.1% vs. 11.2+/-3.5%; P=0.04), but not at 1 y. CGM analyses revealed higher minimum overnight glycemia in DAISY children (119 vs 72 mg/dL; P=0.01). While there was no difference in the frequency of severe hypoglycemia, DAISY participants reported greater health-related quality of life than community controls. Initially favorable patterns of IDAA1c and C-peptide were no longer apparent 1 year from diagnosis; the participants continue to be followed for IDAA1c and hypoglycemia.

Shared Resource for Multiple Projects. Current limited growth of research funding underlines the importance of optimal and extended use of NIH-funded population laboratories, such as the DAISY cohort. DAISY Biobank include aliquots of 30,464 serum, 28,216 plasma, 22,363 throat swabs, 21,475 saliva, 22,305 rectal swabs, (throat, rectal and saliva samples collection was stopped in 2020), 14,493 urine, 7,876 RNA, and 5,892 DNA samples collected over the past 21 years from 2547 subjects. DNA, plasma and serum samples are available from 4,596 of parents and older siblings of DAISY participants. Cord blood plasma and DNA is available from 31,811 general population newborns typed for HLA-DR,DQ. DAISY has and will continue to provide data and samples to multiple investigators in areas beyond diabetes research. Examples of projects that have used DAISY as the core grant in 2011-15 include T1D studies: R01 DK32083 (Eisenbarth), R01 DK49654 and R01 DK104351 (Norris), R01 DK069878 (She) R01 DK068001 (Dabelea); celiac disease R01 DK-50979 (Rewers) and R21 DK084568 (Liu); rheumatoid arthritis R01 AR051394 (Norris); multiple autoimmune disorders U19 AI50864 (Nepom, Eisenbarth, and Holers) and JDRF 33-2008-396, 17-2013-535 (Rewers), 2-SRA-2015-51(Yu); and training grants K12 DK63722, K12DK094712 as well as ADA and JDRF Career Development Awards. DAISY has screened over 6,500 relatives for DPT-1 and TrialNet and has a mutual follow-up protocol for children participating in both studies. DAISY multiplex families and non-diabetic controls participate in the T1D Genetics Consortium (T1DGC).

RESEARCH PLAN

Specific Aim 1. Determine the natural history of islet autoimmunity and to explore the heterogeneity of diabetes and other autoimmune phenotypes in young adults through follow-up of the existing cohort .

SIGNIFICANCE: Prospective birth cohorts of high-risk children[6;32-34] have revolutionized our understanding of the natural history of T1D. While chronic autoimmune destruction of pancreatic β-cells often starts in the initial years of life, it can be triggered at any age up to 15 y; however, no data exist beyond that age. The progression from onset of IA to diabetes usually does not follow a predictable linear model[35]. In most children, IA appears to follow a remitting-relapsing course[36], perhaps due to reactivations of persistent viral infection or recurrent infections of the islets[37-39] and/or the exocrine pancreas[40;41]. Autoantigen epitope spreading may be balanced by partial restoration of tolerance slowing β-cell damage. Ebb and flow in insulin sensitivity and secretion with overweight and puberty are also among potential factors contributing to variable T1D phenotypes (**Figure 4**).

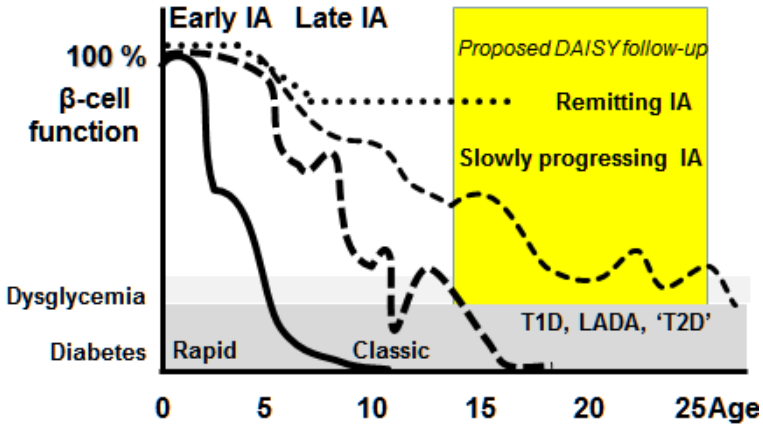


Figure 4: Multiple phenotypes of childhood islet autoimmunity (IA) and their outcomes

Early IA: may lead to a rapid progression ___ to T1D or a classic presentation of T1D _ _ _ school-age with a discernible remission.
Late IA: may lead the classic presentation _ _ or slowly progress to diabetes in adulthood - -

Either early or late IA may remit . . . . without progression to diabetes, but may leave a β-ce defect contributing to type 2 diabetes in adults

Late-onset IA, slowly progressing to diabetes, may lead to overt diabetes in adults, including classical T1D, LADA, GDM and phenotypes masquerading as T2D. It is unknown what proportion of adult diabetes represents slowly progressing or transient IA. The heterogeneity of diabetes in young adults poses a perplexing research and clinical challenge. Extended follow-up of DAISY participants for dysglycemia and diabetes will help to improve classification, diagnosis and possibly therapeutic targets for these patients.

Interestingly, the development of IA continues unabated beyond early childhood (Fig 1). This proposal will extend the incidence estimates to age 25 and will increase the precision of estimates by ~20%. IA levels fluctuate from positive to negative and back to positive in some children. While we do not fully understand the biological reasons for the remitting-relapsing course of IA[36], our findings show that a much larger number of children than previously thought have an extended period of IA, without rapidly progressing to T1D. These children may develop diabetes as adults or escape diabetes altogether by re-establishing tolerance to islet autoantigens. If the latter is true, understanding the pathomechanisms could greatly help development of prevention and, possibly, cure.

The proposed 4-year renewal of this project will extend the scope of this study to T1D cases expected between ages 15-25 y - an age-range previously not studied by prospective cohorts. During follow-up, ~50% of the study participants will achieve 25 y of age allowing for a comprehensive assessment of the natural history of T1D throughout adolescence and early adulthood. Identification of the environmental cause(s) of T1D requires prospective assessment of multiple exposures prior to and after the development of IA. Individuals with persistent IA may benefit from interventions to prevent T1D or if unavoidable, limit risk of life-threatening DKA at diagnosis[42]. Persons who do not progress to T1D, despite long-term persistent IA, may help us to understand mechanisms for regaining tolerance to autoantigens or islet regeneration. T1D shares genetic determinants with several common autoimmune diseases: celiac disease, autoimmune thyroid disease, and the less frequent rheumatic arthritis and Addison's disease. Progress in multiplexing immune assays and genotyping has made feasible an integrated screening for of these diseases. The DAISY population is ideal for collection of data needed to model the cost-effectiveness of such a program, based on different sets of biomarkers, age at the initial screening/ re-screening, and the outcomes of early detection and treatment.

INNOVATION: We are proposing to continue this highly productive prospective study to gain information concerning the development of T1D and other autoimmune diseases in young adults – an age group that has never been studied. DAISY is serving as a large vanguard study for the international TEDDY consortium, which will not be able to address the questions posed in this application for another 10 years. DAISY has discovered that IA can develop in older children and adolescents and that the predictors of late- vs. early-onset IA differ. This observation can be of a paramount importance to our understanding of the etiology of T1D diagnosed after age 18 (~half of T1D patients) and may have a profound effect on design of future screening and prevention programs. Screening of young children for several polygenic autoimmune diseases with variable incubation period, poses significant logistic questions, surpassing those previously encountered in

newborn screening programs[43;44]. This study is likely to exert a sustained, powerful influence on diabetes research by using prospectively defined immuno-genetic phenotypes to advance both personalized medicine and public health approaches to screening and early interventions.

APPROACH: We will follow the already established cohort of youth at high risk of T1D and other autoimmune diseases (current n=1149, median age 17.2 y, IQR 13.5-20.2 y) to estimate the overall burden of pre-clinical and clinical T1D, celiac, thyroid, adrenal, rheumatic and parietal autoimmune disease in Colorado by age 25. This will inform future screening and prevention programs. We will further explore the apparent heterogeneity of IA and its implications for adult-onset diabetes.

Study Population: Since 1994, DAISY has studied two large cohorts of high-risk children: A) 1,424 general population newborns with high-risk HLA-DR,DQ genotypes; and B) 1,123 non-diabetic relatives of patients with T1D. Cohort to date: 222 persistent IA and 94 progressed to T1D, 5 died, 3 disenrolled due to medical conditions, 265 lost to follow-up; 2180 participants are being tracked and 1149 complete study visits.

General Population High-risk Children During 1993-2004, 31,881 newborns from the general population of Denver, Colorado were screened for HLA-DR,DQ genotypes carrying susceptibility to T1D[17;45]. 90% of all children born at the St. Joseph's Hospital have participated, representing closely the ethnic breakdown of the Denver Metropolitan Area and including children classified by their mothers as non-Hispanic white (56%), Hispanic (30%), African American (7%), Asian American (2%) or biracial/other (5%). Throughout this application, "DR4" denotes DR4,DQB1*0302. All children with DR3/4, DR3/3 and DR4/4 and a sample of those with DR4/x or DR3/x were invited to participate in follow-up at ages 9, 15, and 24 months and yearly thereafter; 1,424 completed at least one visit.

Non-diabetic Relatives Non-diabetic offspring and siblings of patients with T1D were invited to participate through the BDC clinics and regardless of their HLA genotype; 1,123 enrolled, at median age 1.4 y and are followed identically to the general population cohort.

Islet autoantibodies to insulin (IAA), GAD65 (GAA), IA-2 (IA-2A), and ZnT8 (ZnT8A) were measured at all visits and if positive, in 3-6 month intervals. IA is defined as the presence of one or more of the autoantibodies on at least two consecutive visits 3-6 months apart or development of T1D after a positive test. Diabetes is diagnosed using the ADA criteria [46].

Progression from Islet Autoimmunity (IA) to T1D

Hypothesis 1.1: Persistent IA triggered after age 6 y results in T1D diagnosis in early adulthood

Hypothesis 1.2: Transient childhood IA does not increase the risk of diabetes in early adulthood

As of 10/2015, 94 children participating in DAISY follow--up have developed T1D. All but one had expressed IA prior to or at diagnosis. Over 70% of children expressing two or more autoantibodies progressed to T1D in 10 years since seroconversion[3], compared to 15% of those with one autoantibody. There was little difference in progression by family history of T1D or presence of the HLA-DR3/4,DQB1*0302 genotype. PTPN22 and UBASH3A and INS predicted diabetes, controlling for family history of T1D and presence of HLA-DR3/4-DQB1*0302 genotype. In general population children, risk alleles at these 5 genetic markers conferred combined HR of 13 for IA and >40 for T1D[11]. INS, UBASH3A, and IFIH1 were significantly associated with progression from IA to T1D, while PTPN22 and IL27 showed a borderline association[47]. Age of appearance of first autoantibody and mean IAA levels during the follow-up were major determinants of the age at diagnosis. In contrast, levels of GAA or IA-2A did not predict time to progression to diabetes[10].

New assays for insulin/proinsulin and GAD autoantibody detection: Serial serum samples from DAISY participants followed from birth who progressed to T1D were essential in development of novel assays for insulin/proinsulin autoantibodies (IAA), GADA and TGA. These nonradioactive electrochemiluminescence-based assays have been developed on a platform from MesoScale Discovery (MSD). The ECL-IAA assay detects onset of IA earlier than any current standard radioassays in 24% of children, thus defining more precisely the onset of IA and timing of potential triggering events[48;49]. Both ECL-IAA and ECL-GADA assays are more specific for high-affinity autoantibodies that predict progression to T1D[48-51]. Next, the individual ECL assays were adapted to measure four antibodies in a single well using a small blood volume (6ul). Using a MSD QuickPlex 4-Spot plate or U-Plex plate, we successfully combined ECL-IAA, GADA and IA-2A with transglutaminase autoantibodies (ECL-TGA) in a single well of a 96 well plate. The multiplex ECL

assay was more sensitive than current radioassays for IAA and TGA and will facilitate high-throughput screening for T1D and celiac disease risk.

Risk of T1D in subjects with a single islet autoantibody

Nearly a half of children with IA persistently express a single autoantibody: 46% in DAISY, 45% in combined DAISY, BABYDIAB and DIPP cohorts[3] and 40% in TEDDY[52]. Fewer than 15% of these children progress to diabetes in 10 years[3].

Those who do progress, typically have high-affinity islet autoantibodies[48;53;54]. The ECL-based IAA and GADA assays detect only high-affinity autoantibodies and efficiently classify single islet autoantibodies into high-risk vs. no risk (**Figure 5**).

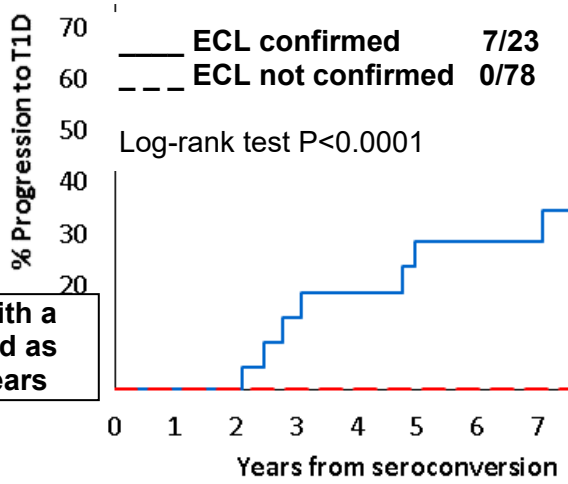


Figure 5: High rate of progression to T1D among subjects with a single persistent islet autoantibody by radioassay, confirmed as high-affinity by ECL; 29% and 35% risk of T1D in 5 and 10 years

Laboratory Serum samples are stored at -70°C prior to testing. All measurements are performed on coded samples with blinded duplicate testing of all positive and 5% of negative samples. Islet autoantibodies are measured[55-57], in the laboratory of Dr. Eisenbarth (the reference laboratory for TrialNet, TEDDY and T1DGC). In the Diabetes Autoantibody Standardization Program (DASP) workshop, sensitivity and specificity were, respectively, 72% and 99% for GAA, 62% and 99% for IA-2A, 62% and 99% for ZnT8A, and 48% and 100% for IAA. Transplacental autoantibodies are excluded. TG IgA autoantibodies[58] are measured at all visits (by R01 DK-50979). TPO (with TSH and T4, if positive), autoantibodies to 21OH (adrenal) and to cyclic citrullinated peptide (rheumatoid arthritis) - will be re-measured only once at age 20 or the end of the follow-up. Standard methods will be used to measure: HbA1c, blood glucose, fasting total-, HDL-cholesterol and triglycerides. Adiponectin, insulin and C-peptide will be measured using radioimmunoassays.

Assessment of dysglycemia Participants with persistent IA are asked to complete an oral glucose tolerance test (OGTT) annually. Earlier diagnosis and treatment of T1D may be also possible by applying continuous glucose monitoring (CGM). In a case-control study nested in DAISY[59], we documented that CGM can detect early hyperglycemia in children with persistent IA and no symptoms of diabetes. More than 20% of CGM time spent above 140 mg/dL predicted progression to diabetes. The IA cases had mean HbA1c of 5.5% (37 mmol/mol), however, their average maximum daytime glucose value was higher, and they had increased glycemic variability on CGM. We will record CGM for 1-2 weeks after each clinic visit in all participants with IA to monitor evolution of dysglycemia during progression to T1D (letter of collaboration from T. Walker, Dexcom).

<u>Hypothesis 1.3: Sedentary life-style and insulin resistance predict faster progression from persistent IA to T1D</u>

SIGNIFICANCE: The environmental factors influencing the risk and rate of progression from IA to T1D remain poorly understood; accelerated growth[60] or insulin resistance[61-63], due to obesity[64], puberty or physical inactivity, could contribute to risk. Habitual physical activity is associated with lower HbA1c, higher c-peptide, and lower insulin requirements at diagnosis of T1D and up to 2 y later[65]. To fully ascertain predictors of progression to T1D (related also to non-autoimmune forms of adult diabetes), anthropometry, dietary factors (glycemic index, intake of fat and added sugars), pubertal stage, and physical activity must be measured prospectively and repeatedly. We hypothesize that IR, increased by puberty, central obesity, poor diet and physical inactivity, will predict progression from IA to T1D.

INNOVATION: Most prior studies of growth, obesity, insulin resistance and progression to T1D have examined children prior to or at puberty. However, less is known about the relationship

between the evolution of these factors that tends to occur among youth and young adults and progression to diabetes. As DAISY study participants reach that age, this is an ideal time to examine how changes in adiposity, physical activity and insulin sensitivity may contribute to the risk of T1D.

APPROACH: ***Preliminary studies*** DAISY added measurement of waist circumference and physical activity questionnaire[66] at each visit starting in June 2010; these measurements are available on 1,090 subjects. Pubertal stage has been determined based on self-reported Tanner Stage at each visit [67;68]. In 2011, we began collecting objective physical activity data using an accelerometer - a small device worn on the hip or lower back to measure change in velocity over time[69] - worn for 7 days. We have identified 126 antibody positive subjects as eligible for this more intensive measurement of physical activity. We have obtained accelerometer data for 82 subjects, 69 of which have worn the accelerometer more than once; 8 have later developed T1D. We have also obtained accelerometer data one time on 55 antibody negative controls. Multivariate linear regression analysis, adjusted for age, sex, BMI, and total time recorded has shown that IA subjects spent a higher percentage of time sedentary than controls (p=0.045), less time engaged in vigorous physical activity (p=0.01), and took fewer total steps (p=0.02) and steps per minute (p=0.03). In a cross-sectional analysis, these differences translated into IA subjects spending over 110 extra minutes sedentary and 70 fewer minutes engaged in vigorous physical activity per week, compared to controls. In March of 2020 DAISY stopped offering accelerometers to participants that are antibody positive or are antibody negative controls.

Prospective dietary data We will continue to measure dietary intake of macronutrients, glycemic index and added sugars using a validated Youth/Adolescent Questionnaire[74] completed by both parents and child between ages 10-12 and by the participant after age 12 on the entire DAISY cohort. In the IA positive subjects and the controls,

Anthropometry Weight will be measured using a calibrated digital scale to the nearest 0.1 kg, height using a wall-mounted stadiometer to the nearest 0.1 cm. Minimal waist circumference will be measured using a Figure Finder tape (Novel Products, Rockton, IL)[78]. Body fat percentage will be measured using bioelectrical impedance analysis (TANITA Dc-430 U Dual frequency Total Body Composition Analyzer), which is a fast, inexpensive and low-risk alternative to DXAy[79]. Fat distribution may be important in the development of type 1 diabetes (T1D). By measuring body fat in DAISY, we can better analyze factors related to being overweight or obese.

Specific Aim 2. Validate candidate proteomic biomarkers of IA and T1D, in a nested-case study of 213 youth with persistent IA and 213 controls.

SIGNIFICANCE: Progress in proteomic technology has made it applicable to epidemiologic studies. Prospective evaluation of the proteome contemporary with development of IA, or progression to T1D can help to discover underlying pathways and potential therapeutic targets. Cross-sectional studies of serum proteome at diagnosis of T1D have suggested, not surprisingly, involvement of inflammation and innate immunity[83-85]. However, samples used in these studies were collected shortly after T1D diagnosis, reflecting late stages of β-cell destruction and hyperglycemia. In contrast, prospective cohorts allow profiling of the proteome as children develop IA and progress to T1D and in aged-matched controls[86]. Several published and unpublished discovery studies have generated multiple protein candidates, but little consistency. This highlights the need for replication in larger studies, able to control for heterogeneity of pre-T1D phenotypes and confounders, e.g., age and sex, but also for genetic background and environmental exposures. DAISY has collected a unique dataset and serum/plasma samples for the high-power targeted protein analysis proposed below. The goal of this study is to help form consensus regarding proteomic biomarkers of IA and T1D and underlying pathways. The analytical platform proposed here is capable of quantification of low-abundant proteins, likely to be found in the earlier stages of disease, that can be masked by more abundant proteins[87].

INNOVATION: So far, none of the 'omics biomarkers reported to predict progression from IA to T1D have been confirmed by independent studies. Using targeted proteomics we will validate candidate peptides/proteins reported from several discovery studies. The analyses will cover an

unprecedented age-span of diabetes diagnosis (0-25 y) while controlling for an apparent heterogeneity of IA phenotypes, age, sex and ethnicity.

APPROACH: We will characterize differences in the abundances/profiles of ~75 target proteins in a nested case-control study of 213 youth persistently positive for IA and 213 controls, frequency matched on age, sex and ethnicity (79% NHW). As of 9/2015, 89 cases have developed T1D (at mean age 9.6±4.6 y), 61 were persistently positive for ≥2 islet autoantibodies or one high-affinity autoantibody and 63 were positive for one low-affinity autoantibody (**Figure 6**). These three subgroups of youth represent heterogeneity of the IA phenotypes that show, respectively, rapid, slow and no progression to T1D (see Figures 4 and 6, above).

Hypothesis 2.1: A reproducible set of proteomic biomarkers will predict development of IA independently of demographic factors and HLA class II genotypes.

Hypothesis 2.2: A reproducible set of proteomic biomarkers will predict the rate of progression from IA to T1D independently of demographic factors, HLA class II genotypes.

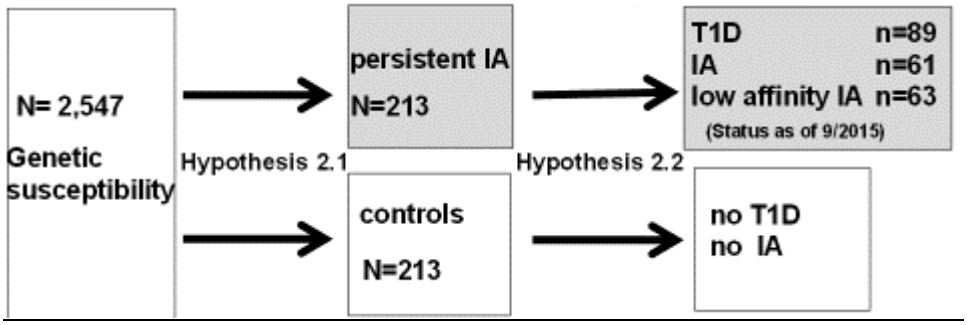


Figure 6:
Proposed nested case-contro study of targeted proteomics.

Serial samples collected at 4- time points per subject will be analyzed.

Preliminary data: Two nested cases-controls DAISY proteomics discovery studies have identified proteins associated with development of IA and/or progression to T1D (**Table 3**). In ***DAISY1***, one of the 4 proteins identified by the aptamer-based approach[84] as strongly associated with IA and T1D (CNDP1, p=1.22e $^{-07}$, q-value 0.00013) was independently confirmed in a different laboratory using the LC-MRM/MS approach[88].

Table 3: Sources of candidate protein targets for validation in this project

	DAISY1		DAISY2	DIPP[86]	Additional sources
Population	**25 T1D 20 IA+ and non-DM 25 controls**		**11 T1D 15 IA+ non-DM 10 controls**	**13# + 6* T1D 13# + 6* controls**	**TEDDY discovery[89] BetaMarker[90] other [83;85;91-94] BABYDIAB (Hauck et al. unpublished)**
Time points	**4**		**9**	**5-11**	
Material	**serum**		**plasma**	**serum**	
Approach	**aptamers**	**LC-MRM /MS**	**10-plex iTRAQ LC-MS/MS**	**8-plex iTraQ# LFQ* LC-MS/MS**	
Laboratory	**Somalogic**	**PNNL**	**PNNL/UNCG**	**Turku CB**	
Candidate targets	**4**	**1**	**142**	**15**	

In ***DAISY2***, with Dr. Qibin Zhang (previously at PNNL) from the University of North Carolina at Greensboro, we identified 2118 protein groups at false discovery rate of 1%, and quantified 1947 proteins based on at least two razor/unique peptides; 682 proteins were identified and quantified in all samples. Half (1045) of the proteins identified were also found in cultured islet cells[93] and the pancreatic head tissues (Liu at al. unpublished data). The expression patterns of 142 proteins differed ($p<0.05$) between control and T1D cohort, including 31 that showed clear difference before the appearance islet autoantibodies. Pathway enrichment analysis of those 142 protein groups, using DAVID bioinformatics tool and all 2118 plasma proteins identified indicated involvement of

proteasome (p=0.015) and complement and coagulation cascades (p=0.048); the latter has important roles in innate immunity[84;86].

The only similar study published so far for pre-T1D from DIPP[86] has identified 658 proteins at 5% FDR and used 220 proteins for quantification. Comparison of the DAISY2 and DIPP findings showed some overlap with ADIPOQ, TGFBI, C4B, CNDP1, GPX3, APOA4, IGFBP2 and PGLYRP2 as possible candidate protein for a validation study. However, this list will grow as additional targets emerge (**Table 3**, Additional sources)

Plasma proteomics of The Environmental Determinants of Diabetes in the Young (TEDDY) consortium. Dr. Metz's PNNL lab is currently analyzing the plasma proteome of 368 individuals enrolled in the TEDDY program, searching for signatures of pancreatic β-cell death, systemic responses to autoimmunity and development of type 1 diabetes. So far, samples from 180 individuals were analyzed by LC-MS/MS, which led to the identification of over 19,000 peptides derived from approximately 1,500 proteins. Data capture is still ongoing, and a statistical analysis has not yet been performed.

***Signatures of pancreatic β-cell stress - Beta Marker[90]* -** NIH-funded project of the Human Islet Research Network (HIRN) (R Mirmira & T Metz, PIs). To study pancreatic β-cell stress and potential mechanisms of death, a proteomic analysis of pancreatic islets treated with the cytokines interferon-γ and interleukin 1β is being performed. Preliminary data from the analysis of samples from 4 donors led to the identification of 4,810 proteins, of which 47 were differentially abundant after cytokine treatment.

Target selection. Targeted proteins will be selected based on three criteria: (i) ***high priority*** candidates that are consistently statistically significant in multiple cohorts (and DAISY, DIPP, TEDDY, BetaMarker, BABYDIAB and others), (ii) ***moderate priority*** candidates significantly different in only one study, but with a high fold change, and (iii) ***low priority*** candidates based on significant differences only in one or more studies. We anticipate that a subset of the peptide/protein markers (approximately 150 peptides corresponding to 75 proteins) will be assayed in targeted measurements.

Targeted proteomics analysis will be carried out at the laboratory of Dr. Thomas Metz at the Pacific Northwest National Laboratory (PNNL) in Richland, WA (see letter of collaboration). For the targeted proteomics analyses proposed in this study, we will use selected reaction monitoring (SRM) mass spectrometry (MS). SRM measurements will be performed using liquid chromatography coupled with triple quadrupole MS and SRM transitions specific to each targeted peptide. Transitions are defined by selecting peptide precursor ions in the first quadrupole, fragmenting them in the second quadrupole, and selecting specific peptide fragment ions in the third quadrupole. Such selection of peptide precursor and fragment ions greatly reduces background noise; combined with peptide retention times from analyses of authentic, stable isotope labeled standards, this method ultimately improves measurement specificity and sensitivity and enables the detection of proteins in the range of ng/mL in serum/plasma[95;96].

Robust quality control (QC) practices Dr. Metz's laboratory has developed and employed various quality control practices to ensure the accuracy and reproducibility of proteomics data collected in the laboratory. In addition to these routine quality control practices, stable-isotope labeled reference peptides are included as internal standards in each sample to be analyzed depending on the type of analysis and experimental design, thus obviating the need for separate quality control samples within sample batches.

Informatics tools All collected instrument raw data files will be automatically captured by the PNNL Proteomics Laboratory data management system[97]. Acquired SRM datasets will be processed with Skyline software[98], and ratios between the endogenous peptides and their SIS analogs will be used as peptide abundances for statistical analysis.

Limitations/potential pitfalls and alternative approaches Possible limitations/pitfalls include (i) peptides that do not provide ideal transitions for SRM measurements and (ii) lack of sensitivity. For peptides without ideal SRM transitions, we plan to test multiple peptides per protein if needed and will choose only peptides with good response. In case of low abundant candidates, an alternative approach would be to analyze samples by PRISM (high-pressure, high-resolution separations coupled with intelligent selection and multiplexing). PRISM is based on pre-fractionation of peptides prior to LC-SRM measurements, which immensely increase the sensitivity of the analysis enabling

the detection of as low as picograms of proteins per milliliter of plasma/serum[96;99] . However, PRISM requires much more instrument time, limiting the number of samples that can be analyzed. Thus, it will be only used in cases of very promising biomarker candidates.

Specific Aim 3. Develop an integrated comprehensive model of the relapsing-remitting process leading to T1D in youth using prospectively collected DAISY 'omics data.

SIGNIFICANCE: Machine learning algorithms are currently used for predictive modeling across an array of biomedical domains. Some of the most popular include Random Forests (RF) and Support Vector Machines (SVM); however, traditional algorithms such as Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and K-Nearest neighbors are still often used. These algorithms have various levels of success when faced with qualitative variables and missing data[102-104]. Dr. Webb-Robertson has developed a methodology that uses the benefits of simple naïve Bayesian integration with the power of individualized algorithms best suited to specific datasets often referred to as meta-learning or ensemble methods[105-109]. The probability of the outcome of an individual is the combined evidence from multiple datasets, each modeled independently. This approach allows predictors from different algorithms to be utilized in another machine learning algorithm, NB in this case, to relate the datasets to one another in the context of a common outcome. Bayesian integration ensemble feature finder (BIFF) uses a statistical optimization algorithm known as simulated annealing. A major issue with feature selection algorithms is they return a single solution via greedy optimization of a defined criteria, such as the area under a ROC curve. However, this is only one of likely many potential solutions, especially when the feature space is large and disparate. The BIFF approach allows the algorithm to move out of local minima by updating the solution at each iteration, based on the current feature state and sampling in proportion to including or excluding the variable of interest[110]. Integrative Bayesian modeling, based on a set of disparate features e.g., gene variants, proteins, or metabolites, will be used to generate individualized prediction algorithms in IA progressors vs. non-progressors. The DAISY population with its rich longitudinal data set is ideal for the development of statistical machine learning models that can explore individualized and subgroup predictors of IA development and progression from IA to T1D. In addition, pairing clustering with selected features will offer a new approach to evaluate features in the context of pathways. The major advantage of this study is the prospective characterization of autoimmunity over a prolonged period of time, with high-frequency of repeat measurements of biomarkers. This approach will improve prediction of T1D and may identify new targets for prevention.

INNOVATION: As the number of features increases the standard approach to feature selection as a component of predictive modeling becomes computationally challenging and prone to over-fitting, i.e., for a dataset with k features, there are 2^k possible subsets of markers. The BIFF algorithm is an innovate approach to sample the feature space in the context of the integrated solution in combination with repeated cross-validation. BIFF allows each clinical feature or biomarker to be quantified by the likelihood that it would be included in the model. Teaming these feature selection uncertainty estimates with clustering can identify key subsets of features that can be used for pathway modeling in lieu of frequentist statistical thresholds.

APPROACH: Statistical analyses will be carried out by Dr. Bobbie-Jo Web-Robertson (PNLL) – see letter of collaboration. In addition to the proteomic data described in Specific Aim 2, we will have extensive data on the metabolome, epigenome and transcriptome of these 213 cases and 213 controls, which has been separately funded (R01 DK104351, J. Norris, PI). Briefly, the **metabolome** is being measured at the UC Davis (O. Fiehn) using three panels (CSH-QTOF MSMS method for general lipidomics, TOF MS for primary metabolism, and HILIC-QTOF MS/MS for specific and dietary metabolites. **Epigenetic profiling** is being done using the Illumina Infinium 450k methylation array platform at the Genomics and Microarray Core Laboratory at the University of Colorado. The **transcriptome** of some of the cases and controls has been previously characterized[111] using HumanRef-8 Expression BeadChip (Illumina, San Diego, CA) with validation using high-throughput Real-Time PCR. Additional expression studies on the 213 cases and controls are being conducted at the University of Colorado core laboratory using the same technology. **Genome:** DAISY has genotyped all study participants for HLA-DR,DQ, 108 non-HLA

SNPs[11;47] and ImmunoChip's >200,000 SNPs in non-Hispanic white participants with HLA-DR3/4. In a separately funded Helmsley Trust project, DAISY will apply the whole genome sequencing (WGS) by Illumina X Ten sequencing (Macrogen, Rockville, MD) to the 213 cases and controls included in Specific Aim 2. The interpretation of WGS results will be informed by the results of the Type 1 Diabetes Genetic Consortium (T1DGC) WGS (also by Macrogen) and genetic architecture of T1D[112]. The analyses will be performed by Dr. S. Rich, from the Center for Public Health Genomics at the University of Virginia and past PI of the T1DGC. His lab has completed the DAISY ImmunoChip analyses and we also collaborate on TEDDY[113]. **Dietary and infectious exposure** data collected by DAISY will be used in the analyses as a distinct set of features to explore in relation to the above 'omic domains.

Hypotheses: A limited number of features and time-points derived from high-resolution 'omics data can reliably predict, on an individual basis, 3.1: development of IA; and 3.2: the rate of progression from IA to T1D

Preliminary Evaluation As a proof-of-principle, modeling was performed on data from the DAISY1 (Table 3) nested case-control study of subjects who developed diabetes (T1D, n=25), persistent IA but not T1D (IA+, n=20), and 25 healthy age-matched controls. Four time points were studied: T1-the earliest available sample, T2-just prior to IA, T3-just after development of IA, and T4-prior to diagnosis of T1D or the most recent. Demographics (age, sex, ethnicity, family history of T1D), genetic (HLA-DR,DQ and 108 non-HLA SNPs), immune, metabolic and proteomic biomarkers were compared between the groups across the 4 time points.

Metabolites The global metabolic profiling combined three independent *platforms*: ultrahigh performance liquid chromatography/tandem mass spectrometry optimized for basic and acidic species, and gas chromatography/mass spectrometry (Metabolon, Durham, NC). A total of 382 named metabolites were included in this analysis. An example of this study results is shown in **Figure 7**.

BIOCHEMICAL	T1			T2			T3			T4		
Cofactors and Vitamins	AbPos C	T1D C	T1D AbPos	AbPos C	T1D C	T1D AbPos	AbPos C	T1D C	T1D AbPos	AbPos C	T1D C	T1D AbPos
Vitamin C	0.63	0.41	0.66	1.07	1.46	0.86	1.01	0.75	0.75	0.83	1.19	1.66
threonate	0.94	0.7	0.75	1.02	1.07	1.05	1	1.03	1.04	1.01	0.94	0.93
beta-tocopherol	1.62	1.34	0.83	1.12	0.96	0.86	0.89	1.16	1.31	0.89	1.54	1.72
delta-tocopherol	1.46	0.94	0.64	1.12	0.79	0.71	0.64	0.89	0.93	0.77	0.97	1.26

Figure 7: Lower levels of vitamin C at T1 (infancy) predict development of IA.

Proteins/peptides Relative abundance of 1001 serum proteins measured by aptamers (Somalogic, Boulder, CO)[84] and 49 peptides (representing 24 proteins) measured by LC-MRM/MS[88] in DAISY1 study were included.

We first compared BIFF to standard Recursive Feature Elimination (RFE) approaches in combination with cross-validation (CV). RFE is a method using extensively in biology in combination with various machine learning algorithms, such as LDA and SVMs[107;108;114-117]. RFE is readily available in most statistical programming languages and is simple to implement. It is a greedy algorithm that sequentially eliminates features to maximize AUC.

In respect to the hypotheses of interest, comparing the early time points, T2 and T1, for development of IA, the control group was compared to the combined group of subjects with IA+ and T1D (Hypothesis 3.1); for progression to T1D (Hypothesis 3.2), T3 and T4 were compared between the IA+ and T1D groups.

Figures 8A and B show the results of these comparisons, respectively, as a ROC curve for the SA approach at a 50% frequency selection. For comparison purposes we also evaluated these two comparisons with standard repeated recursive feature elimination (RFE) and selected features similarly based on percentage of times the features were selected at either 25% or 50% frequency, as well as all features. There is a clear advantage to the integrated feature selection approach of BIFF (SA_50). Features selected are quantified based on the percentage of times selected, and ordered based on this percentage. This allows the features to be evaluated in the context of the integrated model to determine the key biomarker panel in lieu of specific datasets.

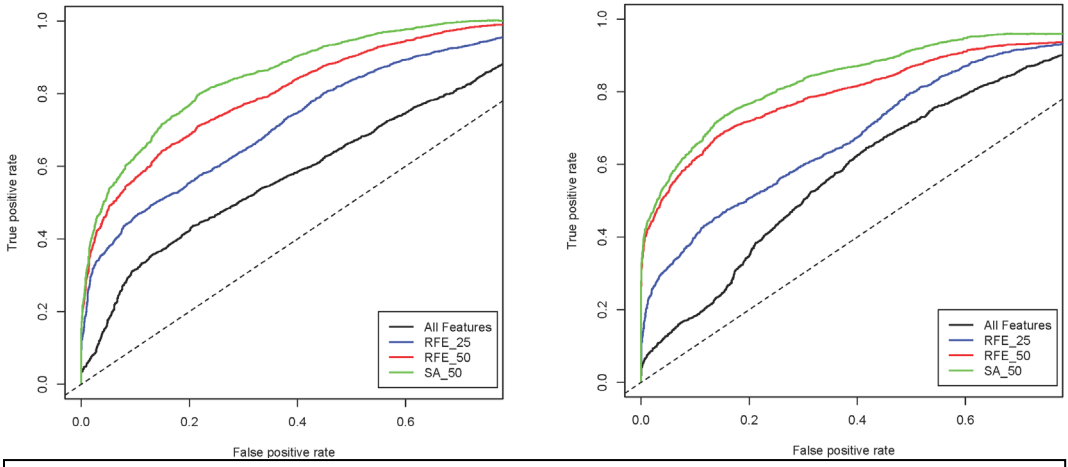


Figure 8: ROC curves comparing A: controls to IAB+ and T1D subjects at early time points (T2,T1); B: IA+ compared to T1D at late time points (T4,T3).

The top features predicting **development of IA** included mostly proteins, however metabolites were the top two features (**Table 5**). While we would caution against over-interpretation of these very preliminary data, selection of PTPN22 and CTLA-4 SNPs was reassuring, while the high ranking of vitamin C levels (consistent with Figure 7 data) and 4-hydroxyhippurate – a microbial-origin xenobiotic was intriguing. Metabolic features predicting **progression to T1D** were consistent with dysglycemia and lipolysis. Complement and coagulation cascade proteins lit up as potential predictors of either endpoint, consistent with previous reports[84;86] and DAISY2 results (above).

Table 5: The top features, a mixture of metabolites, proteins and SNPs (metabolites highlighted)

Development of IA		**Progression to T1D**	
Feature	**% selected**	**Feature**	**% selected**
vitamin C	**100**	glucose	**100**
3-methyl-oxobutyrate	**94**	ADS fibrinogen	**100**
ENTP5	**93**	C3adesArg	**100**
4-hydroxyhippurate	**91**	mannose	**99**
IL-4 sR	**90**	contactin-5	**98**
rs2476601_G PTPN22	**89**	RAC1	**98**
pyroglutamine	**89**	MAPK2	**97**
SPTA2	**89**	rs3117103_A	**92**
Factor H	**87**	rs7221109_11SN_C	**91**
ULBP-3	**86**	plasmin	**90**
PTPN22_R620W_C	**85**	ribose	**89**
rs3087243_G CTLA-4	**84**	FYN	**89**

IV. Research Methods

A. Outcome Measure(s):

Primary outcome measures of the DAISY study is pancreatic beta-cell autoimmunity, islet autoimmunity (IA). Antigen-specific radioimmunoassay have been developed and improved, in

specificity and sensitivity, throughout the course of the DAISY study. The detection of GAD, IAA and BDC512 autoantibodies from serum is an indicator of a physiological change and the triggering of the autoimmune process.

The secondary outcome measure is development of TID, diagnosed by standard criteria[144], determined by fasting blood glucose level or 2 hour oral glucose tolerance test level and hemoglobin A1c measurement. Other signs and symptoms related to diabetes may also be indicated: increased thirst, frequent urination, recent unexplained weight loss.

B. Description of Population to be Enrolled

Inclusion Criteria

General Population Cohort Screening Criteria

- Child born at St. Joseph Hospital between 1/1/1993 and 6/30/2006
- Informed consent for genetic screening obtained
- Cord blood sample available
- No other severe co-existent condition
- At least one parent/guardian speaks English

General Population Cohort Follow Up Criteria

- Informed consent for Follow Up obtained
- Initial telephone interview at 3 months of age, initial blood draw at 9 months of age
- High or moderate risk of T1D, as determined from genetic screening (300 low risk children also enrolled as a control group)

Risk categories-odds of T1D by the age of 20 yrs

- High- 1:16
- Moderate- 1:230
- Low- 1:300

Sibling/offspring Cohort Enrollment and Follow Up Criteria

- Informed consent for follow up obtained
- Age < 4 yrs (< 7 yrs prior to 1997)
- No other severe co-existent condition
- At least one parent/guardian speaks English

Family Members Follow Up Criteria

- Informed consent obtained
- First degree relative of either a sibling/offspring participant or general population newborn participant
- No other severe co-existent condition
- Ages 12 months to 65 years

TEDDY Study Siblings Follow Up Criteria

- Informed consent obtained
- Full or half sibling of a TEDDY Study participant
- Demonstration of IA through the parent/sibling DNA protocol with the TEDDY Study
- Ages 12 months to 65 years

TEDDY participants Follow Up criteria

- Age 15 years or older
- Informed consent obtained
- Demonstration of IA through the protocol with the TEDDY study

Exclusion Criteria

Individuals were not eligible to be enrolled into the DAISY study General Population cohort if cord blood was not available for genetic screening, if other severe co-morbidities were

indicated, if parents refused consent to screening or follow-up, if at least one parent or legal guardian did not speak English or if T1D had been diagnosed.

Individuals were not eligible to be enrolled into the DAISY study Sibling/offspring cohort if they did not have a first degree relative with T1D, if other severe co-morbidities were indicated, they were older than 4 years of age (older than 7 years of age, before 1997), if parents refused consent to follow-up, if at least one parent or legal guardian did not speak English or if T1D had been diagnosed.

Individuals that do not agree to the long-term storage of data and specimen samples will be excluded.

Duration of the Study

Each subject will participate in the study until the index child develops T1D, reaches the age of 40 years- or until funding is no longer available. At age 20, if the subject has developed IA, the study will continue to follow them and monitor IA status, as these individuals will continue to be at an increased risk for T1D.

Current Study Population

The study population was selected from among the eligible population in the Denver Metropolitan are as described in the section above. Approximately 50% of the participants are female. The ethnic distribution in the general population newborn cohort is representative of the general Denver Metro area population: 7% African American, 30% Hispanic, 56% non-Hispanic white, and the remaining of biracial or other ethnicity. The ethnic distribution of the sibling/offspring cohort is representative of T1D in the general U.S. population: 5% African American, 10% Hispanic and 85% non-Hispanic white.

C. Study Design and Research Methods

1. Recruiting Methods:

Data and sample collection began in January 1994 and has been carried out in the study clinic in the Department of Preventive Medicine & Biometrics, UCHSC (currently the UC Denver AMC School of Medicine-Barbara Davis Center in Aurora, CO) and in St. Joseph's Hospital in Denver. Study laboratories are located in Denver (the Barbara Davis Center and the Pediatric Infectious Diseases), in Alameda, CA (HLA typing lab) and in Atlanta, GA (CDC).

Sibling/offspring cohort (SOC)

Recruitment. Starting February 11, 1994, young siblings and offspring of a person with T1D were recruited from: i) families of diabetic children aged 0-17 identified in Colorado between 1978 and 1991 (Colorado IDDM Registry); ii) families of diabetic children seen in the Barbara Davis Center or The Children's Hospital after 1991; and iii) media publicity.

General population newborn cohort (GPNC)

Newborn HLA screening. HLA screening was open to all children born in St. Joseph's Hospital in Denver from January 20, 1994 until June 30, 2006. The newborns screened were representative of the general population of the Denver Metropolitan Area. About 75% of the newborns were members of the Kaiser Permanente Colorado HMO, while the others represented a mix of Medicaid, uninsured and privately insured families. DAISY has developed a reliable and valid genetic screening for HLA markers associated with T1D. The screening uses cord blood and the PCR/SSO technology and is acceptable to 94% of parents in the multi-ethnic Denver population[30]. This work has yielded the first reliable estimates of the population frequencies of the HLA genotypes associated with T1D susceptibility and protection in the four major U.S. ethnic groups[30]. After a routine clinical sample of cord blood was collected, approximately 10 ml of cord blood was obtained for DAISY screening and storage. Informed consent for screening was obtained from the

parents in the hospital after the birth. Blood samples of newborns for whom informed consent was not obtained were destroyed. Cord blood was aliquotted into two 0.5 ml samples: 1) the primary typing sample remains at room temperature for a period of 1-14 days before DNA preparation; and 2) the quality control sample was refrigerated. Samples were sent in bi-weekly batches to the HLA typing laboratory at Roche Molecular Systems, Inc., in Alameda, CA. Whole blood (15-25 μl) was used in PCR amplification with two primer sets to co-amplify the DRB1 and the DQB1 locus. The details of PCR amplification and hybridization with oligonucleotide probes have been published[30].

Family Members

Parents and siblings of the General Population cohort and sibling/offspring cohort are generally studied only once since they are not eligible for prospective follow-up. However, parents or ineligible siblings who are IA positive are offered autoantibody, HbA1c and random glucose testing every 3-6 months to avoid acute onset of diabetes. Families will receive home glucose monitoring education and will be provided with a glucometer to test blood sugars as recommended. The family member may be offered a continuous glucose monitor (CGM) for close monitoring of blood glucose levels if they have an HbA1c >5.7 The CGM data collected will not be used be used in research analysis. Ages of participants range from 12 months to 65 years.

TEDDY Siblings

Islet autoantibody (IA) positive TEDDY siblings whose parents have consented to the parent/sibling DNA protocol with the TEDDY study are not eligible for the DAISY prospective follow-up. However, these TEDDY siblings are offered autoantibody, HbA1c and random glucose testing every 3-6 months to avoid acute onset of diabetes. Families will receive home glucose monitoring education and will be provided with a glucometer to test blood sugars as recommended. The sibling may be offered a continuous glucose monitor (CGM) for close monitoring of blood glucose levels if they have an HbA1c >5.7. The CGM data collected will not be used be used in research analysis. Ages of participants range from 12 months to 65 years.

TEDDY Participants >=15 years of age

Islet autoantibody (IA) positive TEDDY study (COMIRB 04-0361) participants ages 15 and over who have completed the TEDDY protocol will be eligible for the DAISY prospective follow-up. They will be offered autoantibody, HbA1c and random glucose testing every 3-6 months to avoid acute onset of diabetes. Families will receive home glucose monitoring education and will be provided with a glucometer to test blood sugars as recommended. Also a continuous glucose monitor (CGM) or OGTT may be offered for close monitoring of blood glucose levels.

2. Consent Procedures:

Study participants in the sibling/offspring cohort are recruited by telephone/mail or by their physician and asked to come into a study clinic. If the child is younger than 9 months (age of the initial clinic visit) and eligible for telephone interviews, the consent process is completed over the phone and the parents are asked to mail back a copy of the signed consent form prior to the first scheduled telephone interview. These clinics are located on the UC Denver AMC campus at the Barbara Davis Center. Upon arrival, the study participant is given the consent form to read. This is a one-step informed consent to follow-up and genetic testing. The study coordinator or the research nurse goes over the consent form with the study participant sentence by sentence. The research nurse invites questions from the study participant's parent or legal guardian and answers these

questions. The research nurse reads the consent form to the subject's parent or legal guardian when there is a question about the literacy level of the parent or legal guardian. Once the study participant's parent or legal guardian signs the consent form, they are given a copy of the signed consent form, and the data collection begins.
Enrollment of study participants in the general population cohort is in two steps. Consent process for screening is carried out in person by one of the DAISY recruiters at St. Joseph Hospital. The consent to follow-up is identical to that for the sibling/offspring cohort and is completed at the time of the first clinic visit or preceding telephone interview.

Enrolled subjects who develop T1D related autoantibodies or with an HbA1c of >5.7, will be asked to participate in the Continuous Glucose Monitoring (CGM) portion of the DAISY study. The parent or legal guardian will complete a separate consent, entitled "Subject consent for Continuous Glucose Monitoring". We will enroll negative controls into the CGM portion of the study.

DAISY family members and TEDDY siblings who are IA positive and wish to be monitored for progression to T1D will complete a separate consent, DAISY Subject Consent for Family Members, in person with the research nurse, in the same manner as described above, at the next DAISY participant follow up clinic visit.

Training for obtaining written consent: The research clinic staff will undergo training in standard procedures in how to obtain informed consent. Dr. Rewers and the Research Nurse, Michelle Hoffman, have attended the mandatory COMIRB training for investigators. In addition, Dr. Rewers and Ms. Hoffman are responsible for training the Professional Research Assistants in obtaining the consent, and discussing with them the importance of fully informing subjects of the nature of the study and of complying with COMIRB and OPRR regulations. The clinic staff practices reading the consent form verbatim, in full and role-play in answering questions regarding the consent form and the study. A script has been developed for responding to questions. Any questions that the clinical staff cannot answer are referred to the principal investigator (Dr. Rewers). Training is completed in approximately 2 hours. The principal investigator and/or research nurse intermittently monitor the consent process in person to assure that protocols and regulations are adhered to.

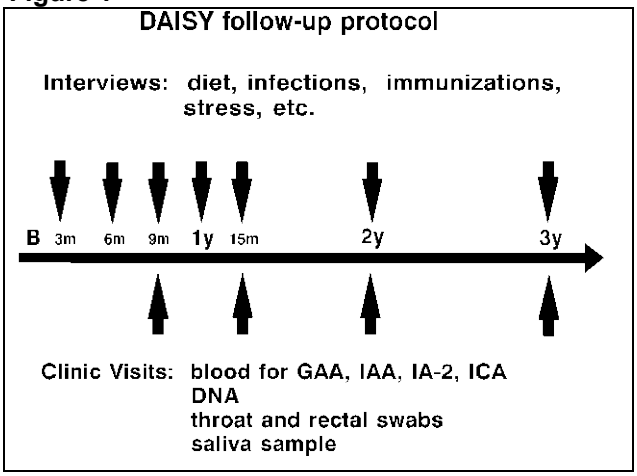
3. Treatment, intervention or observation

This is an observational study. No treatment or intervention will be administered. The prospective follow-up protocol is summarized in the figure and further described below.

Siblings/offspring older than 8 months of age are invited to study clinic for a baseline blood draw and clinic data collection immediately after enrollment. They are seen at the clinic at ages 9, 15, 24 months and annually thereafter. If a child is found to be positive for any of the islet autoantibodies, he/she is placed on an accelerated protocol where the child is seen at the DAISY clinic every 3-6 months to track the natural history of these autoantibodies. Those relatives who meet the entry criteria for the TrialNet Study are followed by TrialNet with intravenous glucose tolerance tests as well as ICA and IAA measurements. As a part of a study ancillary to TrailNet, we continue to follow these children using the regular DAISY protocol.

General population newborns are followed-up only if they meet specific study criteria (see below). The results of newborn HLA screening and an interpretation are mailed to the parents. In addition, parents of moderate and high-risk infants are telephoned to provide genetic risk counseling. To date, all high risk and a sample of moderate and low risk children have been invited to participate in follow-up to determine the age- and genotype-specific incidence of autoimmunity, evaluate its candidate environmental causes as well as promoters of progression from IA to T1D. Eligible children are enrolled after a separate informed consent is signed. When an eligible child's parents decline participation, they are asked to allow telephone contact and medical records review every year, for the purpose of determining the incidence of T1D.

Figure 1



Information concerning dietary, infectious and other exposures is collected prospectively over the telephone when the child is 3, 6, 9, 12 and 15 months old (Fig.1). Blood, and urine, are obtained for measurements of exposure and islet autoantibodies. Weight and height are recorded during clinic visits at the ages of 9, 15, 24 months, and yearly thereafter. Body composition (weight and kilograms of body fat) will be measured at every DAISY visit on subjects coming for study visits at the Barbara Davis center or those who are seen at our off site clinic locations using the TANITA® DC-430U Dual Frequency Total Body Composition Analyzer. Blood amount drawn is 15 cc in children whose weight is <10 kg and never exceeds 3 cc/kg in those over 10 kg. If a child is found to be positive for any of the islet autoantibodies, he/she is placed on an accelerated protocol where the child is seen at the DAISY clinic every 3-6 months to track the natural history of these autoantibodies.

DAISY family members and TEDDY siblings are only monitored if previously IA positive, per enrollment criteria above. Approximately 10 cc of blood from these individuals will be collected at the time of the DAISY or TEDDY participants' follow-up clinic visits and tested for plasma glucose level, hemoglobin A1c and IA. No other exposure data will be collected.

D. Description, Risks and Justification of Procedures and Data Collection Tools:

Measurement of islet autoantibodies: Blood samples obtained by venipuncture or a finger stick will be used. We are currently using radioassays for insulin, GAD_{65}, and ICA512 (IA-2) autoantibodies, IAA, GAA, and ICA512AA respectively. IAA are measured in a standard radioimmunoassay[145] incorporating competition with unlabeled insulin and precipitation with polyethylene glycol and in a new microassay. The former requires 600 μl of serum (150 μl duplicates with and without unlabeled insulin). The inter-assay coefficient of variation is 10.3% at low positive levels[145]. In the most recent proficiency testing this assay gave 100% sensitivity and specificity. The microassay, based on protein A/G precipitation, requires 5 μl of serum and appears to be more specific than the large volume assay. GAA are measured in triplicate using a modification of the radioassay with in vitro transcribed and translated human GAD65 and precipitation with protein A-sepharose[146]. The results are expressed as an index: GAA index = (sample cpm - negative control cpm) / (positive control cpm - negative control cpm). The inter-assay coefficient of variation is 10.7% (n=24). In the most recent Immunology of Diabetes Society Workshop (IDS1995) the assay gave 82% sensitivity and 99% specificity using sera from new onset diabetic patients aged less than 30 years. The ICA512 (IA-2) autoantibodies are measured in a system similar to the GAA assay with the recombinant human ICA512 protein transcribed and translated in vitro with ^{35}S labelling[147]. The most recent modification of the assay utilizes clone consisting of aminoacids 256 to 979 of human

ICA512. The inter-assay coefficient of variation is 9.6% (n=12). In IDS1995 the assay gave 73% sensitivity and 100% specificity. Over 90% newly diagnosed T1D patients and ICA positive relatives followed to diabetes[33-36] are positive for either GAD_{65} or ICA512 using the combined assay. The GAA and ICA512AA are measured in a combined assay using ^{3}H-labelled GAD_{65} and ^{35}S-labelled ICA512 autoantigens and Packard Top beta-counter with protein A sepharose and 96 well filtration.

Quality control All samples with GAA or ICA512AA levels exceeding the 99th percentile of the distributions in 198 unrelated non-diabetic persons ages 0.4 - 67.5 years (i.e., GAA >0.032 and ICA512AA>0.16, respectively) and a random 10% of the remaining samples are blindly retested for quality assurance. The IAA testing were repeated if the levels are above 79 nU/ml, rather than above the 99th percentile of non-diabetic distribution (42 nU/ml) because of the large serum required relative to that obtained in these young children. With development of IAA microassay, all positive samples (index >0.01) are re-tested using blinded duplicates.

Primary endpoint for these analyses will be islet autoimmunity (IA) defined as presence of at least one autoantibody (IAA, GAA, ICA512AA), measured according to the protocols described in Specific Aim 1. All such children identified in the sibling/offspring cohort and in the newborn cohort are eligible. The two cohorts will not be combined for analyses. Persistent autoimmunity is defined as presence of at least one autoantibody on at least two separate blood tests. T1D, defined according to the WHO criteria[148] is a secondary endpoint. Hemoglobin A1c and random blood glucose are measured with each blood sample in children positive for any islet autoantibody and the results are reviewed by pediatric diabetologist. If the value is >6% (nl 4.3-6.4%), the parents are asked to measure several fasting and 1-2 hr post prandial blood sugars at home. If the fasting glycemia is over 120 mg/dl and/or the post-prandial levels are >180 mg/dl on a home glucose monitor, a standard 2 hr oral glucose tolerance test is performed. Families of children persistently positive for more than one autoantibody are provided with keto-diastix strips and taught how to test urine for presence of glucosuria and ketonuria.

Surveillance of the whole newborn cohort screened at St. Joseph Hospital: The whole cohort is followed for development of IDDM. Further surveillance of the whole cohort is necessary to produce unbiased estimates of the risk of IDDM associated with different HLA genotypes, and predictive value of the genetic markers. At enrollment, parents of all participants authorize release of medical information. Since most of the HLA screened children are members of Kaiser Permanente HMO, hospital, clinic, and pharmacy databases of Kaiser Permanente will be linked yearly with DAISY database, using unique Kaiser ID. The initial linkage of the databases identified Kaiser records for 76% (5981/7849) of the HLA screened children (we are attempting to resolve 107 mismatches for children thought to be Kaiser members). Thus, we can efficiently access, through database linkage, clinical records including ICD codes and insulin prescriptions of at least 76% of the cohort. Membership turnover rate is about 5%/yr. Families of the remaining 24% of the children is contacted yearly by mail, to update the demographic, address and diabetes status information (Surveillance Form). Computerized patient database of the Barbara Davis Center, where over 90% of Denver Metro Area newly diagnosed T1D children are seen, serves as the validation data source for diabetes ascertainment in the whole cohort. Currently, 75 cases of diabetes have been ascertained and 1919 have developed persistent IA, among the initial 37,683 children screened.

This unique cohort is already serving as population laboratory for several additional studies. The celiac disease project (RO1 DK-50979, 9/95-3/2015) allowed us to screen for islet autoantibodies a group of children with the "low risk" HLA-DR3/x genotypes and a group of relatives of patients with celiac disease who are also at an increased risk of T1D. The DR3/x genotype groups represent 15% of T1D cases in the general population. Investigation of Vitamins in Youth (IVY, R01 DK 49654, J. Norris, PI) studies a subset of participants for the associations between vitamin supplements and development of IA and T1D. A supplement to the main study: "Immunizations and type 1 diabetes (DK32493 16S1) investigates the potential link between routine childhood vaccination and IA and T1D. Finally, recently funded Autoimmunity Center of Excellence includes a clinical trial component, where 300 of additional SOC participants will be

screened for the HLD-DR3/4, DQB1*0302 genotype and those positive, offered participation in a trial of immunological "vaccination" with small amounts of insulin to prevent development of IA and T1D.

HLA typing: The HLA complex, and in particular the DQ and DR loci account for approximately 50% of the familial aggregation of T1D. DAISY is screening all participants for the major T1D susceptibility (HLA-DRB1*03,*04 and DQB1*0302) and protection (DQB1*0602) alleles. In addition, cases of islet autoimmunity and appropriate controls are fully genotyped for DRB1 and DQB1 subtypes. There are several reasons why we are proposing to continue this subtyping.

First, the PCR-based typing of cord blood is prone to contamination with maternal blood. Eight cases of contamination (0.1%) were inferred from the presence of third allele in the cord blood of the initial 7,537 children screened, however, there may be undetected cases. For instance, if a child is DR1/8 and the mother is DR4/8, and mother's DNA significantly contaminated the cord blood, it would be typed as DR4/x. In the case of homozygosity, e.g., if the child is DR3/3 and the mother is DR3/x their mixed blood would type as DR3/x. Although contamination did not appear to be a significant problem in the screening of cord blood, its exact rate should be determined by retyping the HLA-DR/DQ genotypes on blood samples obtained directly from the child. We propose to accomplish this through subtyping of the DRB1 and DQB1 loci in all children selected for the proposed nested case-control studies using blood sample collected at the 15 months visit. This will also assure the validity of the HLA data used in these studies.

Second, it is unclear whether the associations of different DRB1*04 subtypes with islet autoimmunity are the same as with T1D. We will determine whether the DRB1*0403 allele protects from autoimmunity or rather from progression from autoimmunity to diabetes. We also need to subtype the DQB1 alleles of children identified with the usually protective DR2 haplotype. In the Denver population, 34% of children with the DR2/4 DQB1*0302 genotypes do not have the protective DQB1*0602 allele on their DR2 haplotypes. Finally, subtyping of the DRB1 and DQB1 alleles may help to understand why more Hispanics (24%) than non-Hispanic whites (17%) have moderate risk HLA genotypes, while the incidence of T1D in Hispanics is only 60% of that in non-Hispanics[83]. On the other hand, among persons with DR4/DR2, more Hispanics than non-Hispanic whites carry a neutral DR2 haplotype, rather than the protective DR2(DRB1*1501),DQB1*0602 haplotype and are, in fact, at an increased risk of T1D.

The goal is to collect at least 100 complete families with an IA index child and both parents available (diabetic and non-diabetic siblings will also be asked to participate). DNA will be extracted and aliquotted for these and future studies. We do not request funding for setting up cell lines from these samples given the expense and the amounts of DNA realistically required for the proposed studies. Approximately 20 μg of genomic DNA is obtained from 1 ml of whole blood. With PCR based typing, and at least 10 ml of whole blood per individual, we expect to have sufficient DNA for more than 4,000 PCR based tests per person. If family members are unable to come into the clinic or unwilling to have their blood drawn we will offer to have them collect their DNA using an OROGene DNA self-collection kit.

Quality control The procedures included blinded duplicate HLA typing of the same blood sample for all high and moderate risk genotypes and a random 10% of the low risk genotypes (reduced later to 6%). We found only six discordant cases and four occurred in the first ten weeks of the study. None of the 104 samples typed as high risk genotype gave a discordant result on retyping. Repeat typing gave discordant results for 4 out of 519 moderate and 2 of the 354 low risk genotypes. Thus, the reproducibility of the screening was over 99%.Viral studies We will continue to use the method of Pallansch et al.[149] to determine the levels of IgM and IgA antibodies to 17 different enteroviruses (Coxsackie B1-6 and A9, ECHO 4, 6, 9, 11, 30 and 34, enterovirus 71 and poliovirus 1, 2 and 3). The testing will continue to be carried out in the Respiratory and Enteroviral Laboratory at the Centers for Disease Control, Atlanta, GA, under the supervision of Dr. Mark Pallansch. In addition, samples of serum, saliva and rectal swab viral medium were tested for the presence of enteroviral RNA using a well validated assay[150] in the laboratory of Dr. Harley Rotbart in the Department of Pediatrics, University of Colorado, Denver, CO. In addition, appearance of infectious diseases in the previous three months are asked during telephone interviews at ages 3, 6, 9, 12, 15, and 24 months and yearly thereafter.

RNA microarray studies of gene expression*: In collaboration with Dr. Jin-Xiong She, Director of Diabetes Research and NIDDK Biotechnology Center at the University of Florida, we will initiate RNA microarray studies of gene expression in non*-diabetic children with BCA compared to their parents and age, HLA-DR3/4 matched children.

Identification of other biomarkers: A large nested case-control proteomics study based on the DAISY population has identified seven potential protein/peptide biomarkers of progression to T1D. The biomarkers were cross validated using two approaches: 1) the liquid chromatography-mass spectrometry (LC-MS)-based proteomics analyses approach in collaboration with colleagues from the Pacific Northwestern National Laboratory (PNNL); and 2) the aptamer approach (Somalogic, Boulder, CO). Additional protein biomarkers have been reported from the DIPP, nPOD and other studies. We are proposing to apply the newest version of the SOMAscan™ assay that can measure up to 3,000 proteins simultaneously across a wide range of concentrations. This SOMA scan assay is available through the Genomics and Microarray Core Laboratory at the University of Colorado and can be customized to capture all targets pertinent to this proposal. Additionally, we have utilized this same case-control cohort to identify metabolomics signatures characteristic of islet autoimmunity and progression to T1D. In collaboration with Metabolon, the analysis of serial plasma samples may identify up to 2,200 named metabolites. The study is powered to confirm previously identified associations as well as identify potential new metabolic pathways involved in the development of IA and T1D. Finally, longitudinal changes in the metabolic profiles identified in the study participants will be related to detailed dietary records, infections, and other environmental stressors, e.g., antibiotic use, during the follow-up period leading to diabetes.

T-cell study: T-cell reactivity to autoantigens will be studied in autoantibody positive siblings and negative control children with HLA DR3/4 DQB1*0302 as well as non-diabetic parents of these children. This work will utilize a standard proliferation assay system using unfractionated peripheral blood mononuclear cells.

Infant diet Under the current DAISY protocol, infant diet data is collected prospectively via periodic telephone interviews conducted by a professional interviewer at 3, 6, 9, 12 and 15 months. Minor modifications to this protocol are proposed for the additional general population newborns to be enrolled. The questionnaire (UPDATE3_15) is included. The infant diet portion of this questionnaire is a modification of a questionnaire that was found to have relatively good agreement with pediatric records[54]. During each interview, parents will be asked to detail their infant's diet over the past 3 months (i.e. during a 6-month interview, the mother would be asked to recall the infant diet in the 3rd, 4th and 5th months). The interviewer records the date a milk/food on the list was introduced into the diet. For infant formulas, the type, including brand name is recorded. The types of cereal (e.g. rice, oatmeal) and fresh milks (e.g. cow, goat) are also recorded separately. Then the frequency of consumption of that food is averaged over a month and is recorded as number of servings per day. For breast-milk, this would be number of feedings per day; for formulas and milks, it would be recorded as the number of baby bottles per day and for solid food, it would be number of discrete servings per day, without distinguishing serving size. If the milk/food was introduced mid-month, the frequency of consumption is recorded as the average from the start-date to the end of the month. In addition, frequencies of consumption are recorded separately for juice, fruit, vegetables, meat, breads, dairy products, eggs, sweets and snacks. Vitamin supplementation during infancy, including brand name and dose, will also be collected.

Early childhood diet will be measured annually using the Willett FFQ that has been altered for use in pre-school children[151]. The Willett FFQ has been found to be reliable in children while using parental recall[152]. All mothers of children will be given a food frequency questionnaire asking them to record their child's diet in the previous year. As part of the FFQ, extensive information is collected on vitamin supplementation. This food frequency questionnaire will be first administered when the participating child is 2 years old and repeated annually. In addition, two 24-hr recalls per

year will be administered to the participants. The diet recall forms are included. The diet recall will be used to calibrate the FFQ for each participant[153]. From the age of 10 years on, we ask the children to complete the Youth/Adolescent Questionnaire (YAQ), which is a modification of the Willett FFQ that is meant to be filled out by the child alone.

Dietary antigen exposure: All FFQ forms will be sent to the Channing Laboratory for data entry and analysis. All training, data entry and analyses for the 24 hr recall will be performed by the Energy Balance Laboratory of the Center for Human Nutrition (James Hill, PhD, Director) at the University of Colorado. Nutrient profiles for both methods will be produced as a result of this analysis. Some of the variables of interest for the proposed study include total protein intake (i.e. grams per day), dairy protein intake, vegetable protein intake and nitrate/nitrite intake. Vitamin intake (i.e. antioxidants) will also be of interest as a potential effect modifier or protective factor. Total vitamin intake will be calculated from the vitamin supplementation data and vitamin content of consumed foods. The infant diet exposures will be calculated based on the infant diet food frequency (contained in UPDATE3_15). In order to calculate exposure quantities, an average serving size across subjects is assumed. Based on the frequency of servings, exposures to specific dietary proteins, i.e. dairy, animal, vegetable, wheat, etc., will be calculated on a monthly basis. In order to do this, we will use the Minnesota Nutrition Data System (NDS) to assign average nutrient quantities (i.e. grams of protein per day, etc) to the milks and foods in each category listed in the infant food frequency questionnaire. Additional measurements of dietary and viral exposures are described under Specific Aim 2.

Immunizations: Immunization records will be obtained, participants will be asked to sign a separate authorization form which will allow us to gather this information.

Self Tanner-Stage Questionnaire: Administration of a questionnaire designed to capture subject's Tanner stage of puberty. Age at Menarche will also be collected for the girls. DAISY subjects >= 8 years of age will be asked to complete this questionnaire. Girls will be asked to complete the female Tanner staging questionnaire, and boys will be given the male Tanner staging questionnaire to complete. These questionnaires will be given and collected at clinic visits, and will not be mailed to subjects.

Prostaglandins: Prostaglandins are lipid molecules derived from arachidonic acid via the enzyme cyclooxygenase PG synthase (PGS). The pro-inflammatory prostaglandins (PGE_2) produced by macrophages and monocytes expressing PGS_2 are potent modulators of the immune response and tolerance mechanisms. Ingestion of fish oils that contain n-3 PUFAs will result in a decrease in membrane arachidonic acid levels and a concomitant decrease in the capacity to synthesize prostaglandins from arachidonic acid (Calder, 1998). Prostanglandins (PGE2) will be measured in serum from subjects using ELISAs specific for each eicosanoid. Using these data, we will be able to see whether prostaglandin levels are associated with risk of BCA.

Physical activity: As DAISY subjects approach puberty we are interested in examining factors related to insulin resistance, such as weight and physical inactivity. To measure these factors we will collect a waist circumference on each participant and a short questionnaire on physical inactivity.

Puberty Hormones: Inhibin B, Testosterone, and Estradiol will be measured in blood samples from subjects in order to determine the onset and progression of puberty.

Cytokines: Adiponectin, Leptin, Interleukin-6, Interleukin-8, Monocyte Chemotactic Protein 1, Tumor Necrosis Factor Alpha, and Interferon gamma will be measured in stored plasma samples. Using this data, we will be able to examine whether obesity-related inflammation, insulin-resistance and beta-cell toxicity are related to IA and T1D.

Nutrient Biomarkers Vitamins, carotenoids, and red blood cell membrane fatty acid content will be measured on stored blood samples, in order to obtain a biological measure of nutrient exposure.

Oxidative Stress F_2 isoprostanes will be measured on stored urine.

Development of automated reliable high-throughput methods, using dried blood spots, to screen for the autoantibodies against GAD_{65} and ICA512

Venipuncture, separation of serum or plasma, and frozen storage of the material are important limitations for large-scale population screening for autoantibodies. The filter paper technology using dried blood has been employed in automated screening for antibodies to HIV-1[154]. Islet autoantibodies are remarkably stable in dried blood spots and the volume required for the recombinant GAA/ICA512AA radioassays is well within the range of 50-100 μl of whole blood contained in a standard "Guthrie spot", collected in routine genetic newborn screening programs nationwide. To pilot the feasibility and accuracy of GAA and ICA512AA testing using dried blood spots, we collected blood samples from 46 T1D patients with a wide range of positive GAA and ICA512AA levels and from eight healthy controls[130]. Coded sera obtained from these 54 persons were tested for GAA and ICA512AA using our routine combined assay. In addition, a few drops of capillary blood were collected from each of these 54 person on standard filter paper cards used by Colorado Newborn Screening Program (coded differently than serum), dried, stored at room temperature for a few days and then in 4° C for over a month. A standard piece of the whole blood blot, representing approximately 10 μl of serum, was eluted using the standard GAA/ICA512AA assay washing buffer, and the autoantibodies were measured in the eluent using the routine method. Specificity of the dried whole blood method was 100% for both GAA and ICA512AA, compared to the routine method using serum. Sensitivity for GAA was lower only for those samples that were weakly positive in our routine assay (below 5 SD score of mean of 205 controls). Sensitivity for ICA512 antibodies was 94% (1/23 of patients false negative by filter paper assay). The correlation (Spearman) with the routine method was r=0.89 for GAA and r =0.94 for ICA512 ($p<10^{-6}$ for both). Based on this pilot study, we are confident that we can accurately measure GAD_{65} and ICA512 autoantibodies using dried blood filter paper spots.

The filter paper method will be used in all cases where a sufficient amount of venous blood cannot be obtained, where there is a strong parental or child's preference for capillary vs. venous blood collection, or where blood is collected in a remote location, rather than in the DAISY or Kaiser Permanente clinic, and there is a strong preference on the part of the family or health provider collecting/shipping the blood sample, for filter paper vs. venous blood sample. Using estimated loss to follow-up we expect that 75% of the siblings and offspring otherwise lost (n=94) and 50% of the lost general population participants in the existing (n=117) and new (n=105) cohorts will accept this alternative. This will significantly reduce loss of the follow-up outcome information. In addition, if the filter paper method is highly valid and reliable (as expected) and as predictive of strong islet autoimmunity progressing to diabetes as the currently used combination of GAA, ICA512AA and IAA, this new method would provide an attractive inexpensive alternative for large general population screening programs.

Continuous Glucose Monitoring (CGM)

IA subjects or those that have had an HbA1c test >5.7 will be asked to complete a 7-10 day period of CGM up to 4 times per year. The study will also recruit participants negative for diabetes autoantibodies matched by age and gender to the cases to compare CGM results to subjects with positive autoantibodies

A CGM transmitter and receiver, sensors and in-person training by the study research nurse will be provided to the study participants at the Barbara Davis Center pediatric clinical research center on the day of the visit. In addition, the research nurse will be available by pager to answer questions during the 7-10 days of CGM. Certain medications are not allowed for one day prior to and during the time of CGM measurement. Acetaminophen (paracetamol, Tylenol) is prohibited for one day prior to and during the time of CGM measurement. glucose levels by measuring and transmitting glucose information from interstitial fluid every 5 minutes to the pager-sized receiver, which is kept in a pocket or clipped onto a belt of the subject. The CGM monitoring will be blinded to the subject

during the–7-10 day period. The results will be reviewed by a study physician at the end of monitoring for patient safety.
Measures of glucose control will include HbA1c, the overall mean of glucose values, % of values within target range (70-180 mg/dl), % of values <70 mg/dl (hypoglycemic range), and % of values >180 mg/dl (hyperglycemic). Primary variables to characterize glycemic variability will include the overall standard deviation (SD), the mean of daily differences (MODD) and the mean amplitude of glycemic excursions (MAGE). While multiple additional computed measures have been proposed in this dynamically evolving analytical field, they do not appear to offer a particular advantage and we will limit the number of comparisons to minimize the chance of type I error.

Blood Glucose meter
IA subjects or those that have had an HbA1c test >5.7 will be given a glucose meter to monitor blood glucose levels. The study doctor or study nurse will review the results from the meter and subjects will be notified if significant hypoglycemia or hyperglycemia is observed.

Risks

Genetically susceptible children are being currently identified by cord blood screening. Screening for genetic markers associated with (but not diagnostic for) a severe and currently incurable disease, such as T1D, raises important ethical issues. Our study placed special emphasis on: 1) voluntary participation ensured by informed consent process; 2) full disclosure of the results of the screening to parents, combined with education about T1D and genetic risk counseling; and, 3) confidentiality of genetic information that cannot be disclosed to health providers or other parties without parental consent. The informed consent process for the follow-up phase of the study has been and will continue to be separated from the consent to screen and involved additional genetic counseling. Since the prognostic significance of these markers is currently uncertain, psychologic support of the families will be made available as needed, so that undue anxiety about developing T1D is not invoked. Psychologic studies have found minor transient increases in anxiety with rapid appropriate coping for subject's positive for genetic of immunological risk markers and their relatives[158]. Parents of only 6% of the children screened were concerned that a positive result of the screening could adversely affect the insurability of the child. Colorado law specifically prohibits insurance institutions from seeking and using results based on DNA testing for determination of eligibility of conditions of coverage.

Subjects: This is not a therapeutic/intervention study, and so we do not anticipate serious adverse events, however, for those children enrolled in the study, potential risks are from blood draws and psychological stress. The blood draw volume has been kept low, so that other than possible hematoma formation, minimal risks are foreseen. Participants may feel brief pain at the time of the needle stick for the blood draw. In about 10% of cases, a small amount of bleeding under the skin will produce a bruise (hematoma formation). The risk of temporary clotting of the vein is about 1% and the risk of infection of the bruise or significant external blood loss is less than 1 in 1,000. Occasionally, individuals may experience light-headedness or fainting during the blood draw process. This occurs in less than 1% of individuals undergoing a blood draw or blood donation. DAISY participants undergo blood draws in a phlebotomy chair with arm and back rests to minimize the risk of fall injury. If participants experience light-headedness or fainting during a blood draw, the reaction is documented and subjects are asked to lie down during subsequent blood draws.

There is the possibility of increased anxiety that may occur with the study of factors that could trigger T1D. Parents may become concerned that any illness or change in diet will be crucial. Extensive training of study staff in methods to counsel families has been instituted and ongoing since the beginning of the study. In addition, provision is made for access to counseling at the Barbara Davis Center, which is very experienced in dealing with these fears. These fears exist without this study. The participation in this protocol and the DAISY protocol may lessen such concerns, since children who test negative for autoantibodies are at lower risk of T1D, and since

children who become positive will be followed closely and counseled concerning the best treatment. Thus, the likelihood that diabetes will develop abruptly, leading to significant morbidity or mortality, will be lessened.

Continuous glucose monitoring (CGM) requires the insertion of a small plastic tube with a needle that is then removed. The small plastic tube stays in place for 7-10 days. There is a low risk of developing a local skin infection at the site of the sensor needle replacement. Itchiness, redness, bleeding and bruising at the insertion site may occur as well as local tape allergies.

The risk to loss of confidentiality is extremely low (none in 7 years with nearly 25,000 participants), computer files are accessible only to study staff with appropriate passwords, individual identifiers are kept in a separate file, all subject records are kept in ID access areas which allows entry only for staff with specific permission and all analyses are done anonymously.

Investigators
There are no known risks to investigators.

Benefits

There are no benefits to the subjects from participating in this research study. DAISY participants are compensated $20 for each clinic visit.~~,~~ Participants who complete the CGM procedure are compensated $60 for each one-week wear period. Family members are compensated $20 for a one time blood draw, to complete genetic HLA typing and T1D related autoantibody testing.

Based upon studies in identical twins, the genetic risk of developing T1D to an identical co-twin is approximately 20-30%. Thus, a large component of risk must be environmentally mediated. Since little can be done at present to modify genetic risk, further identification of environmental factors that could be modified through dietary advice or immunization (primary prevention) is imperative, even though it must include very young children, since this is the critical time period where the disease develops. Participation in this protocol allows for the earliest possible detection of pre-clinical diabetes amenable to intervention (secondary prevention). Risks and benefits associated with specific types of secondary prevention will be presented to interested parents. It will only be parents who will ultimately determine if these long-term benefits outweigh the short-term risks, through their participation.
These studies are critical for development of cost-effective population screening for T1D. The disease is a candidate for routine screening, because: 1) it is quite prevalent, affecting 0.5-1% in the U.S. population, but the clinical diagnosis is follows several years of asymptomatic autoimmunity with destruction of insulin producing cells; 2) it usually causes severe morbidity and premature death; 3) early detection is feasible using a sensitive and specific autoantibody tests; and 4) a number of early treatment protocols are being evaluated in large clinical trials involving relatives. The long-term health outcomes and cost-benefit ratio for early diagnosis have not been studied, because of the paucity of population studies, especially in the U.S.

Changes from Usual Treatment
All laboratory tests and procedures undertaken in this study are for research purposes only, and would not be part of the usual clinical care of the subject. However, values of HbA1c and blood glucose exceeding normal levels will be reported to the participants and/or their primary care providers.

E. Potential Scientific Problems:

Cohort Retention: We have been able to retain study participants over an extended time period; 85% of the cases and 62% of the original cohort have been retained. Frequent telephone contact with well-trained and flexible staff has encouraged continued participation. Further loss to follow-up will be minimal because the majority of the population has already completed multiple visits and

drop-out is limited to moving out of the study area. We offer long-distance participation protocol which has significantly reduced drop-out rates.
Limitations: As the study participants get older, some potentially exciting studies become not feasible. For instance, participating families told us that collection of stool samples, of interest in microbiome research, would decrease their further participation. This area is better left to TEDDY that has collected monthly stool samples in young participants. We have also refrained from collecting T-cells, on a large scale, because of cost, poor performance of assays, and large blood volume needed relative to child's age. However, DAISY has participated in a number of smaller T-cell studies, using fresh blood (in collaboration with Drs. B. Kotzin, P. Gottlieb, H. Reijonen and G. Nepom) and we will strongly consider adding T-cell studies to the follow-up if our current collaborative experiments with Drs. J. Kappler and B. Roep or others show promising results.

F. Data Analysis Plan:

Plan for monitoring adverse events: This is not a therapeutic/intervention study, and so we do not anticipate any adverse events. There will be no independent data safety monitoring board, since this is not a therapeutic or intervention study. The study coordinator and the Principal Investigator will conduct ongoing monitoring of the study.

Specific Aim 1. Determine the natural history of islet autoimmunity and to explore the heterogeneity of diabetes and other autoimmune phenotypes in young adults through follow-up of the existing cohort for 5 more years.
We will follow already established cohort of youth at high risk of T1D and other autoimmune diseases (current n=1149, median age 17.2 y, IQR 13.5-20.2 y). We will estimate overall burden of pre-clinical and clinical T1D, celiac, thyroid, adrenal, rheumatic and parietal autoimmune disease in Colorado by age 25. This will inform future screening and prevention programs. We will further explore the apparent heterogeneity of IA and its implications for adult-onset diabetes. Selected hypotheses include:

- Hypothesis 1.1: Persistent IA triggered after the age of 6 y results in T1D diagnosis in early adulthood
- Hypothesis 1.2: Transient childhood IA does not increase the risk of diabetes in early adulthood
- Hypothesis 1.3: Sedentary life-style and insulin resistance predict faster progression from persistent IA to T1D

Statistical Analyses and Power Cox PH analysis will be used to test Hypothesis 1.1 and 1.2, controlling for HLA-DR,DQ, family history of T1D, sex and ethnicity. Similarly, we will estimate adjusted hazard ratios (HR) for sedentary lifestyle and insulin resistance, adjusted for Tanner Stage and diet (Hypothesis 1.3). Assuming an additional 5% of children will seroconvert, we will have 80% power at a α=0.05 to detect HR>1.8 or < 0.56 /SD.
Joint latent class models. Using fixed and time-varying covariates from the 1-year period following conversion to IA, we will develop a joint latent class model that captures the dependencies between our fixed and longitudinal covariates (age, HLA risk group, family history, BMI, glucose, physical activity, insulin resistance, development of additional IAb) and time from IA conversion to T1D. We assume that heterogeneity across our cohort can be distilled into homogeneous latent subgroups with common covariate trajectories or patterns and common risk of T1D. We will implement the method using the *lcmm* R package, which will also allow us to summarize dynamic predictive accuracy of the joint model over time[80;81].
Analytic methods to examine age-related heterogeneity will include testing age*exposure interaction terms with age as a continuous variable in a Cox PH model. If these are significant, suggesting violation of the assumption of proportionality, we will explore age cut-points using B-spline functions[82], and adjusting for variables found to be associated with IA in the previous investigation [6]. We will then incorporate these age cut-points into the interactions terms in the proportional hazards model to test the age-related heterogeneity.

Power analysis. A two-sided log rank test with an overall sample size of 2,547 subjects (Hypotheses 1.1 and 1.2) with an overall IA rate of 9% in 15 years achieves 80% power at a 0.05 significance level to detect a HR=0.66 assuming that the proportion of children lost to follow-up is 0.20. These results are based on the assumption that the hazard rates are proportional over time. Adjusting for covariates in a Cox model and assuming that the regression of cohort membership on covariates has an R^2 of 0.5, the minimally detectable hazard ratio is 0.56. For the Hypothesis 1.3, among IA positive children (n=98), we will have 80% power to detect a hazard ratio of 1.87 for each SD increase in time spent sedentary, waist circumference, and BMI, assuming 20% of children with IA will develop T1D. Expected SDs: time spent sedentary=90 minutes/day; waist circumference=15cm; BMI=4.3 kg/m^2.

Specific Aim 2. Validate candidate proteomic biomarkers of IA and T1D, in a nested-case study of 213 youth with persistent IA and 213 controls.

Statistical analyses will be carried out by Dr. Bobbie-Jo Web-Robertson (PNNL) – see letter of collaboration. The relative quantification data will be normalized to the reference sample, log transforming (base 2), and then median centered. First mixed effects linear models, with a normal conditional distribution and identity link function, will be fit separately to cases and controls using R and the lme4 package[100;101]. Age and sex will be used as the fixed effects in these models including linear and quadratic models to examine trends in abundance over time and random effects for each subject will be included to account for the subject variability and non-independent nature of the data over time. Likelihood ratio tests will be conducted with (α = 0.05). A difference in trends between cases and controls will be quantified by fitting the mixed effects linear/quadratic models and including a fixed effect to account for different mean intercepts, slopes and quadratic terms, where applicable, for each group. Additionally, a likelihood ratio test between a model including and not including case-control status as an effect can be used to determine differences in intercept, slope, and/or quadratic mean trends. The highest degree model that is statistically significant for at least one group will be fit with fixed effects for age and group and an interaction term between age and group (allowing for separate slopes for each group) and a random effect for subject.

Detecting a significant difference in trends cases and controls is by testing if the group regression parameter is equal to 0, which is rejected when the true population parameter is different from 0. In the case of a linear trend, this involves the following t-test: $t_{\alpha,n-4} = b/s_b$ where α is the type 1 error rate, *b* is the estimated regression parameter, s_b is the standard error of the regression parameter estimate. The standard error of the regression parameter estimate depends on the overall variability of the data, the number of subjects (n) and the number of time points per subject (T). A power analysis was conducted based on a simulation study. From the previous DAISY1 study, the 75th percentile absolute value estimated slope due to group membership was *b* = 0.005 and the overall error variances were calculated for each peptide. **Table 4** gives the power of the above test for select combinations of number of subjects and time points per subject for the 90th percentile of error variances. There will be a sufficient power to test the hypotheses with 4 or more samples per subject even in a subgroup analysis with 25 subjects per subgroup, For instance, a sample size of 20 subjects per group and 4 time points per subject will provide at least 82% power at the 90th percentile of the observed error variances.

Table 4: Power calculation for Specific Aim 2

Number of subjects/group	Number of time points/subject					
	3	**4**	**5**	**6**	**7**	**8**
30	.89	.99	1	1	1	1
25	.77	.90	.96	.99	1	1
20	.60	.71	.82	.89	.95	.98
15	.34	.43	.57	.52	.70	.84

The mixed model approach is typically preferred to ANCOVA with repeated measures for this type of data. ANCOVA's limitations include sensitivity to missing observations and restrictive variance-covariance structures. Additionally, mixed models naturally handle uneven spacing of

repeated measurements, tend to be more interpretable and will lead to natural tests for comparing groups.

Quality Control and Protocol Adherence for all specific aims will remain largely the same and include: standard interviewer training and retraining, inter-interviewer reliability testing; precisely written protocols, random review of the study forms by the investigators, computing error rates from doubly entered screens, blind duplicates for laboratory measures at a sampling rate of 5%, monthly QC logs of clinical and laboratory data, equipment calibration data, re-certification data, listing of incomplete exams, and adverse events. The distribution of data for key variables will be reviewed. Interviewers, examiners, and laboratories will be queried about unusual, illogical or out-of range values and missing values. The study investigators meet monthly.

Data management and confidentiality

The questionnaire data, physical exam and laboratory results will be recorded in an Access database maintained by the study data manager. Standard security procedures will be employed to assure patient information confidentiality. Subjects will be identified only by a study ID code in the database, and all identifying information will be kept in a separate, password-protected file. All paper records will be kept in ID access areas which allow entry only for staff with specific permission.

The study design requires that data be collected in a variety of locations: St. Joseph Hospital, DAISY Clinic, the Barbara Davis Center and 6 laboratories (3 local, and 3 out of state). In order to facilitate data entry and on-site quality control measures, a distributed data management system in MS Access was developed and implemented. Clinical and interview data are entered by study personnel in the UC Denver AMC School of Medicine-BDC, using custom designed data entry screens. The relational data base structure allows easy retrieval, extensive report generation and security. Data from remote laboratories, such as Roche Molecular Systems, the CDC, or Barbara Davis Center, are downloaded electronically as coded files via e-mail, bi-weekly. Decoding and filtering procedures have been implemented to assure data security and integrity. Inconsistencies in data are reported back to the data source. Once cleaned, data is released to investigators for analyses in the format of ASCII or SAS files. Specific data components which are reported to the participants are exported to automatically generate results letters.

The electronic data generated from the CGM will be recorded in a SQL Server database maintained by the study data manager. Standard security procedures will be employed to assure patient information confidentiality including login user IDs and passwords and restricted access to the data based on user roles. All paper records will be kept in ID access areas which allow entry only for staff with specific permission.

Genomic Data Sharing Plan

All the genetic testing as well as any relevant associated data (phenotype or exposure data) will be deposited in a controlled-access Database of Genotypes and Phenotypes (dbGaP), in compliance with the NIH genomic data sharing policy. This means that only researchers who are given permission to use the information for a specific research project will be able to access the information. All deposited data will be de-identified. Researchers approved to access information in the database have agreed not to attempt to identify subjects. The DAISY study has applied to the National Institutes of Health for a Certificate of Confidentiality to further protect the privacy of our subjects by preventing compelled disclosure of any personally identifiable information.

The Diabetes Autoimmunity Study in the Young (DAISY) and the Barbara Davis Center for Diabetes is committed to the open and timely dissemination of research outcomes. The Principal Investigator (PI), Dr. Marian Rewers, will facilitate data and specimen sharing and distributing knowledge amongst experts of various disciplines for the discovery of mechanisms and development of future interventions for prevention of type 1 diabetes. The ease in which data can be shared is critical to the success and mission of DAISY. The reuse of data will also encourage

further areas of study and support general knowledge sharing throughout the community. This data sharing plan is designed to enable dissemination of data, while also protecting the privacy of the participants and the utility of the data, by de-identification and masking of potentially sensitive data elements. All requests for access of these data sets must be submitted through written request to the PI. Access for the data will be determined by the PI. All investigators who receive DAISY resources must agree to acknowledge the DAISY Study and all co-investigators. This approach is fully compliant with the NIH public data sharing policy (http://grants.nih.gov/grants/policy/data_sharing). That policy states that the NIH expects and supports the timely release and sharing of final research data from NIH-sponsored studies for use by other investigators and that the definition of "the timely release and sharing" to be no later than the acceptance for publication of the main findings from the final data set.
This policy is construed to include tissues, genetic material, and biological specimens as well as de-identified data. It is noted that biological specimens are stored on-site at the Barbara Davis Center Freezer Facility, key-card access controlled with specific request access. All locally stored data is held on SQL Server at the University of Colorado Denver and protected by a three tier security model. First layer is a secure firewall that allows only a specific range of IP addresses to access the ODBC. Second layer is user authorization and authentication based on the Windows active directory. Third layer is security at the table level (micro) that is managed by a database administrator assigning security roles to each specific user.

It is the DAISY policy that the PI shall retain control and primary rights to the data and biologic specimens developed and obtained under NIH funding, subject to Government rights of access consistent with current HHS, PHS, and NIH policies. This policy further establishes that this right shall cover all data and biological specimens stored locally or retained at any individual laboratory that received them under the auspices of DAISY that contribute to achieving study outcomes. Such primary right shall exist until the conditions in paragraph 1, above, are fulfilled, namely no later than the acceptance for publication of the main findings from the final data set.

The DAISY study is committed to sharing data with the greater scientific community while the study is ongoing and to engage researchers in data analysis. A policy has been established for the request of DAISY data by potential collaborators:

1. The request for DAISY data must be based on a specific intended purpose and the requestor must agree that the use of the data is limited to solely that purpose.
2. Publication or presentation of the results of any analysis of the DAISY data must meet the PI's requirements with respect to scientific merit, authorship, and funding attribution.
3. The requestor explicitly agrees that no attempt will be made to link the DAISY data to any other data source that could potentially violate the confidentiality of any DAISY subject or in any way violate the explicit and implied consent und which the DAISY data has been collected.
4. The requestor must agree to the execution of a Material Transfer Agreement (MTA) between the University of Colorado Denver and the requestor's institute.

G. Summarize Knowledge to be Gained:

The main objectives of DAISY are to fully characterize the natural history of pre-T1D, to find the environmental causes of T1D and how they interact with the host's genome. The overarching goal is primary prevention of T1D and universal screening for several autoimmune diseases. In addition, DAISY will continue to serve as the core population laboratory for multiple NIH-funded studies and as one of the key vanguard projects for the TEDDY consortium.

H. References:

(1) Steck AK, Barriga KJ, Emery LM, Fiallo-Scharer RV, Gottlieb PA, Rewers MJ. Secondary attack rate of type 1 diabetes in Colorado families. *Diabetes Care* 2005;28:296-300.

(2) Insel RA, Dunne JL, Atkinson MA et al. Staging Presymptomatic Type 1 Diabetes: A Scientific Statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care* 2015;38:1964-1974.

(3) Ziegler AG, Rewers M, Simell O et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* 2013;309:2473-2479.

(4) Graves PM, Barriga KJ, Norris JM et al. Lack of association between early childhood immunizations and beta-cell autoimmunity. *Diabetes Care* 1999;22:1694-1697.

(5) Norris JM, Eisenbarth GS, Beaty B et al. Dietary antibodies, infant diet and beta-cell autoimmunity: Diabetes Autoimmunity Study in the Young (DAISY) [abstract]Norris JM, Eisenbarth GS, Beaty B et al. *Autoimmunity* 1995;21:22-A082

(6) Norris JM, Barriga K, Klingensmith G et al. Timing of initial cereal exposure in infancy and risk of islet autoimmunity. *JAMA* 2003;290:1713-1720.

(7) Lamb MM, Frederiksen B, Seifert JA, Kroehl M, Rewers M, Norris JM. Sugar intake is associated with progression from islet autoimmunity to type 1 diabetes: the Diabetes Autoimmunity Study in the Young. *Diabetologia* 2015;58:2027-2034.

(8) Stene LC, Oikarinen S, Hyoty H et al. Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). *Diabetes* 2010;59:3174-3180.

(9) Frederiksen BN, Steck AK, Kroehl M et al. Evidence of stage- and age-related heterogeneity of non-HLA SNPs and risk of islet autoimmunity and type 1 diabetes: the diabetes autoimmunity study in the young. *Clin Dev Immunol* 2013;2013:417657.

(10) Steck AK, Johnson K, Barriga KJ et al. Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: diabetes autoimmunity study in the young. *Diabetes Care* 2011;34:1397-1399.

(11) Steck AK, Wong R, Wagner B et al. Effects of non-HLA gene polymorphisms on development of islet autoimmunity and type 1 diabetes in a population with high-risk HLA-DR,DQ genotypes. *Diabetes* 2012;61:753-758.

(12) Frederiksen B, Kroehl M, Lamb MM et al. Infant exposures and development of type 1 diabetes mellitus: The Diabetes Autoimmunity Study in the Young (DAISY). *JAMA Pediatr* 2013;167:808-815.

(13) Norris JM, Yin X, Lamb MM et al. Omega-3 polyunsaturated fatty acid intake and islet autoimmunity in children at increased risk for type 1 diabetes. *JAMA* 2007;298:1420-1428.

(14) Norris JM, Kroehl M, Fingerlin TE et al. Erythrocyte membrane docosapentaenoic acid levels are associated with islet autoimmunity: the Diabetes Autoimmunity Study in the Young. *Diabetologia* 2014;57:295-304.

(15) Hoffenberg EJ, MacKenzie T, Barriga KJ et al. A prospective study of the incidence of childhood celiac disease. *J Pediatr* 2003;143:308-314.

(16) Liu E, Li M, Bao F et al. Need for quantitative assessment of transglutaminase autoantibodies for celiac disease in screening-identified children. *J Pediatr* 2005;146:494-499.

(17) Emery LM, Babu S, Bugawan TL et al. Newborn HLA-DR,DQ genotype screening: age- and ethnicity-specific type 1 diabetes risk estimates. *Pediatr Diabetes* 2005;6:136-144.

(18) Dabelea D, Mayer-Davis EJ, Saydah S et al. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *JAMA* 2014;311:1778-1786.

(19) WHO. European Observatory on Health Systems and Policies. Policy brief: screening in Europe. *www euro who int/Document/E88698 pdf* [serial online] 2006.

(20) Gesualdo PD, Bautista KA, Waugh KC et al. Feasibility of screening for T1D and celiac disease in a pediatric clinic setting. *Pediatr Diabetes* 2015.

(21) Barker JM, Goehrig SH, Barriga K et al. Clinical characteristics of children diagnosed with type 1 diabetes through intensive screening and follow-up. *Diab care* 2004;27:1399-1404.

(22) Rewers A, Dong F, Slover RH, Klingensmith GJ, Rewers M. Incidence of diabetic ketoacidosis at diagnosis of type 1 diabetes in Colorado youth, 1998-2012. *JAMA* 2015;313:1570-1572.

(23) Frederiksen BN, Kroehl M, Baron A et al. Assessing age-related etiologic heterogeneity in the onset of islet autoimmunity. *Biomed Res Int* 2015;2015:708289.

(24) Hess KR. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Stat Med* 1994;13:1045-1062.

(25) Snell-Bergeon JK, Smith J, Dong F et al. Early childhood infections and the risk of islet autoimmunity: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetes Care* 2012;35:2553-2558.

(26) Frederiksen B, Liu E, Romanos J et al. Investigation of the vitamin D receptor gene (VDR) and its interaction with protein tyrosine phosphatase, non-receptor type 2 gene (PTPN2) on risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *J Steroid Biochem Mol Biol* 2013;133:51-57.

(27) Frederiksen BN, Kroehl M, Fingerlin TE et al. Association between vitamin D metabolism gene polymorphisms and risk of islet autoimmunity and progression to type 1 diabetes: the diabetes autoimmunity study in the young (DAISY). *J Clin Endocrinol Metab* 2013;98:E1845-E1851.

(28) Simpson M, Brady H, Yin X et al. No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia* 2011;54.

(29) Miller MR, Yin X, Seifert J et al. Erythrocyte membrane omega-3 fatty acid levels and omega-3 fatty acid intake are not associated with conversion to type 1 diabetes in children with islet autoimmunity: the Diabetes Autoimmunity Study in the Young (DAISY). *Pediatr Diabetes* 2011;12:669-675.

(30) Lamb MM, Miller M, Seifert JA et al. The effect of childhood cow's milk intake and HLA-DR genotype on risk of islet autoimmunity and type 1 diabetes: The Diabetes Autoimmunity Study in the Young. *Pediatr Diabetes* 2014.

(31) Chan CL, Taki I, Dong F et al. Comparison of Metabolic Outcomes in Children Diagnosed with Type 1 Diabetes Through Research Screening (Diabetes Autoimmunity Study in the Young [DAISY]) Versus in the Community. *Diabetes Technol Ther* 2015;17:649-656.

(32) Ziegler AG, Schmid S, Huber D, Hummel M, Bonifacio E. Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. *JAMA* 2003;290:1721-1728.

(33) Ilonen J, Hammais A, Laine AP et al. Patterns of beta-cell autoantibody appearance and genetic associations during the first years of life. *Diabetes* 2013;62:3636-3640.

(34) Krischer JP, Lynch KF, Schatz DA et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* 2015;58:980-987.

(35) Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med* 1986;314:1360-1368.

(36) von HM, Sanda S, Herold K. Type 1 diabetes as a relapsing-remitting disease? *Nat Rev Immunol* 2007;7:988-994.

(37) Krogvold L, Edwin B, Buanes T et al. Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. *Diabetes* 2015;64:1682-1687.

(38) Morgan NG, Richardson SJ. Enteroviruses as causative agents in type 1 diabetes: loose ends or lost cause? *Trends Endocrinol Metab* 2014.

(39) Pugliese A, Yang M, Kusmarteva I et al. The Juvenile Diabetes Research Foundation Network for Pancreatic Organ Donors with Diabetes (nPOD) Program: goals, operational model and emerging findings. *Pediatr Diabetes* 2014;15:1-9.

(40) Sane F, Caloone D, Gmyr V et al. Coxsackievirus B4 can infect human pancreas ductal cells and persist in ductal-like cell cultures which results in inhibition of Pdx1 expression and disturbed formation of islet-like cell aggregates. *Cell Mol Life Sci* 2013;70:4169-4180.

(41) Rodriguez-Calvo T, Ekwall O, Amirian N, Zapardiel-Gonzalo J, von Herrath MG. Increased immune cell infiltration of the exocrine pancreas: a possible contribution to the pathogenesis of type 1 diabetes. *Diabetes* 2014;63:3880-3890.

(42) Barker JM, Goehrig SH, Barriga K et al. Clinical characteristics of children diagnosed with type 1 diabetes through intensive screening and follow-up. *Diabetes Care* 2004;27:1399-1404.

(43) Farrell PM, Rosenstein BJ, White TB et al. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. *J Pediatr* 2008;153:S4-S14.

(44) Panepinto JA, Magid D, Rewers MJ, Lane PA. Universal versus targeted screening of infants for sickle cell disease: a cost-effectiveness analysis. *J Pediatr* 2000;136:201-208.

(45) Rewers M, Bugawan TL, Norris JM et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 1996;39:807-812.

(46) Standards of medical care in diabetes--2007. *Diabetes Care* 2007;30 Suppl 1:S4-S41.

(47) Steck AK, Dong F, Wong R et al. Improving prediction of type 1 diabetes by testing non-HLA genetic variants in addition to HLA markers. *Pediatr Diabetes* 2014;15:355-362.

(48) Yu L, Miao D, Scrimgeour L, Johnson K, Rewers M, Eisenbarth GS. Distinguishing persistent insulin autoantibodies with differential risk: nonradioactive bivalent proinsulin/insulin autoantibody assay. *Diabetes* 2012;61:179-186.

(49) Yu L, Dong F, Miao D, Fouts AR, Wenzlau JM, Steck AK. Proinsulin/Insulin autoantibodies measured with electrochemiluminescent assay are the earliest indicator of prediabetic islet autoimmunity. *Diabetes Care* 2013;36:2266-2270.

(50) Miao D, Guyer KM, Dong F et al. GAD65 autoantibodies detected by electrochemiluminescence assay identify high risk for type 1 diabetes. *Diabetes* 2013;62:4174-4178.

(51) Miao D, Steck AK, Zhang L et al. Electrochemiluminescence assays for insulin and glutamic Acid decarboxylase autoantibodies improve prediction of type 1 diabetes risk. *Diabetes Technol Ther* 2015;17:119-127.

(52) Krischer JP, Lynch KF, Schatz DA et al. The Six-year Incidence of Diabetes Associated Autoantibodies in Genetically At-risk Children: The TEDDY Study. *Diabetologia* 2014.

(53) Achenbach P, Koczwara K, Knopff A, Naserke H, Ziegler AG, Bonifacio E. Mature high-affinity immune responses to (pro)insulin anticipate the autoimmune cascade that leads to type 1 diabetes. *J Clin Invest* 2004;114:589-597.

(54) Giannopoulou EZ, Winkler C, Chmiel R et al. Islet autoantibody phenotypes and incidence in children at increased risk for type 1 diabetes. *Diabetologia* 2015;58:2317-2323.

(55) Babaya N, Yu L, Miao D et al. Comparison of insulin autoantibody: polyethylene glycol and micro-IAA 1-day and 7-day assays. *Diabetes Metab Res Rev* 2009;25:665-670.

(56) Yu L, Liu Y, Miao D et al. Triple chimeric islet autoantigen IA2-ZnT8WR to facilitate islet autoantibody determination. *J Immunol Methods* 2010;353:20-23.

(57) Bonifacio E, Yu L, Williams AK et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. *J Clin Endocrinol Metab* 2010;95:3360-3367.

(58) Bao F, Yu L, Babu S et al. One third of HLA DQ2 homozygous patients with type 1 diabetes express celiac disease-associated transglutaminase autoantibodies. *J Autoimmun* 1999;13:143-148.

(59) Steck AK, Dong F, Taki I, Hoffman M, Klingensmith GJ, Rewers MJ. Early hyperglycemia detected by continuous glucose monitoring in children at risk for type 1 diabetes. *Diabetes Care* 2014;37:2031-2033.

(60) Vehik K, Hamman RF, Lezotte D, Norris JM, Klingensmith GJ, Dabelea D. Childhood growth and age at diagnosis with Type 1 diabetes in Colorado young people. *Diabet Med* 2009;26:961-967.

(61) Fourlanos S, Narendran P, Byrnes GB, Colman PG, Harrison LC. Insulin resistance is a risk factor for progression to type 1 diabetes. *Diabetologia* 2004;47:1661-1667.

(62) Bingley PJ, Mahon JL, Gale EA. Insulin resistance and progression to type 1 diabetes in the European Nicotinamide Diabetes Intervention Trial (ENDIT). *Diabetes Care* 2007.

(63) Xu P, Cuthbertson D, Greenbaum C, Palmer JP, Krischer JP. Role of insulin resistance in predicting progression to type 1 diabetes. *Diabetes Care* 2007;30:2314-2320.

(64) Sardinha LB, Andersen LB, Anderssen SA et al. Objectively measured time spent sedentary is associated with insulin resistance independent of overall and central body fat in 9- to 10-year-old Portuguese children. *Diabetes Care* 2008;31:569-575.

(65) Martinez-Ramonde T, Alonso N, Cordido F et al. Importance of exercise in the control of metabolic and inflammatory parameters at the moment of onset in type 1 diabetic subjects. *Exp Clin Endocrinol Diabetes* 2014;122:334-340.

(66) Kann L, Kinchen SA, Williams BI et al. Youth risk behavior surveillance--United States, 1999. *MMWR CDC Surveill Summ* 2000;49:1-32.

(67) Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child* 1970;45:13-23.

(68) Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969;44:291-303.

(69) Vanhelst J, Beghin L, Turck D, Gottrand F. New validated thresholds for various intensities of physical activity in adolescents using the Actigraph accelerometer. *Int J Rehabil Res* 2010.

(70) Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 1998;30:777-781.

(71) Bergman BC, Howard D, Schauer IE et al. Features of hepatic and skeletal muscle insulin resistance unique to type 1 diabetes. *J Clin Endocrinol Metab* 2012;97:1663-1672.

(72) Schauer IE, Snell-Bergeon JK, Bergman BC et al. Insulin resistance, defective insulin-mediated fatty acid suppression, and coronary artery calcification in subjects with and without type 1 diabetes: The CACTI study. *Diabetes* 2011;60:306-314.

(73) Duca LM, Maahs DM, Schauer IE et al. Development and validation of a method to estimate insulin sensitivity in individuals with and without type 1 diabetes Schauer,I.E. Bergman,B.C. Nadeau,K.J. Bjornstad,P. Rewers,M. Snell-Bergeon,J.K. *J Clin Endocrinol Metab* 2015;In press.

(74) Perks SM, Roemmich JN, Sandow-Pajewski M et al. Alterations in growth and body composition during puberty. IV. Energy intake estimated by the youth-adolescent food-frequency questionnaire: validation by the doubly labeled water method. *Am J Clin Nutr* 2000;72:1455-1460.

(75) Harray AJ, Boushey CJ, Pollard CM et al. A Novel Dietary Assessment Method to Measure a Healthy and Sustainable Diet Using the Mobile Food Record: Protocol and Methodology. *Nutrients* 2015;7:5375-5395.

(76) Schap TE, Zhu F, Delp EJ, Boushey CJ. Merging dietary assessment with the adolescent lifestyle. *J Hum Nutr Diet* 2014;27 Suppl 1:82-88.

(77) Boushey CJ, Harray AJ, Kerr DA et al. How willing are adolescents to record their dietary intake? The mobile food record. *JMIR Mhealth Uhealth* 2015;3:e47.

(78) Fernandez JR, Redden DT, Pietrobelli A, Allison DB. Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J Pediatr* 2004;145:439-444.

(79) Kabiri LS, Hernandez DC, Mitchell K. Reliability, Validity, and Diagnostic Value of a Pediatric Bioelectrical Impedance Analysis Scale. *Child Obes* 2015;11:650-655.

(80) Proust-Lima C, Sene M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Stat Methods Med Res* 2014;23:74-90.

(81) lcmm: Extended Mixed Models Using Latent Classes and Latent Processes. R package version 1.7.2. 2015.
Ref Type: Online Source

(82) Nan B, Lin X, Lisabeth LD, Harlow SD. A varying-coefficient Cox model for the effect of age at a marker event on age at menopause. *Biometrics* 2005;61:576-583.

(83) Zhi W, Sharma A, Purohit S et al. Discovery and validation of serum protein changes in type 1 diabetes patients using high throughput two dimensional liquid chromatography-mass spectrometry and immunoassays. *Mol Cell Proteomics* 2011;10:M111.

(84) Zhang Q, Fillmore TL, Schepmoes AA et al. Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. *J Exp Med* 2013;210:191-203.

(85) Miersch S, Bian X, Wallstrom G et al. Serological autoantibody profiling of type 1 diabetes by protein arrays. *J Proteomics* 2013;94:486-496.

(86) Moulder R, Bhosale SD, Erkkila T et al. Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. *Diabetes* 2015;64:2265-2278.

(87) Anderson NL. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin Chem* 2010;56:177-185.

(88) Lollo B, Steele F, Gold L. Beyond antibodies: new affinity reagents to unlock the proteome. *Proteomics* 2014;14:638-644.

(89) Lee HS, Burkhardt BR, McLeod W et al. Biomarker discovery study design for type 1 diabetes in The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Diabetes Metab Res Rev* 2014;30:424-434.

(90) Biomarkers Of Beta Cell Stress In Type 1 Diabetes (BetaMarker). 2015.
Ref Type: Online Source

(91) Burch TC, Morris MA, Campbell-Thompson M, Pugliese A, Nadler JL, Nyalwidhe JO. Proteomic Analysis of Disease Stratified Human Pancreas Tissue Indicates Unique Signature of Type 1 Diabetes. *PLoS One* 2015;10:e0135663.

(92) Nishida Y, Aida K, Kihara M, Kobayashi T. Antibody-validated proteins in inflamed islets of fulminant type 1 diabetes profiled by laser-capture microdissection followed by mass spectrometry. *PLoS One* 2014;9:e107664.

(93) Schrimpe-Rutledge AC, Fontes G, Gritsenko MA et al. Discovery of novel glucose-regulated proteins in isolated human pancreatic islets using LC-MS/MS-based proteomics. *J Proteome Res* 2012;11:3520-3532.

(94) Crevecoeur I, Rondas D, Mathieu C, Overbergh L. The beta-cell in type 1 diabetes, what have we learned from proteomic studies? *Proteomics Clin Appl* 2015.

(95) Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 2012;9:555-566.

(96) Shi T, Fillmore TL, Sun X et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc Natl Acad Sci U S A* 2012;109:15395-15400.

(97) Kiebel GR, Auberry KJ, Jaitly N et al. PRISM: a data management system for high-throughput proteomics. *Proteomics* 2006;6:1783-1790.

(98) MacLean B, Tomazela DM, Shulman N et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;26:966-968.

(99) Shi T, Gao Y, Quek SI et al. A highly sensitive targeted mass spectrometric assay for quantification of AGR2 protein in human urine and serum. *J Proteome Res* 2014;13:875-882.

(100) *lme4: Linear mixed-effects models using Eigen and S4.* [2014.

(101) R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. 2015 ed. Vienna, Austria: 2015.

(102) Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23:2507-2517.

(103) Saligan LN, Fernandez-Martinez JL, deAndres-Galiana EJ, Sonis S. Supervised classification by filter methods and recursive feature elimination predicts risk of radiotherapy-related fatigue in patients with prostate cancer. *Cancer Inform* 2014;13:141-152.

(104) Webb-Robertson BJ, Wiberg HK, Matzke MM et al. Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res* 2015;14:1993-2001.

(105) Lanckriet GR, De BT, Cristianini N, Jordan MI, Noble WS. A statistical framework for genomic data fusion. *Bioinformatics* 2004;20:2626-2635.

(106) Jarman KH, Kreuzer-Martin HW, Wunschel DS et al. Bayesian-integrated microbial forensics. *Appl Environ Microbiol* 2008;74:3573-3582.

(107) Beagley N, Stratton KG, Webb-Robertson BJ. VIBE 2.0: visual integration for bayesian evaluation. *Bioinformatics* 2010;26:280-282.

(108) Webb-Robertson BJ, Kreuzer H, Hart G et al. Bayesian integration of isotope ratio for geographic sourcing of castor beans. *J Biomed Biotechnol* 2012;2012:450967.

(109) Madl T, Franklin S, Chen K, Montaldi D, Trappl R. Bayesian integration of information in hippocampal place cells. *PLoS One* 2014;9:e89762.

(110) Webb-Robertson BJ, Rewers M, Zhang Q, Waters KM, Metz TO. Bayesian integration ensemble-based feature selection for improved biomarker panel identification. *Appl Bioinformatics* 2015;in press.

(111) Jin Y, Sharma A, Bai S et al. Risk of T1D 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.

(112) Onengut-Gumuscu S, Chen WM, Burren O et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015;47:381-386.

(113) Torn C, Hadley D, Lee HS et al. Role of Type 1 Diabetes-Associated SNPs on Risk of Autoantibody Positivity in the TEDDY Study. *Diabetes* 2015;64:1818-1829.

(114) Rolandsson O, Hagg E, Nilsson M, Hallmans G, Mincheva-Nilsson L, Lernmark A. Prediction of diabetes with body mass index, oral glucose tolerance test and islet cell autoantibodies in a regional population. *J Intern Med* 2001;249:279-288.

(115) Webb-Robertson BJ, McCue LA, Beagley N et al. A Bayesian integration model of high-throughput proteomics and metabolomics data for improved early detection of microbial infections. *Pac Symp Biocomput* 2009;451-463.

(116) Ibanez C, Simo C, Garcia-Canas V, Cifuentes A, Castro-Puyana M. Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics: a review. *Anal Chim Acta* 2013;802:1-13.

(117) Dai Q, Cheng JH, Sun DW, Zeng XA. Advances in feature selection methods for hyperspectral image processing in food industry applications: a review. *Crit Rev Food Sci Nutr* 2015;55:1368-1382.