Examination of all type 2 diabetes GWAS loci reveals *HHEX-IDE* as a locus influencing pediatric BMI

Jianhua Zhao¹, Jonathan P. Bradfield², Haitao Zhang², Kiran Annaiah², Kai Wang², Cecilia E. Kim², Joseph T. Glessner², Edward C. Frackelton², F. George Otieno², James Doran², Kelly A. Thomas², Maria Garris², Cuiping Hou², Rosetta M. Chiavacci², Mingyao Li³, Robert I. Berkowitz⁴,⁵, Hakon Hakonarson¹,²,⁶* and Struan F.A. Grant¹,²,⁶*

¹Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; ²Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; ³Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; ⁴Behavioral Health Center and Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia PA 19104, USA; ⁵Center for Weight and Eating Disorders, Department of Psychiatry, University of Pennsylvania, Philadelphia PA 19104, USA; ⁶Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

*To whom correspondence should be addressed.

Struan F.A. Grant
E-mail: grants@chop.edu

Submitted 2 July 2009 and accepted 13 November 2009.

Additional information for this article can be found in an online appendix at [http://diabetes.diabetesjournals.org](http://diabetes.diabetesjournals.org)

This is an uncopyedited electronic version of an article accepted for publication in *Diabetes*. The American Diabetes Association, publisher of *Diabetes*, is not responsible for any errors or omissions in this version of the manuscript or any version derived from it by third parties. The definitive publisher-authenticated version will be available in a future issue of *Diabetes* in print and online at [http://diabetes.diabetesjournals.org](http://diabetes.diabetesjournals.org).
Objective: A number of studies have found that body mass index (BMI) in early life influences the risk of developing type 2 diabetes (T2D) later in life. Our goal was to investigate if any T2D variants uncovered through genome wide association studies (GWAS) impact BMI in childhood.

Design and Methods: Utilizing data from an ongoing GWAS of pediatric BMI in our cohort, we investigated the association of pediatric BMI with 20 SNPs at 18 T2D loci uncovered through GWAS, consisting of ADAMTS9, CDC123-CAMK1D, CDKAL1, CDKN2A/B, EXT2, FTO, HHEX-IDE, IGF2BP2, the intragenic region on 11p12, JAZF1, KCNQ1, LOC387761, MTNR1B, NOTCH2, SLC30A8, TCF7L2, THADA and TSPAN8-LGR5. We randomly partitioned our cohort exactly in half in order to have a ‘discovery’ cohort (n=3592) and a ‘replication’ cohort (n=3592).

Results: Our data show that the major, T2D-risk conferring G allele of rs7923837 at the HHEX-IDE locus was associated with higher pediatric BMI in both the discovery ($P=0.0013$; and survived correction for 20 tests) and replication ($P=0.023$) sets (combined $P=1.01\times10^{-4}$). Association was not detected with any other known type 2 diabetes loci uncovered to date through GWAS except for the well established FTO.

Conclusions: Our data show that the same genetic HHEX-IDE variant which is associated with type 2 diabetes from previous studies also influences pediatric BMI.
Diabetes mellitus affects an estimated 194 million adults worldwide and more than 18 million in the United States, with the chronic complications including microvascular disease and accelerated development of cardiovascular disease. Approximately 90 to 95 percent of those affected with diabetes have the type 2 diabetes (T2D) form of the disease. Hyperglycemia is a key feature of T2D and occurs through two possible mechanisms: a) abnormal insulin secretion due to pancreatic β-cell defects or b) insulin resistance in skeletal, muscle, liver and adipose tissue.

T2D has been the focus of more genome wide association studies (GWAS) than any other disorder studied to date; such analyses have revealed a number of loci (1-9). The strongest association in European populations has been with a gene established in 2006, namely the Wnt-signaling pathway member transcription factor 7-like 2 (TCF7L2) (10), while in China and Japan the strongest association has been to the gene encoding potassium channel, voltage-gated, KQT-like subfamily, member 1 (KCNQ1) (8; 9). The first batch of such studies (1-6), revealed new loci and with a recent meta-analysis (7) of T2D genome wide SNP genotype data producing another six loci, there are now 17 genes established in the disease, including ADAMTS9, CDC123-CAMK1D, CDKAL1, CDKN2A/B, EXT2, FTO, HHEX-IDE, IGF2BP2, the intragenic region on 11p12, JAZF1, KCNQ1, LOC387761, MTNR1B, NOTCH2, SLC30A8, TCF7L2, THADA and TSPAN8-LGR5 with respect to their correlation with pediatric BMI.

MATERIAL AND METHODS

Research Subjects: Childhood European American Cohort from Philadelphia: All subjects were consecutively recruited from the Greater Philadelphia area from 2006 to 2009 at the Children's Hospital of Philadelphia. Our study cohort consisted of 7,184 singleton children of European ancestry with systematically recorded height and weight. All subjects were consecutively and randomly recruited from the greater metropolitan area of Philadelphia from 2006 to 2009 at the Children’s Hospital of Philadelphia i.e. participants were not specifically targeted for obesity-related traits. The study was approved by the Institutional Review Board of the Children's Hospital of
Philadelphia. Parental informed consent was given for each study participant for both the blood collection and subsequent genotyping.

Genotyping: Illumina Infinium™ assay: We performed high throughput genome-wide SNP genotyping, using the Illumina Infinium™ II HumanHap550 or Human 610 BeadChip technology(Illumina, San Diego), at the Children’s Hospital of Philadelphia’s Center for Applied Genomics, as described previously (15). The overall genomic control value was 1.036. The SNPs analyzed survived the filtering of the genome wide dataset for SNPs with call rates <95%, minor allele frequency <1%, missing rate per person >2% and Hardy-Weinberg equilibrium $P < 10^{-5}$.

Most loci described from GWA studies published to date have been found using either the Affymetrix or Illumina platform. In the event a locus was reported using both the Illumina and Affymetrix arrays, we used the SNPs present on the Illumina array. In the event of a signal only being described on the Affymetrix array, we either already had the SNP on our Illumina array or we identified and used the best surrogate SNP available based on the CEU HapMap (Supplementary Table 1 which can be found in an online appendix at http://diabetes.diabetesjournals.org). We utilized two SNPs at the $CDKAL1$ (rs4712523 and rs7756992; $r^2 = 0.677$) and $HHEX$-IDE (rs1111875 and rs7923837; $r^2 = 0.698$) loci as the association with T2D from various GWA studies reported different SNPs which were in imperfect LD with each other. rs3751812 at $FTO$ was included as a positive control as we have previously reported association with this SNP and both pediatric obesity and pediatric BMI previously(16; 17).

Analysis: Normalization of BMI: BMI percentiles were defined using the standard Center for Disease control (CDC) growth chart $z$-scores that take in to account age and gender. All subjects were biologically unrelated and were aged between 2 and 18 years old. All subjects were between -3 and +3 standard deviations of CDC corrected BMI i.e. outliers ($n=356$) were excluded to avoid the consequences of potential measurement error or Mendelian causes of extreme obesity.

Association: We queried the data for the SNPs of interest in our pediatric sample. All statistical analyses were carried out using the software package PLINK version 1.05(18). We applied the PLINK to the generation of genome-wide IBS estimates between all subjects and then generated multi-dimensional scaling (MDS) plots for visual examination of population outliers. To help interpret the population genetic analysis, we have included 924 HapMap3 individuals from 11 populations as positive controls into the MDS analysis. The individuals of European ancestry were selected by the principal component 1 of more than 0.04 and principal component 2 of more than 0.01. Comparing self-identified ancestry with the MDS-inferred ancestry confirmed the reliability of MDS to identify genetically inferred individuals of European ancestry.

By treating the normalized BMI $z$-score as a quantitative trait, association analysis for each SNP was carried out using linear regression (additive model) with the SNP included as an independent variable (coded as 0, 1, and 2). With 3592 subjects in the discovery cohort, the powers to detect 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.8% and 1% variation at the $\alpha=0.0025$ level were 27.0%, 49.0%, 68.2%, 82.0%, 90.6%, 97.9% and 99.6%, respectively.

RESULTS

In our analysis, twenty SNPs corresponding to the eighteen T2D loci previously discovered in GWAS of the disorder were investigated, namely $ADAMTS9$, $CDCA123-CAMK1D$, $CDKAL1$, $CDKN2A/B$, $EXT2$, $FTO$, $HHEX$-IDE, $IGF2BP2$, the intragenic region on 11p12,
JAZF1, KCNQ1, LOC387761, MTNR1B, NOTCH2, SLC30A8, TCF7L2, THADA and TSPAN8-LGR5 (Table 1).

We randomly partitioned our cohort exactly in half in order to have a ‘discovery’ cohort (n=3592) and a ‘replication’ cohort (n=3592). Five of these twenty SNPs yielded at least nominally significant association to BMI ($P < 0.05$) in the discovery cohort, representing four different independent loci.

Of these four loci, the minor allele of rs3751812 at the FTO locus yielded the strongest association with $P=3.81 \times 10^{-5}$ and tracked with higher BMI. The direction of effect was also readily replicated in the additional cohort ($P=5.56 \times 10^{-6}$) yielding a combined $P=1.05 \times 10^{-9}$.

The major T2D-conferring G allele of rs7923837 at the HHEX-IDE locus was associated with higher pediatric BMI in both the discovery (unadjusted $P=0.0013$; Bonferroni correction for 20 variants threshold $P \leq 0.0025$) and replication (unadjusted $P=0.023$) sets (combined unadjusted $P=1.01 \times 10^{-4}$). The major C allele of rs1111875 at the same locus was also trending with higher pediatric BMI but did not survive the Bonferroni correction for multiple testing in the discovery cohort.

As for the other two nominally significant loci in the discovery cohort, rs4402960 at IGF2BP2 ($P=0.05$) and rs11257622 at CDC123-CAMK1D ($P=0.024$), they failed to replicate in the additional cohort. Association was not detected at all with any of the other T2D loci uncovered to date through GWAS.

We also analyzed males and females separately but the effect of the G allele rs7923837 at the HHEX-IDE locus on pediatric BMI did not vary by gender (Supplementary Table 2). However, we did look at different age bins and found that the variant was associated with higher pediatric BMI most strongly in the 2-6 years old age bin (Supplementary Table 3). Breaking the ages down further in to individual years, nominal significant association for this HHEX-IDE variant in the same direction was observed at ages 3, 7, 14 and 16 years old (Supplementary Table 4). However we did not observe an overall statistical interaction with age, with the interaction $P$-values for rs1111875 and rs7923837 being 0.2507 and 0.1076 respectively.

**DISCUSSION**

If a genomic variant is well established to be associated with a trait which is the consequence of a defect of recognition of insulin by the body or by a fault in the amount of insulin released for the pancreatic islets, i.e. type 2 diabetes, then if these defects are operating at all in childhood one might expect there to be an impact on BMI in childhood.

With this notion in mind, we queried the existing dataset from our ongoing GWAS of pediatric BMI if any of the T2D loci uncovered in GWAS to date play a role in our trait of interest; it should be noted that PPARG, KCNJ11 and of WFS1 were not included as their discovery with respect to being T2D loci pre-dates GWA studies and thus have already been more extensively investigated. Our data in fact do show that the same genetic HHEX-IDE variant which is significantly associated with T2D from previous studies also influences pediatric BMI. Indeed, the major G allele of rs7923837 at the HHEX-IDE locus was associated with higher pediatric BMI in both the ‘discovery’ and ‘replication’ cohorts, which is the same allele that has been reported to confer risk of T2D. This mirrors very well what is seen with the much more established FTO gene, seen here and in other studies.

SNP rs7923837 yielded the fourth strongest association to T2D in a Canadian/French GWAS carried out on the Illumina HumanHap platform (1). SNPs rs1111875 and rs7923837 yielded the
strongest association at the HHEX-IDE locus but it should be noted that they are far from being in perfect LD with each other ($r^2=0.698$) thus the inclusion of both in the current study; however, despite the lack of complete concordance and the large sample size, we were unable to separate the effects of these SNPs as they cannot be considered to be totally independent signals either.

One hypothesis could be that the fetal genotype for rs7923837 is primarily associated with birth weight, as reduced birth weight is often reported to be associated with increased BMI and T2D later in life. However, this doesn’t appear to be the case as we have already investigated and reported the role of these T2D loci in the context of birth weight in our cohort. Although we have agreed with previous studies that CDKAL1 is a birth weight associated gene, we have not observed such an association with HHEX-IDE(19). Further, although there is no CDC categorization for the under 2 year old age group, following our own normalization we do not observe association between rs7923837 and BMI in this age category (data not shown). The correlation between birth weight and BMI in later childhood in less correlated than in earlier stages, suggesting the HHEX-IDE variant exerts its physiological influence directly rather than as a consequence of a ‘knock-on’ effect from a primary impact on birth weight. However, we do acknowledge, of the age bins studied, the strongest effect was observed in the 2-6 years age bin (effect size(SE)=0.12(+/0.04)) (Supplementary Table 3) but is not the whole story due to the fact that at the individual age level, although more limited in terms of power, the impact continues to be observed in to the mid-teens (Supplementary Table 4).

The assumption in this study is that deficient insulin secretion mediates the effect on childhood BMI but it is also possible that higher childhood BMI results in impaired insulin secretion later in life. There could indeed be pleiotropic associations from multiple independent mechanisms; however we were not able to address this as we do not have insulin secretion/sensitivity measures in our study.

From our analysis, apart from FTO it is clear only one of the loci previously reported from T2D GWAS plays a role in our phenotype of interest i.e. pediatric BMI. While this recently discovered locus unveils a new biomolecular pathway not previously studied in the context of T2D and obesity, it is also important to note this and other genetic associations with childhood obesity explain very little of the genetic risk for the pathogenesis of the trait(17); indeed, an estimate of the explained variance of the HHEX-IDE and FTO loci combined is only 0.98%, suggesting the existence of additional loci whose number and effect size remain mainly unknown. Current knowledge concerning the impact of genetic factors in the determination of pediatric BMI may still be very limited due to both the lack of availability of large pediatric cohorts with GWAS data and methodological difficulties in the analysis of the phenotype that changes with age and depends on many other contributing factors. Once our GWAS is complete, we will have the opportunity to look for other variants in the genome associated with BMI in childhood.

ACKNOWLEDGEMENTS

We would like to thank all participating subjects and families. Elvira Dabaghyan, Hope Thomas, Kisha Harden, Andrew Hill, Kenya Fain, Crystal Johnson-Honesty, Cynthia Drummond, Shanell Harrison and Sarah Wildrick provided expert assistance with genotyping or data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnsson of Raförinn ehf for their extensive software design and contribution. This research was financially
supported by the Children’s Hospital of Philadelphia.

The study is supported in part by a Research Development Award from the Cotswold Foundation (H.H. & S.F.A.G.) and NIH grant 1R01HD056465-01A1.
REFERENCES


2. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661-678, 2007


18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-575, 2007

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>T2D associate allele</th>
<th>BP</th>
<th>DISCOVERY COHORT</th>
<th>REPLICATION COHORT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Effect Size</td>
</tr>
<tr>
<td>1</td>
<td>rs2793831</td>
<td>C</td>
<td>120235944</td>
<td>3592</td>
<td>0.03508</td>
</tr>
<tr>
<td>2</td>
<td>rs7578597</td>
<td>T*</td>
<td>43644474</td>
<td>3592</td>
<td>0.01896</td>
</tr>
<tr>
<td>3</td>
<td>rs4411878</td>
<td>C*</td>
<td>64678705</td>
<td>3591</td>
<td>-0.02394</td>
</tr>
<tr>
<td>3</td>
<td>rs4402960</td>
<td>T</td>
<td>186994389</td>
<td>3587</td>
<td>-0.05843</td>
</tr>
<tr>
<td>6</td>
<td>rs4712523</td>
<td>G</td>
<td>20765543</td>
<td>3592</td>
<td>0.01223</td>
</tr>
<tr>
<td>7</td>
<td>rs756992</td>
<td>G</td>
<td>20787688</td>
<td>3591</td>
<td>0.02923</td>
</tr>
<tr>
<td>7</td>
<td>rs1635852</td>
<td>C*</td>
<td>27962651</td>
<td>3590</td>
<td>0.009886</td>
</tr>
<tr>
<td>8</td>
<td>rs13266634</td>
<td>C*</td>
<td>118253964</td>
<td>3590</td>
<td>0.003039</td>
</tr>
<tr>
<td>9</td>
<td>rs2383207</td>
<td>A*</td>
<td>22105959</td>
<td>3591</td>
<td>0.04088</td>
</tr>
<tr>
<td>10</td>
<td>rs11257622</td>
<td>C</td>
<td>12335345</td>
<td>3580</td>
<td>-0.08373</td>
</tr>
<tr>
<td>10</td>
<td>rs1111875</td>
<td>C*</td>
<td>94452862</td>
<td>3592</td>
<td>0.08005</td>
</tr>
<tr>
<td>10</td>
<td>**rs7923837</td>
<td>G*</td>
<td>94471897</td>
<td>3592</td>
<td>0.0913</td>
</tr>
<tr>
<td>10</td>
<td>rs7903146</td>
<td>T</td>
<td>114748339</td>
<td>3592</td>
<td>-0.01407</td>
</tr>
<tr>
<td>11</td>
<td>rs163171</td>
<td>C*</td>
<td>2777641</td>
<td>3588</td>
<td>-0.03288</td>
</tr>
<tr>
<td>11</td>
<td>rs9300039</td>
<td>C*</td>
<td>41871942</td>
<td>3585</td>
<td>-0.08931</td>
</tr>
<tr>
<td>11</td>
<td>rs7480010</td>
<td>G</td>
<td>42203294</td>
<td>3591</td>
<td>0.02035</td>
</tr>
<tr>
<td>11</td>
<td>rs729287</td>
<td>C*</td>
<td>44236666</td>
<td>3592</td>
<td>-0.0188</td>
</tr>
<tr>
<td>11</td>
<td>rs1387153</td>
<td>T</td>
<td>92313476</td>
<td>3592</td>
<td>-0.003922</td>
</tr>
<tr>
<td>12</td>
<td>rs1353362</td>
<td>C</td>
<td>69899543</td>
<td>3581</td>
<td>-0.01765</td>
</tr>
<tr>
<td>16</td>
<td>**rs3751812</td>
<td>T</td>
<td>52375961</td>
<td>3587</td>
<td>0.1159</td>
</tr>
</tbody>
</table>

BP: Base pair position (dbSNP build 125); N: number of individuals tested; Effect Size: regression coefficient for the test SNP; SE: standard error of the regression coefficient; Test statistic: additive model; P: unadjusted two-sided trend test P-value. The direction of effect is shown for the T2D-risk allele in each case. *The T2D-risk allele is the major allele; **P≤0.0025 in the discovery cohort i.e. survive Bonferroni correction for number of variants tested.