



The First Genome-Wide Association Study for Type 2 Diabetes in Youth: The Progress in Diabetes Genetics in Youth (ProDiGY) Consortium

Shylaja Srinivasan,¹ Ling Chen,² Jennifer Todd,³ Jasmin Divers,⁴ Samuel Gidding,⁵ Steven Chernausek,⁶ Rose A. Gubitosi-Klug,⁷ Megan M. Kelsey,⁸ Rachana Shah,⁹ Mary Helen Black,¹⁰ Lynne E. Wagenknecht,¹¹ Alisa Manning,^{2,12} Jason Flannick,^{12,13} Giuseppina Imperatore,¹⁴ Josep M. Mercader,^{2,12,15} Dana Dabelea,¹⁶ and Jose C. Florez,^{2,12,15} on behalf of the ProDiGY Consortium*

Diabetes 2021;70:996–1005 | <https://doi.org/10.2337/db20-0443>

The prevalence of type 2 diabetes in youth has increased substantially, yet the genetic underpinnings remain largely unexplored. To identify genetic variants predisposing to youth-onset type 2 diabetes, we formed ProDiGY, a multiethnic collaboration of three studies (TODAY, SEARCH, and T2D-GENES) with 3,006 youth case subjects with type 2 diabetes (mean age 15.1 ± 2.9 years) and 6,061 diabetes-free adult control subjects (mean age 54.2 ± 12.4 years). After stratifying by principal component-clustered ethnicity, we performed association analyses on ~10 million imputed variants using a generalized linear mixed model incorporating a genetic relationship matrix to account for population structure and adjusting for sex. We identified seven genome-wide significant loci, including the novel locus rs10992863 in *PHF2* ($P = 3.2 \times 10^{-8}$; odds ratio [OR] = 1.23). Known loci identified in our analysis include rs7903146 in *TCF7L2* ($P = 8.0 \times 10^{-20}$; OR 1.58), rs72982988 near *MC4R* ($P = 4.4 \times 10^{-14}$; OR 1.53), rs200893788 in *CDC123* ($P = 1.1 \times 10^{-12}$; OR 1.32), rs2237892 in *KCNQ1* ($P = 4.8 \times 10^{-11}$; OR 1.59),

rs937589119 in *IGF2BP2* ($P = 3.1 \times 10^{-9}$; OR 1.34), and rs113748381 in *SLC16A11* ($P = 4.1 \times 10^{-8}$; OR 1.04). Secondary analysis with 856 diabetes-free youth control subjects uncovered an additional locus in *CPEB2* ($P = 3.2 \times 10^{-8}$; OR 2.1) and consistent direction of effect for diabetes risk. In conclusion, we identified both known and novel loci in the first genome-wide association study of youth-onset type 2 diabetes.

Type 2 diabetes (T2D) is a global epidemic and an acknowledged significant population health issue in adults (1). However, until recently, T2D was not thought to be relevant to youth. This paradigm has changed with the rise in obesity among youth worldwide (2). In the U.S., the population-based study SEARCH for Diabetes in Youth showed that the unadjusted incidence rates of T2D in youth 10–19 years of age increased by 7.1% annually between 2002 and 2012, with the increase more dramatic in

¹Division of Pediatric Endocrinology, University of California, San Francisco, San Francisco, CA

²Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

³Division of Pediatric Endocrinology, University of Vermont, Burlington, VT

⁴NYU Langone Health, New York, NY

⁵Genomic Medicine Institute, Geisinger, Danville, PA

⁶Pediatric Diabetes and Endocrinology Section, University of Oklahoma College of Medicine, Oklahoma City, OK

⁷Pediatric Endocrinology, Diabetes, and Metabolism, Case Western Reserve University and Rainbow Babies and Children's Hospital, Cleveland, OH

⁸Pediatric Endocrinology, University of Colorado School of Medicine, Aurora, CO

⁹Pediatric Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA

¹⁰Janssen Pharmaceuticals, Spring House, PA

¹¹Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC

¹²Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA

¹³Department of Pediatrics, Boston Children's Hospital, Boston, MA

¹⁴Centers for Disease Control and Prevention, Atlanta, GA

¹⁵Diabetes Research Center, Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA

¹⁶Department of Epidemiology, University of Colorado School of Public Health, Aurora, CO

Corresponding author: Jose C. Florez, jcflorez@mgh.harvard.edu

Received 30 April 2020 and accepted 18 January 2021

This article contains supplementary material online at <https://doi.org/10.2337/figshare.13607852>.

*A complete list of the ProDiGY Consortium working group members can be found in the supplementary material online.

© 2021 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

minority race and ethnic groups (3). It has become apparent that youth-onset T2D displays an aggressive course and is associated with greater mortality, earlier complications, and unfavorable cardiovascular disease risk when compared with adult-onset T2D or youth-onset type 1 diabetes (4–6). These trends, coupled with the limited treatment options for the management of T2D in youth, makes this an emerging health concern (7).

T2D is a complex disease, influenced by the combination of genetic, epigenetic, environmental, and behavioral factors and their interactions (8). The heritability of T2D in youth is highlighted by the high concordance rates of T2D in identical twins, the presence of a strong family history of T2D in affected youth, and the disproportionate prevalence of T2D in certain racial and ethnic groups, such as Native American, Hispanic, and African American populations (7,9). However, while there have been rapid advances in the understanding of the genetics of T2D in adults, our understanding of the genetics underlying youth-onset T2D has lagged behind (10). While a few studies have examined candidate genes for T2D in youth, the genetics of T2D in youth remain largely understudied (11,12).

Progress in Diabetes Genetics in Youth (ProDiGY) is a collaboration of two pediatric T2D studies, Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) and SEARCH for Diabetes in Youth, along with the Type 2 Diabetes Genetics Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortium, a large collaborative effort to find genetic variants that influence risk of T2D in adults (10). In this report, we set out to comprehensively examine the genetics of T2D in youth in ProDiGY by comparing subjects from TODAY and SEARCH with two control cohorts: adults older than 50 years of age without T2D and youth younger than 18 years of age without diabetes. We also examined the effect of known genetic variants associated with T2D in adults in our youth-onset diabetes cohort.

RESEARCH DESIGN AND METHODS

Description of Participants

ProDiGY is a multiethnic resource that includes data from >3,000 case subjects with youth-onset T2D and 6,000 diabetes-free adult control subjects with both genome-wide genotyping and whole-exome sequence data. The collaboration brings together data from 449 youth with T2D from the TODAY study as well as data from >2,000 youth with T2D from a TODAY ancillary genetics study. It also includes data from 468 youth with T2D from SEARCH for Diabetes in Youth, as well as access to data from >10,000 adult subjects and 10,000 control subjects from T2D-GENES (13). The TODAY and SEARCH studies are described in detail elsewhere (14,15).

Briefly, TODAY was a multicenter randomized controlled trial that enrolled 699 participants (aged 10–17 years) with T2D between 2004 and 2009. Participants were overweight or had obesity (BMI \geq 85th percentile for age, sex, and height), with negative pancreatic

autoantibodies (GAD-65 and tyrosine phosphatase) and a fasting C-peptide concentration >0.6 ng/mL. Participants were randomized to one of three arms—metformin alone, metformin plus rosiglitazone, and metformin plus lifestyle intervention—and followed on average for 3.86 years. The results of the TODAY clinical trial have been described elsewhere (7).

TODAY Genetics is an ancillary study of TODAY distinct from the original clinical trial. To qualify, participants must have been diagnosed with T2D prior to 18 years of age and have a documented BMI \geq 85th percentile at the time of diagnosis. Data and sample collection occurred at a one-time research visit at 1 of 25 clinical sites. During this visit, a self-report questionnaire on family and medical history was administered and blood samples were drawn for DNA extraction and for analysis of glucose, C-peptide, and autoantibodies.

SEARCH is a population-based prospective registry study launched in 2000 that ascertained diabetes in youth diagnosed at <20 years of age in the U.S. at study centers located in five states: South Carolina, Ohio, Colorado, California, and Washington. Youth with T2D were identified by physician report.

In this study, we examined a total of 3,006 youth from TODAY, TODAY Genetics, and SEARCH. Additionally, 856 non-Hispanic White and African American youth without diabetes recruited from Ohio, Colorado, and South Carolina as part of the SEARCH case-control study were used in this analysis (16). The TODAY and SEARCH protocols were approved by the institutional review boards of each participating institution. Participants provided written informed parental consent and child assent, including consent and assent specifically for genetic testing.

Samples for the adult control subjects in T2D-GENES were drawn from 12 studies from T2D-GENES and are described fully in Supplementary Table 1. T2D status was determined according to study-specific criteria. A total of 6,061 adult control subjects older than 50 years of age and free of diabetes were selected from non-Hispanic White, Hispanic, and African American race/ethnicities to match the race/ethnicity distribution in the youth subjects. All individuals provided informed consent, and all samples were approved for use by their institution's institutional review board or ethics committee.

Genotyping, Imputation and Quality Control

Samples for ProDiGY were genotyped on the Infinium genome-wide association study (GWAS) array, a complement to the Nexome platform by the Genetic Analysis Platform at the Broad Institute. The directly genotyped data were called by using the Autocall algorithm. We imputed our genotyped data on the Michigan Imputation Server against the 1000 Genomes Phase 3 v5 panel as the reference. All quality control steps were implemented in PLINK2 and R-3.4. Samples were filtered for sex discrepancies, low sample call rate, and close relatedness; single nucleotide polymorphisms (SNPs) were filtered for minor

allele frequency, low SNP call rate, and lack of compliance with Hardy-Weinberg equilibrium. The imputation threshold (R^2) was set at 0.5. Principal component analysis was performed on the genome-wide identity-by-descent pairwise distances in conjunction with complete linkage clustering of individual after merging with 1000 Genomes data. After cleaning, ~10 million variants were available for analyses (17).

Statistical Analyses

We performed association tests for T2D using a generalized linear mixed model (GLMM) with a genetic relationship matrix to account for population substructure. Covariates included sex and BMI when available. We used the Efficient and Parallelizable Association Container Toolbox (EPACTS) to run GLMM tests to obtain P value estimates and Wald tests to obtain odds ratio (OR) estimates within each racial or ethnic group. A total of 898 of the ProDiGY participants had BMI z -score information. For the purpose of our analyses, BMI z score was calculated in adults according to BMI-for-age charts at 20 years using LMS parameters from https://www.cdc.gov/growthcharts/percentile_data_files.htm/. To account for the putative effect of BMI, a sensitivity analyses was performed adjusting for sex and BMI z score in the GLMM. A binomial test was performed to compare effects of known SNPs associated with T2D in ProDiGY with data previously reported in adults. A GLMM was applied for P value estimates by using EPACTS in each race and ethnic group. Meta-analyses were then conducted by using METAL to combine results from each racial and ethnic group. A threshold of $P < 5 \times 10^{-8}$ was used to define genome-wide significance. To fine-map the novel SNP, a credible set analysis was performed using a Bayesian approach (18). This method is described in further detail in the Supplementary Material.

Data and Resource Availability

The data sets generated analyzed in the current study are available in dbGap (dbGaP Study Accession: phs001511.v1.p1, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001511.v1.p1). No applicable resources were generated or analyzed during the current study.

RESULTS

Baseline demographics of case and control subjects are summarized in Table 1. The analyses included 3,006 youth with T2D and 6,061 adult control subjects who were selected to be older than 50 years old and free of diabetes. In secondary analyses, we also conducted a GWAS with 856 youth control subjects without diabetes who were recruited as part of the SEARCH case-control study. Mean age of the affected youth was 15.1 ± 2.8 years compared with 54.2 ± 12.4 years for the adult control subjects. The majority of participants were female and non-White. As expected, the youth had high levels of obesity, with a mean BMI z score of 2.17 ± 0.6 compared with 1.08 ± 0.8 for

Table 1—Baseline demographics of ProDiGY participants

	Case subjects	Adult control subjects	Youth control subjects
Total, n	3,006	6,061	856
TODAY	449		
TODAY Genetics SEARCH	2,084		
	468		
Age, years	15.1 ± 2.8	54.2 ± 12.4	14.5 ± 2.3
Female, %	62	57	52
Race/ethnicity, n (%)			
Non-Hispanic			
White	664 (22)	1,434 (24)	542 (63)
African American	1,068 (36)	1,068 (18)	314 (37)
Hispanic	1,274 (42)	3,559 (59)	0
BMI z score*	2.17 ± 0.6	1.08 ± 0.8	0.61 ± 1.03
TODAY	2.21 ± 0.5		
TODAY Genetics	N/A		
SEARCH	2.12 ± 0.7		
Fasting glucose (mg/dL)	154.36 ± 88.99	87.93 ± 10.61	84.20 ± 8.49
Fasting insulin (μ U/mL)	31.16 ± 26.55	N/A	N/A
HbA _{1c} %	6.75 ± 1.86	N/A	N/A

Data are mean \pm SD unless otherwise indicated. N/A, not available. *BMI available in 898 TODAY and SEARCH participants.

the adult control subjects. Matching of youth subjects with adult control subjects was successful, as evidenced by little systematic deviation ($\lambda_{GC} = 1.08$) of the observed distribution from the expected distribution under the null hypothesis of no association in the quantile-quantile (QQ) plot (Fig. 1A); the QQ plot for youth subjects versus youth control subjects is comparable and shown in Fig. 1B ($\lambda_{GC} = 1.09$).

Figure 2 shows the Manhattan plot from the trans-ethnic meta-analysis. We identified seven genome-wide significant findings, including variants in or near *TCF7L2*, *MC4R*, *CDC123*, *KCNQ1*, *IGF2BP2*, *PHF2*, and *SLC16A11* (Table 2). The association signal in *PHF2* is novel, whereas the remaining six variants have been associated with T2D in adults; relative effect sizes are compared in Table 2. To fine-map the region of the novel *PHF2* signal, a credible set analysis was performed (reference). The 99% credible set contained a total of 38 possible causal variants. The 99% credible set of variants for this region is summarized in Supplementary Table 2. In the subset of 898 participants in whom BMI data were available, signals were attenuated but remained nominally significant after adjustment for BMI (Table 3). Association tests for youth subjects versus youth control subjects identified a novel genome-wide significant variant, rs2604566 in *CPEB2* ($P = 3.2 \times 10^{-8}$; OR 2.1) (Fig. 3). This variant has been shown to be nominally associated with T2D in adults (OR 1.11; $P = 0.014$ in Joint T2D-CHD GWAS) (19).

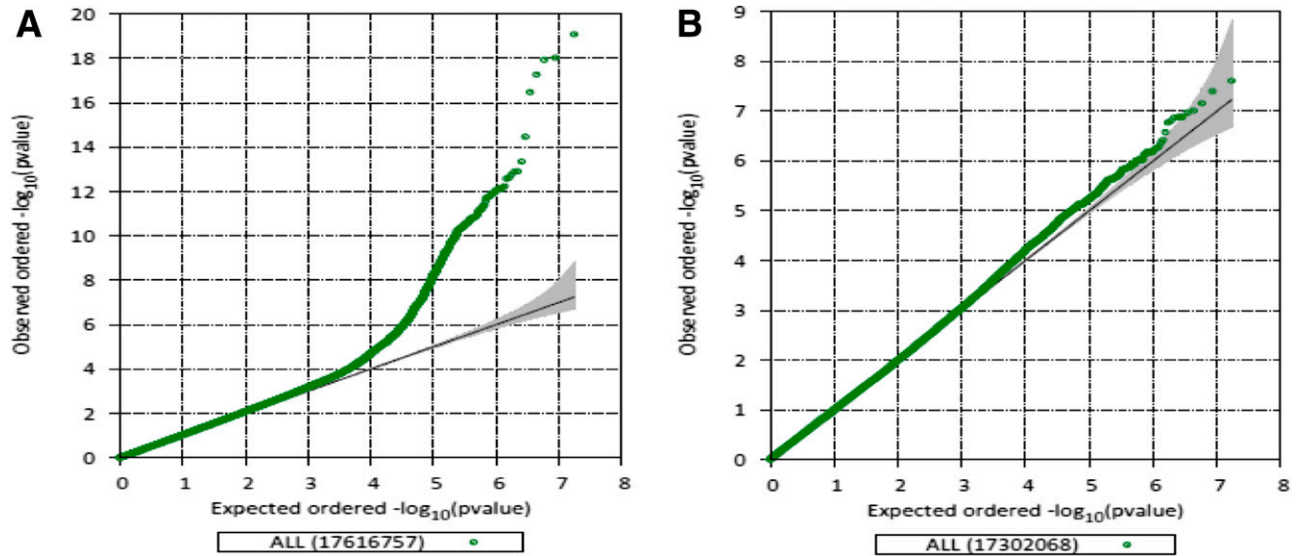


Figure 1—A: QQ plot for case subjects with youth-onset T2D vs. adult control subjects without diabetes. The x-axis shows the expected distribution and the y-axis shows the observed distribution of findings. $\lambda_{GC} = 1.08$. B: QQ plot for case subjects with youth-onset T2D vs. youth control subjects without diabetes. The x-axis shows the expected distribution and the y-axis shows the observed distribution of findings. $\lambda_{GC} = 1.09$.

Our lead SNP rs10992863 in *PHF2* is also associated with height and BMI (<https://t2d.hugeamp.org/variant.html?variant=rs10992863>). However, Locus Zoom plots for height and BMI at this locus demonstrate that rs10992863 is not the top SNP for height and BMI and is in modest disequilibrium ($r^2 = 0.1282$) with rs9650755, the top SNP for height and BMI in this locus. This suggests that the youth-onset T2D-associated SNP rs10992863 may also be associated with BMI, but it is not the same signal. To confirm this, we conducted a colocalization analysis that shows that rs10992863 (the most significant SNP for T2D) and rs9650755 (the most significant SNP in this locus for BMI) are independent of each other

(Supplementary Fig. 1). To test whether the association of rs10992863 with BMI is primarily driven by the BMI-associated variant rs9650755, we tested the association of rs10992863 with BMI in the UK Biobank, conditioning on rs9650755 (Supplementary Table 3). Our results confirm that rs10992863 is associated with BMI ($P = 5.1 \times 10^{-7}$; $\beta = 0.062$), but the association is significantly weakened in the presence of rs9650755 ($P = 2.5 \times 10^{-2}$; $\beta = 0.030$). However, rs9650755 still remained highly significant after conditioning on rs10992863, with a threefold strengthening of the effect size ($P = 1.9 \times 10^{-11}$; $\beta = 0.090$). To further validate these findings, we performed conditional analyses of the above SNPs in ProDiGY (Supplementary

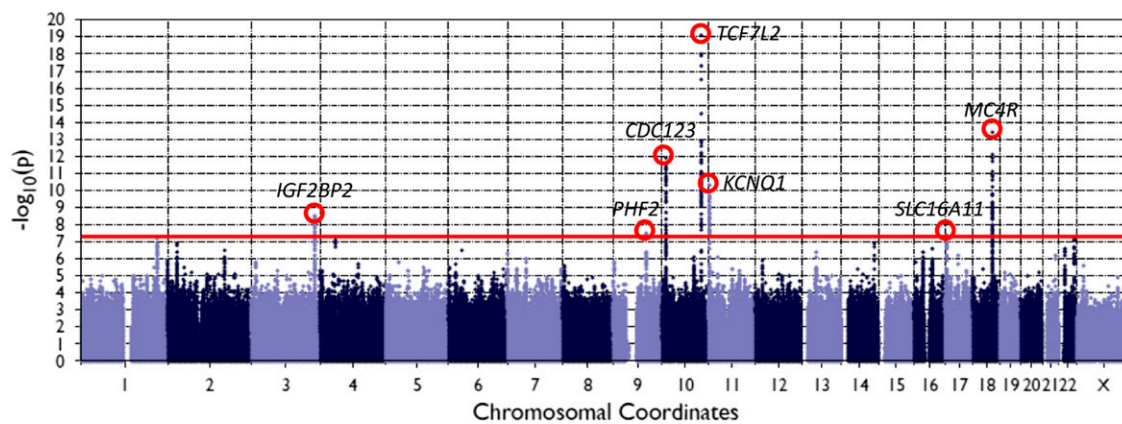


Figure 2—Manhattan plot for youth case subjects with T2D vs. adult control subjects without diabetes. The red horizontal line in the plot indicates the genome-wide significance P value threshold of 5×10^{-8} . The closest genes for the seven genome-wide significant findings are shown circled in red.

Table 2—Genome-wide significant findings for youth-onset T2D

SNP	Chromosome	Position	Closest gene	Type	Effect allele	Non-effect allele	OR vs. adult control subjects	<i>P</i> value vs. adult control subjects	OR in adults (20)	<i>P</i> value in adults (20)	OR vs. youth control subjects	<i>P</i> value vs. youth control subjects
rs7903146	10	114758349	<i>TCF7L2</i>	Intron	T	C	1.58	8.0×10^{-20}	1.37	5.8×10^{-447}	1.24	0.00091
rs72982988	18	57802714	<i>MC4R</i>	Intron	A	G	1.53	4.4×10^{-14}	1.05	7.6×10^{-13}	1.20	0.012
rs200893788	10	12283161	<i>CDC123</i>	Intron	T	TAGTA	1.32	1.1×10^{-12}	1.09	1.5×10^{-32}	1.02	0.49
rs2237892	11	2839751	<i>KCNQ1</i>	Intron	C	T	1.59	4.8×10^{-11}	1.12	6.0×10^{-32}	1.25	0.13
rs937589119	3	185537636	<i>IGF2BP2</i>	Insertion	GT	G	1.34	3.1×10^{-9}	1.14	9.0×10^{-36}	1.08	0.52
rs10992863	9	96445803	<i>PHF2</i>	Intergenic	G	A	1.23	3.2×10^{-8}	Novel for T2D*	Novel for T2D*	1.16	0.087
rs113748381	17	6953155	<i>SLC16A11</i>	Intergenic	A	G	1.04	4.1×10^{-8}	1.29 (25)	5.5×10^{-12} (25)	1.36	0.16

A total of 3,006 youth case subjects with T2D were compared with 6,061 adult or 856 youth control subjects. Association tests were adjusted for age, sex, and genetic relationship matrix. **P* value for association with T2D in adults = 8.3×10^{-3} (Type 2 Diabetes Knowledge Portal, <https://t2d.hugeamp.org/variant.html?variant=rs10992863>) (20).

Table 3—Genome-wide significant findings for subjects with youth-onset T2D vs. adult control subjects with and without adjustment for BMI

SNP	Chromosome	Position	Gene	Type	Effect allele	Non-effect allele	OR	P value	BMI-adjusted OR	BMI-adjusted P value
rs7903146	10	114758349	<i>TCF7L2</i>	Intron	T	C	1.52	3.5×10^{-11}	1.59	2.0×10^{-11}
rs72982988	18	57802714	<i>MC4R</i>	Intron	A	G	1.55	2.2×10^{-9}	1.45	2.9×10^{-6}
rs200893788	10	12283161	<i>CDC123</i>	Intron	T	TAGTA	1.35	1.2×10^{-7}	1.40	7.6×10^{-8}
rs2237892	11	2839751	<i>KCNQ1</i>	Intron	C	T	1.61	3.9×10^{-6}	1.70	5.4×10^{-6}
rs10992863	9	185537636	<i>PHF2</i>	Insertion	G	A	1.33	4.0×10^{-6}	1.25	1.2×10^{-4}
rs937589119	3	96445803	<i>IGF2BP2</i>	Intergenic	GT	G	1.32	4.7×10^{-5}	1.37	2.5×10^{-5}
rs113748381	17	6953155	<i>SLC16A11</i>	Intergenic	A	G	0.99	4.1×10^{-3}	0.92	3.6×10^{-2}

Association tests for 898 case subjects with youth-onset T2D in whom BMI information was available were adjusted for sex and genetic relationship matrix and compared before and after BMI z-score adjustment.

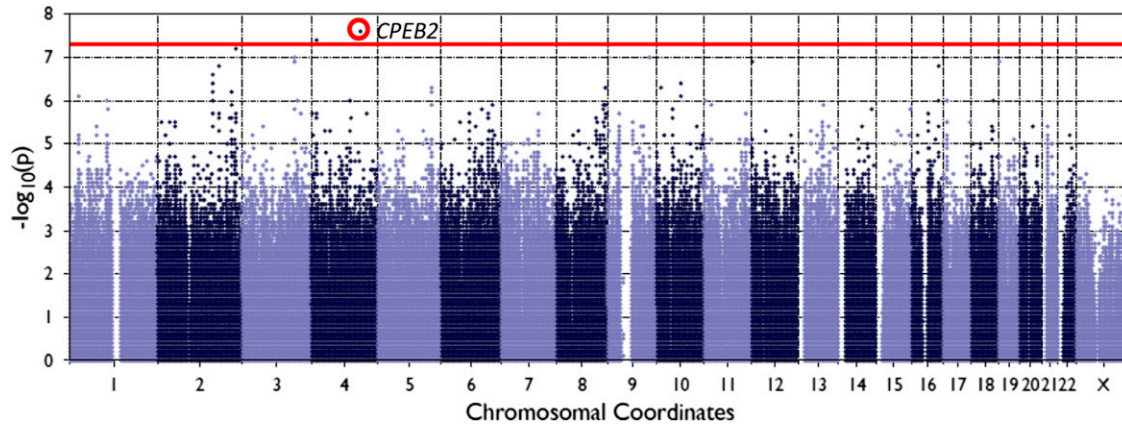


Figure 3—Manhattan plot for youth case subjects with T2D vs. youth control subjects without diabetes. The red horizontal line in the plot indicates the genome-wide significance P value threshold of 5×10^{-8} . The closest gene for the genome-wide significant finding is shown circled in red.

Table 3). When we conditioned rs10992863 on rs9650755, the results remained significant ($P = 4.4 \times 10^{-8}$; $\beta = 0.319$). However, rs9650755 lost the association with T2D when conditioned on rs10992863 ($P = 0.34$; $\beta = 0.046$), ruling out that its BMI association is driving the observed effect on T2D. Altogether, our results confirm that the association of rs10992863 with T2D in ProDiGY is independent of the association with BMI and that the association of rs10992863 with BMI is just a shadow of the association of the main BMI SNP rs9650755 driven by its moderate linkage disequilibrium with rs10992863.

Finally, we examined whether variants previously associated with T2D in adults at genome-wide significant levels were also associated with youth-onset T2D in ProDiGY. Association results in ProDiGY for 303 published variants in adults (20) is shown in Supplementary Table 2. This analysis revealed substantial consistency in direction of effects, with 86% of the SNPs showing the same direction of effect for the analyses of youth subjects versus adult control subjects (binomial test $P < 2.2 \times 10^{-16}$) and 71% of the SNPs showing the same direction of effect for the analyses of youth subjects versus youth control subjects (binomial test $P < 1.3 \times 10^{-13}$) despite limited power to achieve similar levels of statistical significance. Sensitivity analyses done with non-Hispanic White samples also showed a consistent direction of effect (binomial test

$P < 2.2 \times 10^{-16}$ for analyses with adult control subjects and $P < 1.2 \times 10^{-13}$ for analyses with youth control subjects).

To further explore the relationship of genetic variants with youth-onset T2D, we constructed polygenic risk scores in ProDiGY using risk alleles and effect sizes from known T2D genetic variants previously identified in adults (20) (Supplementary Fig. 2). We compared associations of the polygenic risk score between case subjects with youth-onset T2D and adult control subjects, youth case and youth control subjects, and adult subjects with T2D from T2D-GENES with adult control subjects. In all three scenarios, we found that the polygenic risk score was significantly associated with T2D with the same direction of effect (Table 4). Additionally, we found that the OR for T2D was higher in the youth case subject versus adult control analyses compared with the all adult analyses, without overlapping CIs. Altogether, these findings validate our initial findings related to T2D in youth as well as highlight the higher aggregate genetic risk burden of diabetes variants in youth when compared with adults with T2D.

DISCUSSION

Our findings in ProDiGY provide the first large-scale evaluation of the genetics of T2D in youth. Through GWAS, we discovered seven genome-wide significant variants

Table 4—Comparison of polygenic diabetes genetic risk z scores in youth case subjects vs. adult control subjects, youth case vs. youth control subjects, and adult case vs. adult control subjects

	Youth case subjects with T2D vs. adult control subjects			Youth case subjects with T2D vs. youth control subjects		Adult case subjects with T2D vs. adult control subjects		
	EUR	AFR	AMR	EUR	AFR	EUR	AFR	AMR
<i>P</i> value	2×10^{-16}	3.7×10^{-15}	$<2 \times 10^{-16}$	$<2 \times 10^{-16}$	0.0028	$<2 \times 10^{-16}$	5.7×10^{-5}	$<2 \times 10^{-16}$
OR	2.44	1.45	2.49	2.21	1.26	1.73	1.20	1.44
95% CI	2.17, 2.75	1.32, 1.60	2.28, 2.72	1.91, 2.59	1.08, 1.47	1.60, 1.89	1.10, 1.32	1.36, 1.53

AFR, African American; AMR, Hispanic American; EUR, European.

associated with T2D in youth, including a novel variant in *PHF2* shown to be nominally associated with T2D in adults (21). *PHF2* encodes histone demethylase plant homeodomain finger 2, a transcriptional coactivator of the transcription factor ChREBP. Through mouse and in vitro studies, *PHF2* has been shown to be involved in adipogenesis and fat storage through the regulation of CEBP α and peroxisome proliferator-activated receptor γ transcriptional activities in adipose tissue (22,23). Additionally, mice with targeted disruption of *Arid5b* (AT-rich interactive domain D), a specific *PHF2* coactivator partner, show a reduction in their white adipose tissue mass as a result of reduced peroxisome proliferator-activated receptor γ activity, suggesting that *PHF2* could play a role in the regulation of glucose and lipid homeostasis, potentially by exerting its effect during adipocyte development. *PHF2* overexpression in mouse liver leads to hepatic steatosis via epigenetic effects on ChREBP that lead to increased mono-unsaturated fatty acid production (24).

In GTEx (<https://gtexportal.org/home>), our lead SNP rs10992863 is also not an expression quantitative trait locus (eQTL) for *PHF2* in any tissue, although a potential effect on expression of *PHF2* at specific developmental stages has not been ruled out. Interestingly, it is an eQTL in many tissues, including adipose, for *RP11-165J3.6* (all increased expression per copy of the A allele, the risk allele for T2D). This region appears to have no coding exons, a single noncoding transcript, and sits upstream of *PHF2*. Also, rs10992675, which is ~482 kb upstream of rs10992863 and not in linkage disequilibrium with rs10992863 in any population, is an eQTL for *RP11-526D8.11*, a pseudogene region near *RP11-165J3.6*. On chromosome 9, regions are positioned as *RP11-526D8.11*—rs10992675—*RP11-165J3.6*—*PHF2*—rs10992863. Taken together, there may be a regulatory role for rs10992863 of upstream elements, which in turn may regulate expression of *PHF2* in adipose tissue.

The variants in *TCF7L2*, *MC4R*, *CDC123*, *KCNQ1*, *IGF2BP2*, and *SLC16A11* have all been previously associated with T2D in adults (20,25). In addition, a prior study in SEARCH reported that genetic variation in *TCF7L2* is associated with an increased risk of T2D in African American youth, with the OR for diabetes stronger in African American than in non-Hispanic White youth (12). The divergent ethnicity-based results did not replicate in ProDiGY, suggesting that the earlier findings in SEARCH might be due to statistical fluctuations in the context of smaller sample sizes.

The analyses comparing findings in ProDiGY with known variants associated with T2D in adults showed that for the majority of variants, the direction of effect was consistent between adults and youth with T2D. The fact that only 6 of the ≥ 400 genetic variants that have been associated with T2D in adults achieved genome-wide significance in ProDiGY is likely explained by the relatively small sample size compared with adult cohorts (20). Overall, these results demonstrate that there is

significant overlap between the genetics of T2D in adults and youth.

As obesity is a risk factor for the development of T2D, it is important to evaluate the role of BMI in genetic associations of T2D to assess whether a particular genetic variant that influences T2D risk is mediated by change in BMI. We had BMI data available in 898 youth participants with T2D from TODAY and SEARCH, in addition to the adult and youth control subjects. BMI attenuated the T2D association signals in youth, with only rs7903146 in *TCF7L2* remaining genome-wide significant. However, the direction of effect remained consistent for our top findings, despite the reduction in sample size. One of the top findings in ProDiGY was rs72982988 in *MC4R*, a gene well-known to influence obesity with the effect size for T2D in youth greater than reported in adults (20). rs72982988 is associated with T2D in adults at genome-wide level of significance (OR 1.05; $P = 1.4 \times 10^{-10}$), even after adjustment for BMI (26). It is possible that the role of obesity-related genes in youth varies in comparison with adults and is a topic to be explored further.

In secondary analyses, we conducted a separate GWAS comparing our 3,006 youth subjects with 856 youth control subjects free of diabetes from non-Hispanic White and African American ancestries. We identified a genome-wide significant variant, rs2604566 in *CPEB2* on chromosome 4 ($P = 3.2 \times 10^{-8}$; OR 2.1), that has not been previously reported in the literature. Given that this finding was only uncovered in the youth analyses, the likely significance is that genetic variation in *CPEB2* is only active in youth and displays an age-dependent phenomenon that occurs early in development. *CPEB2* encodes cytoplasmic polyadenylation element binding protein 2, an mRNA binding protein that regulates cytoplasmic polyadenylation of mRNA. *CPEB2*-knockout mice show reduced uncoupling protein 1 level and impaired thermogenesis in brown adipose tissue (27). Functional studies are needed to further explore the role of this gene on T2D risk in youth.

Clinically, youth-onset T2D displays a more aggressive disease course than adult-onset diabetes, with more rapid β -cell decline (4–6). We therefore hypothesized that there might be significant genetic differences between youth and adult-onset diabetes or that genetic effects might be more prominent in youth. Our results suggest that there is significant overlap in the genetic architecture of T2D in youth and adults. In our cohort, genetic effect sizes are stronger in youth for some of our top findings; however, CIs around the ORs in youth and adults largely overlap due to the smaller sample size of the youth GWAS. Another potential difference compared with adults is that several of the top findings in our analysis seem to influence T2D risk through obesity, potentially suggestive of the stronger impact that obesity-mediated genetic effects have on diabetes risk in the younger population. This also supports the potential of these variants influencing obesity and accelerating T2D onset at a younger age in these youth.

Further work is needed to understand the genetic determinants of T2D in youth and to explore reasons for differences compared with adults. In addition, the genetic determinants of medication response in the TODAY study could be examined. All ProDiGY samples have also undergone whole-exome sequencing to enable the analysis of rare variants, which is a planned next step. However, large-scale studies in adults have shown that rare variants only explain a small fraction of the heritability of T2D and that extremely large sample sizes are needed to uncover exome-wide significant findings, which presents a challenge considering our sample size (13). In addition, other potential modulators, such as gene \times environment interactions and epigenetic modifiers, remain to be explored.

The biggest strength of our study is that it brings together two large available youth-onset T2D studies, TODAY and SEARCH, along with the TODAY Genetics ancillary study and the large adult T2D consortium T2D-GENES, compiling the largest collection of subjects with youth-onset T2D and adult control subjects known to us. While previous smaller scale genetic studies of youth-onset T2D have been limited in their genetic diversity, our study is multiethnic, a key advantage considering that the disease predominantly affects non-White youth. A weakness of the study is the limited BMI data because BMI was not collected in the TODAY Genetics cohort. Additionally, whereas our sample size is large compared with other pediatric studies, it is still limited from a GWAS standpoint, and further collaborations are needed to increase our sample size for more effective genetic exploration. Another limitation is that in line with the usual practice in GWAS for T2D in adults, case subjects with monogenic diabetes were not screened out. As part of a separate ongoing project, we have estimated that the prevalence of these subjects is only 2% in ProDiGY, and thus, it is unlikely that removal of such a small number of subjects would drastically influence results.

In conclusion, in ProDiGY, we have discovered seven genome-wide significant associations with T2D in youth, one novel and the remaining known, providing initial insight into the genetic architecture of T2D in youth. Importantly, ProDiGY has established a cohort of subjects with youth-onset T2D who serve as a valuable resource for future genetic investigation.

Acknowledgments. The TODAY and SEARCH for Diabetes in Youth studies are indebted to the many youth and families and the health care providers, whose participation made this study possible.

Funding. S.S. is supported by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant 1K23DK120932-01A. J.M.M. is supported by American Diabetes Association Innovative and Clinical Translational Science Award 1-19-ICTS-068. SEARCH 3/4 was supported by the South Carolina Clinical & Translational Research Institute at the Medical University of South Carolina, National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS) grants UL1 TR000062 and UL1 TR001450; Seattle Children's Hospital and the University of Washington, NIH/NCATS grant UL1 TR00423; University of Colorado Pediatric Clinical and Translational Research Center,

NIH/NCATS grant UL1 TR000154; the Barbara Davis Center for Childhood Diabetes at the University of Colorado at Denver, Diabetes Endocrinology Research Centers NIH grant P30 DK57516; the University of Cincinnati, NIH/NCATS grants UL1 TR000077 and UL1 TR001425; and the Children with Medical Handicaps Program managed by the Ohio Department of Health. TODAY and TODAY Genetics was completed with funding from NIDDK and the NIH Office of the Director through grants U01-DK61212, U01-DK61230, U01-DK61239, U01-DK61242, and U01-DK61254; the National Center for Research Resources General Clinical Research Centers Program grants M01-RR00036 (Washington University School of Medicine), M01-RR00043-45 (Children's Hospital Los Angeles), M01-RR00069 (University of Colorado Denver), M01-RR00084 (Children's Hospital of Pittsburgh), M01-RR01066 (Massachusetts General Hospital), M01-RR00125 (Yale University), and M01-RR14467 (University of Oklahoma Health Sciences Center); and the National Center for Research Resources Clinical and Translational Science Awards grants UL1-RR024134 (Children's Hospital of Philadelphia), UL1-RR024139 (Yale University), UL1-RR024153 (Children's Hospital of Pittsburgh), UL1-RR024989 (Case Western Reserve University), UL1-RR024992 (Washington University in St. Louis), UL1-RR025758 (Massachusetts General Hospital), and UL1-RR025780 (University of Colorado Denver). SEARCH for Diabetes in Youth (SEARCH 3) is funded by the Centers for Disease Control and Prevention (PA numbers 00097, DP-05-069, and DP-10-001) and supported by the NIDDK. The SEARCH for Diabetes in Youth Cohort Study (1UC4DK108173) (SEARCH 4) is funded by the NIH/NIDDK and supported by the Centers for Disease Control and Prevention. The Population Based Registry of Diabetes in Youth Study (1U18DP006131, U18DP006133, U18DP006134, U18DP006136, U18DP006138, and U18DP006139) is funded by the Centers for Disease Control and Prevention and supported by the NIH/NIDDK. The sites for SEARCH 1–4 and funding are as follows: Kaiser Permanente Southern California (U18DP006133, U48/CCU919219, U01 DP000246, and U18DP002714), University of Colorado Denver (U18DP006139, U48/CCU819241-3, U01 DP000247, and U18DP000247-06A1), Cincinnati Children's Hospital Medical Center (U18DP006134, U48/CCU519239, U01 DP000248, and 1U18DP002709), University of North Carolina at Chapel Hill (U18DP006138, U48/CCU419249, U01 DP000254, and U18DP002708), Seattle Children's Hospital (U18DP006136, U58/CCU019235-4, U01 DP000244, and U18DP002710-01), and Wake Forest University School of Medicine (U18DP006131, U48/CCU919219, U01 DP000250, and 200-2010-35171). Sequencing for T2D-GENES cohorts was funded by NIDDK grant U01DK085526 (Multiethnic Study of Type 2 Diabetes Genes) and National Human Genome Research Institute grant U54HG003067 (Large Scale Sequencing and Analysis of Genomes). Sequencing for ProDiGY cohorts was funded by NIDDK grant U01DK085526. Analysis was supported by NIDDK grant U01DK105554 (AMP T2D-GENES Data Coordination Center and Web Portal). The Mount Sinai Institute for Personalized Medicine Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies. The research from the Korean cohort was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grants HI14C0060 and HI15C1595). The Framingham Heart Study is funded by NIH contracts N01-HC-25195 and HHSN2682015000011 and NIH grants U01 DK085526 and R01 DK078616. Additional support for the Framingham Heart Study is provided by the National Heart, Lung, and Blood Institute (NHLBI) (N01-HC-25195 and grant R01 NS17950) and National Institute on Aging (AG08122 and AG033193). This research was conducted in part using data and resources from the Framingham Heart Study of the NHLBI of the NIH and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource project. This work was partially supported by the NHLBI's Framingham Heart Study (contract number N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (contract number N02-HL-6-4278). This research was partially supported by NIDDK grant 1R01DK8925601. A portion of this research used the Linux Cluster for Genetic Analysis. The Mexico City Diabetes Study has been supported by the following grants: R01 HL24799 from the NHLBI; Consejo Nacional de Ciencia y Tecnología 2092, M9303, F677-M9407, 251M, 2005-C01-14502, and SALUD 2010-2151165;

and Consejo Nacional de Ciencia y Tecnología (Fondo de Cooperación Internacional en Ciencia y Tecnología C0012-2014-01-247974). The Diabetes in Mexico Study was supported by Consejo Nacional de Ciencia y Tecnología grant S008-2014-1-233970 and by Instituto Carlos Slim de la Salud, A.C.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the NIDDK. This study includes data provided by the Ohio Department of Health, which should not be considered an endorsement of this study or its conclusions. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

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

Author Contributions. S.S. assisted in analysis of results, interpretation of data, wrote the first draft of the manuscript, and edited and approved the manuscript. L.C. and J.M.M. conducted the analysis and edited and approved the manuscript. J.T., J.D., S.G., S.C., R.A.G.-K., M.M.K., R.S., M.H.B., L.E.W., A.M., J.F., G.I., J.M.M., and D.D. assisted in forming the analyses plan, interpretation of data, and edited and approved the manuscript. J.C.F. supervised analyses and interpretation of data and edited and approved the manuscript. J.C.F. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. This study were presented in oral form at the 78th Scientific Sessions of the American Diabetes Association, Orlando, FL, 22–26 June 2018 and the Annual Meeting of the Pediatric Endocrine Society, Baltimore, MD, 24 April–1 May 2019.

References

- International Diabetes Federation. *IDF Diabetes Atlas*, 8th edition, 2017. Accessed 10 November 2018. Available from https://diabetesatlas.org/upload/resources/previous/files/8/IDF_DA_8e-EN-final.pdf
- Ogden CL, Carroll MD, Lawman HG, et al. Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *JAMA* 2016;315:2292–2299
- Mayer-Davis EJ, Dabelea D, Lawrence JM. Incidence trends of type 1 and type 2 diabetes among youths, 2002-2012. *N Engl J Med* 2017;377:301
- Constantino MI, Molyneaux L, Limacher-Gisler F, et al. Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. *Diabetes Care* 2013;36:3863–3869
- Narasimhan S, Weinstock RS. Youth-onset type 2 diabetes mellitus: lessons learned from the TODAY study. *Mayo Clin Proc* 2014;89:806–816
- Dabelea D, Stafford JM, Mayer-Davis EJ, et al.; SEARCH for Diabetes in Youth Research Group. Association of type 1 diabetes vs type 2 diabetes diagnosed during childhood and adolescence with complications during teenage years and young adulthood. *JAMA* 2017;317:825–835
- Zeitler P, Hirst K, Pyle L, et al.; TODAY Study Group. A clinical trial to maintain glycemic control in youth with type 2 diabetes. *N Engl J Med* 2012;366:2247–2256
- Langenberg C, Lotta LA. Genomic insights into the causes of type 2 diabetes. *Lancet* 2018;391:2463–2474
- Barnett AH, Eff C, Leslie RD, Pyke DA. Diabetes in identical twins. A study of 200 pairs. *Diabetologia* 1981;20:87–93
- Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature* 2016;536:41–47
- Giannini C, Dalla Man C, Groop L, et al. Co-occurrence of risk alleles in or near genes modulating insulin secretion predisposes obese youth to prediabetes. *Diabetes Care* 2014;37:475–482
- Dabelea D, Dolan LM, D'Agostino R Jr., et al. Association testing of TCF7L2 polymorphisms with type 2 diabetes in multi-ethnic youth. *Diabetologia* 2011;54:535–539
- Flannick J, Mercader JM, Fuchsberger C, et al.; Broad Genomics Platform; DiscovEHR Collaboration; CHARGE; LuCamp; ProDiGY; GoT2D; ESP; SIGMA-T2D; T2D-GENES; AMP-T2D-GENES. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019;570:71–76
- Zeitler P, Epstein L, Grey M, et al.; TODAY Study Group. Treatment options for type 2 diabetes in adolescents and youth: a study of the comparative efficacy of metformin alone or in combination with rosiglitazone or lifestyle intervention in adolescents with type 2 diabetes. *Pediatr Diabetes* 2007;8:74–87
- SEARCH Study Group. SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. *Control Clin Trials* 2004;25:458–471
- Dabelea D, Mayer-Davis EJ, Lamichhane AP, et al. Association of intrauterine exposure to maternal diabetes and obesity with type 2 diabetes in youth: the SEARCH Case-Control Study. *Diabetes Care* 2008;31:1422–1426
- Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284–1287
- Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;81:208–227
- AMP-T2D Program, T2D-GENES Consortium. 2017. Accessed 1 December 2020. Available from <https://t2d.hugeamp.org/variant.html?variant=rs2604566>
- Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–1513
- Guindo-Martinez M, Amela R, Bonàs-Guarch S, et al. The impact of non-additive genetic associations on age-related complex disease. *bioRxiv*. 14 May 2020 [preprint]. DOI: 10.1101/2020.05.12.084608
- Lee KH, Ju UI, Song JY, Chun YS. The histone demethylase PHF2 promotes fat cell differentiation as an epigenetic activator of both C/EBP α and C/EBP δ . *Mol Cells* 2014;37:734–741
- Okuno Y, Ohtake F, Igarashi K, et al. Epigenetic regulation of adipogenesis by PHF2 histone demethylase. *Diabetes* 2013;62:1426–1434
- Bricambert J, Alves-Guerra MC, Esteves P, et al. The histone demethylase Phf2 acts as a molecular checkpoint to prevent NAFLD progression during obesity. *Nat Commun* 2018;9:2092
- Williams AL, Jacobs SB, Moreno-Macías H, et al.; SIGMA Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014;506:97–101
- AMP-T2D Program, T2D-GENES Consortium. 2017. Accessed 2 December 2020. Available from <https://t2d.hugeamp.org/variant.html?variant=rs72982988>
- Chen HF, Hsu CM, Huang YS. CPEB2-dependent translation of long 3'-UTR Ucp1 mRNA promotes thermogenesis in brown adipose tissue. *EMBO J* 2018;37:37