Whole Exome Sequencing Identifies Variation in *CYB5A* and *RNF10* Associated with Adiposity and Type 2 Diabetes

Ke Huang[1], **Anup K. Nair**[1], **Yunhua Li Muller**[1], **Paolo Piaggi**[1], **Li Bian**[1], **Melissa del Rosario**[1], **William C. Knowler**[1], **Sayuko Kobes**[1], **Robert L. Hanson**[1], **Clifton Bogardus**[1], and **Leslie J. Baier**[1]

[1]Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 445 North 5th Street Phoenix, AZ 85004 USA

Abstract

Objective—Few coding variants in genes associated with type 2 diabetes (T2D) have been identified, and the underlying physiologic mechanisms whereby susceptibility genes influence T2D risk are often unknown. The objective of this study was to identify coding variation that increases risk for T2D via an effect on a pre-diabetic trait.

Design and Methods—Whole exome sequencing was done in 177 Pima Indians. Selected variants (N=345) were genotyped in 555 subjects characterized for body fatness, glucose disposal rates during a clamp, acute insulin response to glucose, and 2-hour plasma glucose concentrations during an OGTT, and were also genotyped in up to 5,880 subjects with longitudinal measures of BMI. Variants associated with quantitative traits were assessed for association with T2D in 7,667 subjects.

Results—rs7238987 in *CYB5A* associated with body fatness ($p=7.0\times10^{-6}$). This SNP and a novel SNP in *RNF10* also associated with maximum recorded BMI ($p=6.2\times10^{-7}$ and $p=7.2\times10^{-4}$) and maximum childhood BMI z-score ($p=5.9\times10^{-4}$ and $p=8.5\times10^{-7}$). The BMI increasing alleles increased risk for T2D ($p= 0.01$; OR=1.13 [1.03–1.24] and 9.5×10^{-3}, OR=1.49 [1.10–2.02]).

Conclusions—*CYB5A,* which has a role in stearyl-CoA-desaturase activity, and *RNF10*, with an unknown role in weight regulating pathways, associated with adiposity and nominally increased risk for T2D in American Indians.

Keywords

BMI; Exome sequencing; *RNF10*; *CYB5A*; *SCD1*

Introduction

T2D is heritable, and prior genome-wide association studies have identified common, non-coding variants associated with this disease (1). However, the actual functional variant and the affected metabolic pathway that leads to T2D are often unknown. Body fatness, insulin secretory dysfunction and insulin resistance predict T2D, and their relative hazard ratios (95% CI) are 1.80 (1.29–2.51), 2.33 (1.49–3.70) and 1.64 (1.25–2.00) in Pima Indians (2). To identify potentially functional variants that affect metabolic traits known to increase risk for T2D, we used whole exome sequencing to detect coding variation that associates with a pre-diabetic trait, and then assessed whether the variant is associated with T2D.

Research Design and methods

Study participants and phenotypes

Subjects for Sample 1 genotyping (N=555; 92% were full heritage Pima Indian) were non-diabetic inpatients metabolically characterized in our Clinical Research Center (Table S1). A subset of Sample 1 (N=177 full heritage non-diabetic Pima Indians from different nuclear families) was used for exome sequencing, but all follow-up data was derived from re-genotyped variants in the entire Sample 1. Percent body fatness (PFAT) was measured by underwater weighing (n=305) or total body dual energy X-ray absorptiometry (n=250; DPX-L Lunar Radiation, Madison, WI, USA); measurements using the two methods were made comparable using a previously derived equation (3). Plasma glucose levels (2-hr glucose) were measured at 120 minutes following ingestion of 75g glucose. Insulin action (M) was assessed at physiological insulin concentrations during a hyperinsulinemic–euglycemic clamp with simultaneous glucose tracers (4).

Measurements derived from the clamp were normalized to estimated metabolic body size (EMBS = fat-free mass + 17.7 kg), as previously derived (5). Acute insulin response (AIR) was measured by collecting blood samples before a 25g intravenous glucose infusion over 3 minutes and at 3, 4, 5, 6, 8 and 10 minutes following infusion. AIR was calculated as the mean increment in plasma insulin concentrations from 3 to 5 minutes (4).

Samples 2 and 3 are outpatients from a longitudinal study of the Gila River Indian Community, where community members at age 5 years were asked to come for biennial health examinations which included measuring height and weight to calculate z-scores and BMI, and a 75 g oral glucose tolerance test to determine diabetes status (6). Sample 2 included those individuals whose heritage was full Pima and/or Tohono O'odham (a closely related tribe) ("Full heritage Pima Indian"; N=3,604; Table S2). Sample 3 included all remaining individuals from this study whose heritage was, on average, one half Pima Indian and two-thirds American Indian ("Mixed heritage" American Indian; N=4,063; Table S3). Maximum recorded BMI is defined as the highest BMI calculated from a longitudinal exam at which the subject was non-diabetic and 15 years of age, and maximum childhood z-score is defined as maximum age and sex adjusted z-score before the age of 20 years. Since some individuals were not examined after the age of 20 years, while others developed T2D before the age of 20 years, there were subjects from Samples 2 and 3 (n= 395 and 856, respectively) that had their highest childhood age and sex adjusted z-score and their highest

recorded (unadjusted) non-diabetic BMI calculated from data collected at same longitudinal exam (between the ages of 15–20 years)." Subjects with no measure of BMI prior to developing T2D were excluded from BMI analyses, but included in T2D analyses.

Whole exome sequencing and variant calling

Exon capture (utilized Agilent 38Mb kit) and sequencing was done by Shanghai Bio (North Brunswick, New Jersey, USA). The average read depth of the sequence data was 42X. The methods used for alignment, genotype calling and annotation have been previously published (7). For quality control, only variants with quality score ≥ 50, genotype call rate $\geq 75\%$ and in Hardy Weinberg Equilibrium ($p > 1.0 \times 10^{-4}$) were included.

SNP selection and genotyping

Figure S1 outlines the criteria for selection of coding SNPs for follow-up genotyping (N=345). The SIFT program (8) was used to identify damaging SNPs and the HuGE Navigator (http://www.hugenavigator.net) and IPA programs (www.ingenuity.com) were used to select established or plausible T2D or obesity genes. Variants were genotyped using Illumina BeadXpress System (Illumina, San Diego, CA), or Taqman Open Array (Applied Biosystems, Carlsbad, CA). Quality control for genotyping required a successful call rate of >90% of all samples, lack of deviation from Hardy-Weinberg equilibrium ($p > 1.0 \times 10^{-3}$), and a discrepancy rate of <2.5% for blind duplicate samples (50, 280 and 100 duplicate samples in Samples 1–3, respectively).

Statistical analysis

Statistical analyses were performed with SAS (SAS Institute, Cary, NC). Linear regression models assessed the association between the continuous pre-diabetic traits and genotypes (assuming an additive model) with adjustment for covariates including age and sex (covariates specific for individual traits are reported in the footnote of Table 1). In addition, generalized estimating equations were used to account for sibships and individual estimates of European admixture were used to adjust for ethnicity. These estimates were derived from 45 markers with large differences in allele frequency between populations (9) using the analytical method described in Hanis *et al* (10).

Results and Discussion

Variants identified by whole exome sequencing are summarized in Table S4, and their follow-up, as described below, is depicted in Figure S1. A preliminary association analyses between all 31,441 coding variants and PFAT, M, 2-hr glucose, and AIR in the 177 subjects who were sequenced did not identify an association that achieved exome-wide significance (31,441 SNPs analyzed for 4 traits; which conservatively assumes that all traits are independent, requires a $p < 3.9 \times 10^{-7}$), which was not unexpected unless Pima Indians harbored a coding variant with a substantial effect size on a single quantitative trait. Therefore, 25 SNPs with the strongest "trend" for association with a trait (*p* values ranged from 8×10^{-6}–9×10^{-4}) were re-genotyped in the entire Sample 1 (*n*= 555) to increase the power. In addition, 320 SNPs identified by whole exome sequencing which were either novel and predicted to be damaging, or were missense SNPs in biologic candidate genes

were also genotyped Sample 1. The power to identify variants with modest effect sizes in Sample 1 is still low; we estimate this sample (which includes all available DNA samples with inpatient measures of PFAT, M, 2-hr glucose, and AIR) has 48% power to identify a SNP that explains 1% of the variance. Therefore, the 345 SNPs genotyped in Sample 1 were simultaneously genotyped in Sample 2 (n=2,842; Table S5) and 20 SNPs were further genotyped in Sample 3 (n=3,038; Table S6). Individuals in Samples 2 and 3 have not undergone detailed metabolic phenotyping and therefore are only informative for analysis of BMI (maximum BMI and childhood z-score) as a pre-diabetic trait. For a SNP that explains 1% of the variance in BMI, we estimate 94% power for Sample 2+3 (n=5,880) at $p<10^{-7}$. Samples 2 and 3 are also informative for T2D status.

Table 1 summarizes SNPs with the lowest p values for association with any trait in Samples 1–3; however, only variation in *CYB5A* and *RNF10* had associations that approached genome-wide significance. rs7238987 in *CYB5A* had one of the lowest p values for association with a metabolic trait in the 177 samples that were exome sequenced (p value for association with PFAT=1.6×10^{-5}, data not shown) and the association with PFAT became stronger when the sample size was increased to include all subjects in Sample 1 (n=555 subjects, p=7.0×10^{-6}, Table 1, Figure 1A). Among the larger number of subjects in Samples 2 and 3 with longitudinal data for BMI, this SNP associated with maximum childhood z-score and maximum recorded BMI (p= 5.9×10^{-4} and 6.2×10^{-7} and, Table 1; Figure 1B–C), and individuals with the risk allele for adiposity were at increased risk for T2D (OR= 1.13 [1.03–1.24]; p=0.01, Figure 1D. If the association for T2D was additionally adjusted for BMI, the association became non-significant (p=0.08) suggesting an adiposity mediated effect on risk for T2D. Although rs7238987 in *CYB5A* encodes a P96P synonymous variant, the SIFT program (8) predicts the C to T nucleotide substitution to affect mRNA splicing between exon 3 and intron 3. However, sequencing across exons 1–6 in cDNA from subcutaneous adipose and skeletal muscle biopsies from 65 subjects did not identify differential splicing patterns based on genotype in these two tissues (data not shown). The risk allele (T) for rs7238987 is more common in Pima Indians (frequency 32%) as compared to other ethnic groups that do not suffer from such high rates of obesity (e.g. among Caucasians, Asians and Africans, the T allele frequency is 0.08–0.17); however, this SNP was not associated with BMI or other diabetes related traits in Caucasians from the GIANT and MAGIC consortia (11, 12). In Pima Indians rs7238987 is in perfect linkage disequilibrium with another synonymous Y12Y (not predicted to be functional) and 103 non-coding SNPs across this locus, suggesting that non-coding variation may be giving rise to this association.

Cytochrome b5 type A, encoded by *CYB5A*, is a membrane bound microsomal hemoprotein that acts as an electron carrier for the stearoyl–CoA-desaturase (SCD) complex that facilitates the conversion of saturated fatty acid to mono unsaturated fatty acid (13). SCD, also known as delta 9 desaturase, is a key enzyme in fatty acid and energy metabolism. A prior study in 52 Pima Indians reported that SCD activity is associated with obesity but not insulin action (14). Additional studies have also confirmed that increased SCD activity is associated with obesity and obesity-related diseases (15).

The SNP in *RNF10* (R151H; minor allele frequency 0.03) is novel and predicted to be damaging (Table 1). Sample 1 had insufficient power to detect an association between this low frequency SNP and PFAT (Fig 2A); however, evidence for association with maximum childhood z-score and maximum recorded BMI was observed in the larger Samples 2 and 3 ($p=8.5\times10^{-7}$ and 7.2×10^{-4} and, Table 1; Figure 2B–C), and the allele for higher BMI predicted increased risk for T2D (OR=1.49 [1.10–2.02]; 9.5×10^{-3}; Figure 2D. Adjusting for BMI renders the T2D association non-significant ($p=0.06$), suggesting an adiposity mediated effect on risk for T2D. The R151H has not been reported in any publically available databases, and in our studies has only been detected in American Indians with some degree of Pima heritage. *RNF10* encodes a transcription factor which has not been implicated in a pathway known to affect obesity. However, the Mouse Genome Informatics program lists *rnf10* conditional knockout mice as having increased total body fat, increased body weight and abnormal eating behavior (16), supporting our association data that this gene has a role in body weight regulation.

In conclusion, exome sequencing and follow-up genotyping identified *CYB5A* and *RNF10* as potential new loci for adiposity, where individuals carrying alleles for higher adiposity were at increased risk for T2D. However, given the lack of reproducibility of the *CYB5A* association in Caucasians, and the low frequency of the novel *RNF10* variant, further confirmation and functional studies are necessary to validate these findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. McCarthy MI. Genomics, type 2 diabetes, and obesity. N Engl J Med. 2010; 363:2339–50. [PubMed: 21142536]

2. Bunt JC, Krakoff J, Ortega E, Knowler WC, Bogardus C. Acute insulin response is an independent predictor of type 2 diabetes mellitus in individuals with both normal fasting and 2-h plasma glucose concentrations. Diabetes Metab Res Rev. 2007; 23:304–310. [PubMed: 17001619]

3. Tataranni PA, Ravussin E. Use of dual-energy X-ray absorptiometry in obese individuals. Am J Clin Nutr. 1995; 62:730–4. [PubMed: 7572700]

4. Lillioja S, Mott DM, Spraul M, Ferraro R, Foley JE, Ravussin E, et al. Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. N Engl J Med. 1993; 329:1988–92. [PubMed: 8247074]

5. Lillioja S, Bogardus C. Obesity and insulin resistance: lessons learned from the Pima Indians. Diabetes Metab Rev. 1988 Aug.4:517–40. [PubMed: 3061759]

6. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. Am J Epidemiol. 1978; 108:497–505. [PubMed: 736028]

7. Huang K, Yellspantula V, Baier L, Dinu V. NGSPE: A pipeline for end-to-end analysis of DNA sequencing data and comparison between different platforms. Comput Biol Med. 2013; 43:1171–76. [PubMed: 23930810]

8. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073–81. [PubMed: 19561590]

9. Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, et al. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am J Hum Genet. 2007; 80:1014–23. [PubMed: 17557415]

10. Hanis CL, Chakraborty R, Ferrell RE, Schull WJ. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. Am J Phys Anthropol. 1986; 70:433–41. [PubMed: 3766713]

11. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010 Nov.42:937–48. [PubMed: 20935630]

12. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010; 42:105–16. [PubMed: 20081858]

13. Paton CM, Ntambi JM. Biochemical and physiological function of stearoyl-CoA desaturase. Am J Physiol Endocrinol Metab. 2009; 297:E28–E37. [PubMed: 19066317]

14. Pan DA, Lillioja S, Milner MR, Kriketos AD, Baur LA, Bogardus C, et al. Skeletal muscle membrane lipid composition is related to adiposity and insulin action. J Clin Invest. 1995; 96:2802–08. [PubMed: 8675650]

15. Ntambi JM, Miyazaki M, Stoehr JP, Lan H, Kendziorski CM, Yandell BS, et al. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. Proc Natl Acad Sci U S A. 2002; 99:11482–6. [PubMed: 12177411]

16. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resurse for the genome-wid study of mouse function. Nature. 2011; 474:337–42. [PubMed: 21677750]

What is already known about this subject?

1. T2D is a complex disease for which a genetic contribution is well accepted.

2. Whole exome sequencing has emerged as a powerful technology enabling detection of coding variation not captured by initial GWAS design.

3. Body fatness, insulin secretory dysfunction and insulin resistance are predictive factors for T2D in Pima Indian.
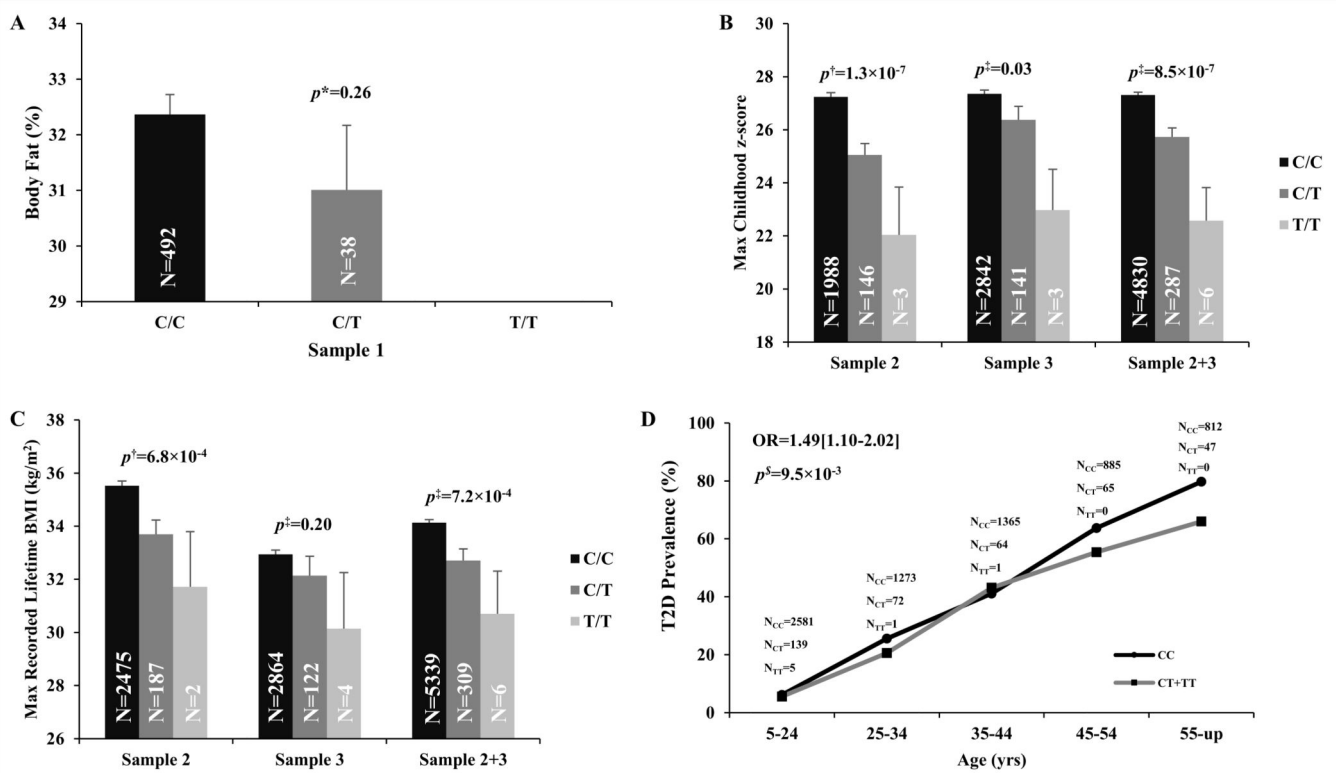
What does this study add?

1. We used whole exome sequencing to detect potentially functional coding variation that increases risk for T2D in American Indians.

2. *CYB5A* and *RNF10* were identified as potential new loci for adiposity and T2D in American Indians.

3. The variant associated with adiposity and T2D in *RNF10* is potentially unique to the Pima Indian tribe, and underscores the importance of identifying susceptibility genes in minority populations that suffer from disproportionately high rates of obesity and T2D.
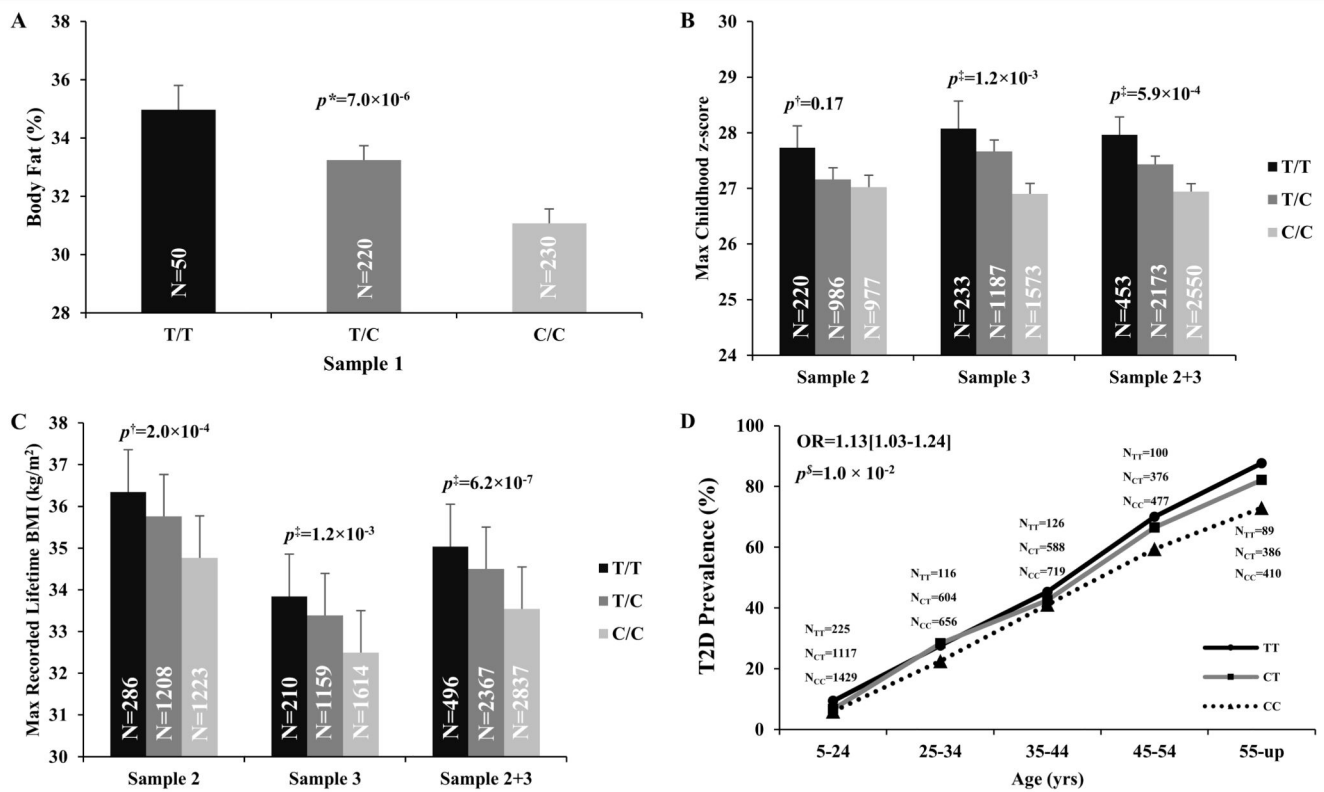
Figure 1. rs7238987 in CYB5A is associated with PFAT (A), maximum childhood z-score (B), and maximum recorded BMI from a longitudinal exam when the subject was non-diabetic and ≥ 15 years of age (C) with the risk allele modestly increasing the risk for T2D (D)

PFAT: Percent body fat, maximum childhood z score is the highest age and sex adjusted z-score from an exam at age <20yrs. For presentation, the z-scores were sex and age (female, 12yr) standardized to a BMI scale. Maximum recorded BMI is defined as the highest BMI recorded from a longitudinal exam when the subject was non-diabetic and ≥ 15 years of age.* *p* values were adjusted for age, sex and nuclear family membership. $ *p* values were adjusted for age, sex, birth year, nuclear family membership and admixture estimates. ‡ *p* values were adjusted for birth year, nuclear family membership and admixture estimates.

Figure 2. R151H, a novel low frequency variant in *RNF10* has a non-significant trend for association with PFAT (A) and is associated with maximum childhood z-score (B), and maximum recorded BMI from a longitudinal exam when the subject was non-diabetic and 15 years of age (C) with the risk allele modestly increasing the risk for T2D (D)

PFAT: Percent body fat, maximum childhood z score is the highest age and sex adjusted z-score from an exam at age <20yrs. For presentation, the z-scores were sex and age (female, 12yr) standardized to a BMI scale. Maximum recorded BMI is defined as the highest BMI recorded from a longitudinal exam when the subject was non-diabetic and 15 years of age.* p values were adjusted for age, sex and nuclear family membership. $ p values were adjusted for age, sex, birth year, nuclear family membership and admixture estimates.‡ p values were adjusted for birth year, nuclear family membership and admixture estimates.

Table 1

SNPs with the lowest *p* values for association with a pre-diabetic trait.

Chr:SNP	Gene	R/NR allele	AA change	fR	beta	*p* value
Maximum recorded BMI (N=5,880, Sample 2+3) [*][$]						
18:rs7238987	*CYB5A*	T/C	P96P	0.30	0.024	6.2×10^{-7}
1:rs913257	*GORAB*	A/G	K320E	0.87	0.022	2.7×10^{-4}
12:120990399(novel)	*RNF10*	C/T	R151H	0.97	0.044	7.2×10^{-4}
Maximum childhood z-score (N=5,316, Sample 2+3) [§]						
12:120990399(novel)	*RNF10*	C/T	R151H	0.97	0.270	8.5×10^{-7}
18:rs7238987	*CYB5A*	T/C	P96P	0.30	0.082	5.9×10^{-4}
PFAT (N=555, Sample 1) [#]						
18:rs7238987	*CYB5A*	T/C	P96P	0.32	2.039	7.0×10^{-6}
2:rs6746030	*SCN9A*	A/G	W1150R	0.11	2.380	2.1×10^{-4}
5:rs62624460	*PCDHA8*	G/C	L64F	0.92	2.662	2.3×10^{-4}
11:rs1064608	*MTCH2*	G/C	A290P	0.5	1.453	3.1×10^{-4}
Log$_{10}$M (N=555, Sample 1) [£]						
18:rs2282632	*ASXL3*	A/G	N954S	0.12	−0.036	2.0×10^{-5}
2:rs10804166	*C2orf80*	A/G	S152G	0.15	−0.027	8.9×10^{-4}
2-hr glucose (N=555, Sample 1) [£]						
1:rs2890565	*UTS2*	C/T	S98N	0.55	6.698	7.0×10^{-5}
11:118244312(novel)	*UBE4A*	G/A	P343L	0.96	17.154	2.7×10^{-4}
19:rs16978738	*ZNF225*	T/A	S679T	0.61	7.053	3.2×10^{-4}
5:rs1432862	*FAT2*	A/G	C574R	0.27	6.851	9.5×10^{-4}
Log$_{10}$AIR (N=297, Sample 1) [¥]						

Chr:SNP	Gene	R/NR allele	AA change	fR	beta	p value
2:rs4674941	DOCK10	C/G	I1251M	0.93	-0.144	2.8×10^{-4}
7:rs2070607	OGDH	G/A	V1018I	0.92	-0.093	3.1×10^{-4}
1:rs3176443	FAM5B	C/G	L390V	0.07	-0.140	5.0×10^{-4}
17:rs12453124	KRT27	C/T	E144K	0.43	-0.072	6.4×10^{-4}
Log_{10} Disposition Index (N=297, Sample 1)[¥]						
7:rs2070607	OGDH	G/A	V1018I	0.92	0.100	2.0×10^{-4}
2:rs4674941	DOCK10	C/G	I1251M	0.93	0.156	4.5×10^{-4}
1:rs3176443	FAM5B	C/G	L390V	0.07	-0.141	5.0×10^{-4}

R: risk allele, NR: non risk allele, fR: frequency of risk allele, BMI: body mass index, PFAT: percentage body fat, Log10M: logarithmic value of glucose disposal rate during insulin infusion, 2-hr glucose: 2-hour plasma glucose concentrations in response to a 75g OGTT, Log10AIR: Logarithmic value of acute insulin response to a 25g intravenous glucose bolus.

[*] p values were adjusted for age, sex, birth year, nuclear family membership and estimate of admixture;

[$] Maximum recorded BMI is defined as the highest BMI measured at a longitudinal exam when the subject was non-diabetic and 15 years of age;

[§] maximum childhood z score is the highest age and sex adjusted Z score from an exam at age<20 yr, p values were adjusted for birth year, family membership and admixture.

[#] p values were adjusted for age, sex and nuclear family membership;

[£] p values were adjusted for age, sex, PFAT and nuclear family membership;

[¥] analysis is restricted to full heritage Pima Indians who are normal glucose tolerant and p values were adjusted for age, sex, PFAT, nuclear family membership and Log10M (for AIR only).