Diabetes Publish Ahead of Print, published online September 10, 2007

Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: Evidence for replication from diabetes-related quantitative traits and from independent populations

Evadnie Rampersaud[1], Coleen M. Damcott[1], Mao Fu[1], Haiqing Shen[1], Patrick McArdle[1], Xiaolian Shi[1], John Shelton[1], Jing Yin[1], Christy Y. Chang[1], Sandra H. Ott[1], Li Zhang[1], Yiju Zhao[1], Braxton D. Mitchell[1], Jeffery O'Connell[1], and Alan R. Shuldiner[1,2]

[1]Division of Endocrinology, Diabetes and Nutrition, University of Maryland, Baltimore MD
[2]Geriatric Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD

Running title: A T2D GWAS in the Amish

Corresponding author:
Alan R. Shuldiner, M.D.
Division of Endocrinology, Diabetes and Nutrition
University of Maryland School of Medicine
660 West Redwood Street, Room 494
Baltimore, MD 21201
Email : Ashuldin@medicine.umaryland.edu

Received for publication 3 April 2007 and accepted in revised form 5 September 2007.

Additional information for this article can be found in an online appendix at http://diabetes.diabetsjournals.org.

ABSTRACT

Objective. Type 2 diabetes (T2D) is the result of complex interactions between genetic susceptibility and environmental factors.

Research Design and Methods. We performed a genome-wide association scan (GWAS) or T2D in the Amish; 93,087 SNPs were tested for association in 124 T2D cases and 295 normoglycemic controls. We tested the most strongly T2D-associated for association with oral glucose tolerance (OGTT) traits in 427 non-diabetic Amish individuals, and replication from GWAS from four independent populations (Caucasians from the Framingham Heart Study, Mexican Americans, Pima Indians, and Caucasians from Scandinavia).

Results. We identified 1,316 SNPs that were associated with T2D in the Amish ($p<0.01$). The strongest association ($p=1.07 \times 10^{-5}$) was for rs2237457, which is located in intron 4 of growth factor receptor-bound protein 10 (*GRB10*), an adaptor protein known to regulate signaling of insulin receptors. Of the T2D-associated SNPs in the Amish, 65 unique SNPs demonstrated internal associations ($p<0.01$) with at least one of five OGTT traits in non-diabetic individuals. Genes notable for internal associations include *ESRRG, FHIT, ADAM12,* and *C10orf59.* In external comparisons, 73 unique SNPs were associated with T2D in the Amish and at least one other population. Among these, SNPs in *SLC8A1* and *BCAT1* were associated with T2D in the Amish and two other populations. Other genes of potential interest include *GPC5, CSN3, INPP4B, LOC730727,* and *C1orf24*.

Conclusions. The results of our 100K GWAS of T2D in the Amish identified a tractable number of novel candidate genes that warrant further investigation.

Abbreviations: AFDS, Amish Family Diabetes Study; DGI, Broad-Lund-Novartis Diabetes Genome Initiative; FHS, Framingham Heart Study; GWAS, genome-wide association scan; HOMA-IR, homeostasis model assessment of insulin resistance; HWE; Hardy-Weinberg equilibrium; IGT, impaired glucose tolerance; ISI, insulinogenic index; LD, linkage disequilibrium; MAF, minor allele frequency; NGT, normal glucose tolerant; OGTT, oral glucose tolerance test; OR, odds ratio; SNP, single nucleotide polymorphism; T2D, type 2 diabetes mellitus

Type 2 diabetes (T2D), a complex disease that is characterized by insulin resistance and impaired beta-cell function, represents a serious global public health problem with over 100 million people worldwide who are affected. While the primary molecular defects in T2D remain largely unknown, it is clear that both genetic and environmental risk factors (including diet and physical inactivity) play critical roles. More than twenty genome-wide linkage scans of T2D have been published, with evidence for linkage reported to a number of loci, including regions on chromosomes 1, 3, 8, 10, 12, 14 and 20 (1-8). Of the numerous candidate genes studied for their functional role in pancreatic beta cell function, insulin action or energy metabolism, as well as positional candidate genes identified under linkage peaks, very few have variants that are consistently associated with T2D. Indeed, common variants in only a handful of genes (*PPARγ, KCNJ11*, *CALPN10, TCF7L2, HNF4A*) have been replicated in multiple populations (9).

The Old Order Amish are a closed founder population who emigrated from Switzerland in the early 1700's. They are a well-suited population for carrying out genetic studies since they live a relatively homogeneous lifestyle and maintain extensive family history records. The Amish Family Diabetes Study (AFDS) was initiated in 1995 with the goal of identifying the genetic determinants of T2D (10). We found that the sibling relative risk (λ_s) of T2D in the Amish is 3.28 (95% CI 1.58-6.80), similar to that observed in other Caucasian populations. Genome-wide linkage analysis of T2D and impaired glucose tolerance (IGT) conducted in AFDS pedigrees (6) revealed regions on chromosomes 1q and 14q, both of which have been implicated in linkage scans from other populations (1-5;7). Specific variants in several well-replicated T2D susceptibility genes are associated with T2D in the Amish, including *TCF7L2* rs7903146 (p=0.008; OR = 1.53) (11) and *HNF4A* rs2425640 (p = 0.03; OR = 1.60) (12). These findings suggest that the common T2D gene variants in the Amish will likely be relevant to more outbred Caucasian populations.

Increased knowledge of common variation in the human genome learned as part of the HapMap initiative, coupled with advances in technologies that make possible the genotyping of hundreds of thousands of single nucleotide polymorphisms (SNPs) in many DNA samples from those with and without disease, have made possible genome-wide association scans (GWAS), a powerful approach to identify novel susceptibility genes for complex diseases (13;14). Recently, four GWAS studies of T2D have identified variants at several novel loci, including *SLC30A8, IGF2BP2, CKDAL1, CDKN2A/CKDN2B,* and *HHEX/IDE*, that show strong replicated association with T2D (15-18). In this paper, we report results from a GWAS of T2D in the Amish using the Affymetrix 100K SNP genotyping platform. We further characterize our findings using diabetes-related quantitative traits measured in nondiabetic Amish individuals. Lastly, we interpret the results of this scan in the context of three recently completed 100K GWAS studies for T2D, as part of the T2D 100K GWAS Consortium along with a publicly available 500K GWAS of T2D recently performed in a Scandinavian population.

METHODS

Study population and phenotype assessment

Individuals with T2D were identified from the Amish Family Diabetes Study (AFDS). Details of the AFDS have been previously described (10). Phenotypic characterization of participants included medical and family history, anthropometry, and a 3-hour, 75-gram oral glucose tolerance test (OGTT) with insulin levels. We based our

primary analyses on 124 T2D cases and 295 normal glucose tolerant (NGT) controls. T2D was defined by fasting plasma glucose level (≥ 7 mmol/l), 2-h OGTT plasma glucose level (≥ 11.1 mmol/l), random plasma glucose level (≥ 11.1 mmol/l), the use of insulin or prescription oral glucose-lowering agents, or a diagnosis of diabetes documented by a physician. To minimize potentially misclassifying subjects with type 1 diabetes as having type 2 diabetes, cases with age of diagnosis <35 years were excluded. NGT controls were >38 years old at the time of study and were selected based on fasting plasma glucose level (<6.1 mmol/l) and 2-h OGTT plasma glucose level (<7.8 mmol/l).

We performed secondary quantitative analyses of our mostly highly associated signals ($p<0.01$) in a set of 427 nondiabetic Amish study participants, 132 of whom had impaired glucose tolerance (IGT) and 295 of whom were part of the NGT control group used in our primary analysis. We estimated the mean levels of two OGTT-derived quantitative glucose traits [fasting glucose (FASTG) and glucose area under the curve (GAUC)], and three insulin traits [insulinogenic index (ISI), insulin area under the curve (IAUC), and homeostasis model assessment of insulin resistance (HOMA-IR)] according to the SNP genotypes in these individuals. Total GAUC and IAUC were calculated based on measurements at 0, 30, 60, 90, 120, 150, and 180 minutes using the trapezoidal method. The ISI was calculated as ([insulin at 30 min – fasting insulin] / [glucose at 30 min – fasting glucose]). HOMA-IR was calculated as (fasting insulin (mU/l) x fasting glucose (mmol/l) / 22.5). Table 1 describes the characteristics of this sample. The study protocol was approved by the Institutional Review Board at the University Of Maryland School of Medicine, and informed consent was obtained from each study participant.

Genotyping

Genomic DNA from leukocytes were genotyped using the Affymetrix GeneChip® Mapping 100K Array set which consists of two microarray chips (*XbaI* and *HindIII*) (Affymetrix, Santa Clara, CA). Total genomic DNA (250 ng) was digested with *XbaI* or *HindIII* restriction enzymes and processed according to the Affymetrix protocol. The GeneChip Genotyping Analysis Software (GTYPE 4.0) was used to generate DM-algorithm derived genotypes that were re-analyzed with the BRLMM genotype calling algorithm (confidence threshold of 0.33) to improve the proportion of heterozygote calls (19) As an initial quality control (QC) measure, BRLMM-generated chip files with call rates <90% for both enzymes across all SNPs were excluded. The resulting median call rate across of the remaining 419 case-control samples was 97.5% (97.6% for *Xba1* and 97.4% for *HindIII*). We further removed individual SNPs with genotype call rates <90%, monomorphic SNPs and SNPs with minor allele frequency (MAF) < 5%, and those deviating from Hardy-Weinberg equilibrium (HWE) in controls ($p < 0.001$). The number of monomorphic and low frequency SNPs (n = 26,816) in the Amish was not appreciably different from that observed in more outbred Caucasians of the HapMap CEU sample. For this report, we focused our analysis on the 82,485 autosomal SNPs that passed our quality control standards.

The concordance rate for 11 QC samples that were run twice on the Affymetrix GeneChip® mapping panel was 97.5%. We also calculated a cross-platform concordance rate of 98% for 419 samples in which 61 SNPs were genotyped using the Affymetrix GeneChip® Mapping 100K panel and an independent Illumina 1536-plex GoldenGate assay. Online Supplementary Table 1 (available at http://diabetes.diabetesjournals.org)

summarizes the quality checks and informativeness of the data.

Association testing and SNP prioritization scheme

Our GWAS analysis and SNP prioritization scheme is shown in Figure 1. We selected the SNPs most highly associated with T2D in our Amish case-control dataset based on *p*-value rankings (*p*-value cutoff < 0.01) and then used two complementary approaches to further prioritize them. In one approach, we evaluated the most highly T2D-associated SNPs for association with diabetes-related quantitative traits in an expanded set of 427 nondiabetic Amish subjects, 295 of whom were NGT controls from the primary T2D association analysis (internal consistency). In a parallel approach, we assessed replication of the most highly associated T2D-associated SNPs in the Amish in four independent GWAS from other populations (external replication).

T2D association analysis

We performed case-control association analysis using a variance component approach as implemented in the SOLAR software (20). Using a liability threshold model, we modeled the probability that the subject was a case or control as a function of the individual's age, sex, and genotype, conditional on the correlations in phenotype among relative pairs. Statistical testing was performed using a likelihood ratio test, in which we compared the likelihood of the data under a model in which the genotype effect was estimated against the likelihood of a nested model in which the genotype effect was constrained to be zero. Odds ratios (OR) were computed from variance components models. We chose to report the additive model as our primary analysis. Supplementary analyses using a dominant or recessive model did not yield any SNP showing genome-wide significance. Of the 82,485 SNPs, 611 had $p < 0.01$ under a dominant model and 569 had $p < 0.01$ under the recessive model. Our complete dataset with results from all models are available online (http://www.medschool.umaryland.edu/amishstudies/index.asp). Pairwise linkage disequilibrium (LD) correlation statistics (r^2) were computed using the Helixtree software version 5.0.2 (GoldenHelix, Bozeman, MT).

Quantitative trait analysis

For quantitative trait analyses performed in nondiabetic Amish subjects, we used the measured genotype approach in which we estimated the likelihood of an additive genetic model given the pedigree structure (21). Prior to analysis, all insulin traits (IAUC, ISI, HOMA-IR) were transformed by their natural logarithm to reduce skewness. Parameter estimates were obtained by maximum likelihood methods, and the significance of association was tested by the likelihood ratio test. Within each model, we simultaneously estimated the effects of age and sex. These analyses were performed using the SOLAR program (20).

Power calculations

Power calculations, based on the Genetic Power calculator of Purcell et al. (22) indicated that our sample would provide 80% power to detect a diabetes susceptibility allele having a genotype relative risk of 1.8 (for allele frequency of 30%, 124 cases and 295 controls, 8% population prevalence of diabetes, assuming a multiplicative model), and 80% power to detect a QTL accounting for 4% or higher of the trait variance for a continuously distributed phenotype (427 subjects).

In silico replication samples

We considered whether our best T2D association signals (*p*-value cutoff <0.01) replicated in at least one of three distinct populations (Framingham Caucasians, Mexican-

Americans, Pima), each with different study designs but performed using the same Affymetrix 100K genotyping platform. Descriptions of each of the T2D 100K GWAS Consortium study populations are provided in accompanying reports (23-25) and in a Joint Summary Supplementary Table 1. We directly checked whether any of the 1,093 SNPs with the best T2D association signals ($p < 0.01$) in the Amish was also associated with T2D based on GEE (generalized estimating equations) and FBAT (family-based association tests) tests in the Framingham Heart Study (FHS), Fisher's exact allelic association test in the Mexican-American study, and case-control and sib-based association tests in the Pima study. We also utilized publicly available pre-released data (March 2007) from a T2D GWAS carried out in a Scandinavian cohort of 1,464 T2D cases and 1,467 matched controls and genotyped using the Affymetrix 500K platform by the Broad-Lund-Novartis Diabetes Genetics Initiative (DGI) (http://www.broad.mit.edu/diabetes/) (18). We specifically checked replication of 295 of 1,093 of our most highly T2D-associated SNPs that were present on both 100K and 500K Affymetrix genotyping arrays. Since LD structure may differ across populations and to limit multiple comparisons, we defined replication only if the same SNP was associated with T2D at $p < 0.05$ with an odds ratios (ORs) in the same direction (i.e., reflective of the same allelic risk).

RESULTS

Following quality-control and HWE checks, 82,485 informative SNPs were included in our analyses. The median physical inter-SNP distance was 11.3 kb and the average distance between SNPs was 29 kb. Under the additive model, a total of 1,093 SNPs, some of which were in linkage disequilibrium (LD), were associated ($p < 0.01$) with T2D (Figure 2). The 50 most strongly T2D-associated SNPs (i.e., lowest *p*-values), are shown in Table 2. The complete dataset is available online at (http://www.medschool.umaryland.edu/amish studies/index.asp). No SNP was associated with T2D at a conservative Bonferroni-corrected level. The strongest association ($p = 1.07 \times 10^{-5}$) was for rs2237457 on chromosome 7, which is located in intron 4 of growth factor receptor-bound protein 10 (*GRB10*), an adaptor protein known to regulate signaling of insulin and insulin-like growth-factor receptors (26-28). In addition to *GRB10*, 15 SNPs were associated with T2D at $p < 1 \times 10^{-4}$ (Figure 2 and Table 2). These SNPs are located in or near *MSH6* (chr 2), *PRKG2* (chr 4), *COL13A1* (chr 10), *MTHFSD* (chr 16), and *SPECC1* (chr 17), none of which are obvious candidate genes for T2D. Adjustment for BMI did not have a large impact on the strength of the associations of these SNPs with T2D (Table 2).

As a measure of internal consistency, we tested whether the 1,093 SNPS associated with T2D ($p < 0.01$) were also associated with OGTT-derived quantitative traits in nondiabetic individuals. In these analyses, we considered two OGTT glucose traits (FASTG and GAUC) and three OGTT insulin traits (IAUC, HOMA-IR and ISI), and $p < 0.01$ as our threshold for significance. Thirty-eight non-redundant ($r^2 < 0.80$) T2D-associated SNPs were also associated with at least one glucose trait and showed the same allelic association as for diabetes (i.e., the diabetes risk allele was also associated with higher glucose levels), while 29 non-redundant T2D-associated SNPs were also associated with at least one insulin-related trait (Table 3). Of the top 16 SNPs associated with T2D at $p < 1 \times 10^{-4}$, rs2237457 in *GRB10* is the only one also associated with an OGTT trait ($p = 0.001$ for GAUC). Two perfectly correlated ($r^2 = 1$) T2D-associated SNPs in *ADAMTS1* (chromosome 5) ($p = 0.004$-0.005) were associated with one glucose trait (p=0.006 for GAUC) and one insulin trait (p=0.007 for IAUC).

We next sought to determine which of our 1,093 most highly T2D-associated SNPs

were also associated with T2D in any of three independent populations for which the same 100K Affymetrix platform was used, or in the DGI Scandinavian population for which the 500K Affymetrix platform was used. We identified 80 non-redundant SNPs for which the same risk allele was also associated with T2D in one of the three studies from the T2D 100K GWAS Consortium ($p < 0.05$) and 11 non-redundant SNPs that showed consistent association in the DGI sample ($p < 0.05$) (Supplementary Table 2). In total, 3 SNPs demonstrated associations in the Amish as well as in two independent populations. The T allele for rs2540317 in *MFSD9* on chromosome 2 was associated with decreased risk of T2D in the Amish (OR 0.72, $p = 0.007$), and showed nominal association in the Pima dataset (case-control OR 0.67, $p = 0.042$; sib-based OR 0.50, $p = 0.043$; summary OR 0.63, $p = 0.016$) and also in the Mexican-American sample (case-control OR 0.75, $p = 0.047$). The G allele in rs10515353 on chromosome 5 was associated with decreased risk of T2D in the Amish (OR 0.61, $p = 0.005$) and also with decreased T2D risk in samples from Mexican-American (OR 0.69, $p = 0.035$) and DGI (OR 0.79, $p = 0.007$). The T allele in rs2242400 in *BCAT1* on chromosome 10 was associated with decreased risk of T2D in the Amish (OR 0.71, $p = 0.004$), and also in the Pima dataset (Sib OR 0.66, $p = 0.019$; summary OR 0.78, $p = 0.034$) and the Mexican-American dataset (OR 0.67, $p = 0.009$); borderline association was also seen in the DGI sample (OR 0.86, $p = 0.051$). The direction of effect was the same for all studies.

Table 4 highlights our most consistent overall findings. We present 21 T2D-associated SNPs in the Amish ($p < 0.005$) that also demonstrated either (a) association with a diabetes-related quantitative trait ($p < 0.005$) in the Amish or *in-silico* replication of T2D association in one independent population ($p < 0.005$). Of interest, the T allele in rs3845971 in *FHIT* was associated with increased risk of T2D in the Amish (OR 1.42, $p = 0.004$) and also in Mexican-Americans (OR 1.46, $p = 0.004$) and with increased GAUC ($p = 4.0 \times 10^{-4}$) in nondiabetic Amish subjects.

DISCUSSION

In this report we described the results of a GWAS of T2D of 82,485 SNPs in the Old Order Amish, a genetically closed founder population with a homogeneous lifestyle. We reasoned that this population is likely to carry a subset of the same common T2D susceptibility variants as in the general population, and that these variants might be easier to identify.

GWAS studies are prone to false positives due to the very large number of statistical tests that must be performed. We were restricted by our relatively modest sample size, and also computationally in our attempts to define a genome-wide significance level for which follow-up was justified (ie. variance components tests were not feasible for the many replications needed for case-control permuted family datasets in the Amish). Thus we relied heavily on a prioritization of SNPs worthy of follow-up by testing for (a) internal consistency of T2D-associated SNPs with OGTT-derived quantitative traits in nondiabetic Amish individuals, (b) external replication of T2D-associations in three independent non-Amish 100K SNP GWAS studies, and (c) external replication in a 500K SNP GWAS of T2D in a large population of Scandinavians.

We found that no single SNP replicated consistently and in the same direction across all GWAS studies, nor were all SNPs associated with T2D also associated with quantitative traits in nondiabetic individuals. This is not particularly surprising since we expect that an appreciable number of T2D-associated SNPs will be false positives. Furthermore, a true susceptibility gene in one

population might not be readily discernible in other populations due to small samples sizes as well as differences in genetic background, linkage disequilibrium, and environmental exposures. Similarly, a true susceptibility gene for T2D might not show association with diabetes-related quantitative traits in nondiabetic individuals, especially since our OGTT-derived traits are only surrogates for gold-standard measures of insulin sensitivity and insulin secretion. Nevertheless, we were able to identify a number of candidate genes and loci that showed evidence for association with T2D in more than one population and/or were also associated with OGTT-derived quantitative traits. These results are intriguing but must be interpreted with caution. None of these loci fall within previously identified linkage regions for T2D (chromosomes 1 and 14) in the Amish.

Our strongest T2D association signal in the Amish was observed on chromosome 7 in a functionally relevant T2D candidate gene, *GRB10*. *GRB10* encodes several different isoforms of growth factor binding protein 10, and has been shown to bind to activated insulin receptor and act as a negative regulator of insulin action and glucose uptake (26-28). Overexpression of *GRB10* in mice causes postnatal growth retardation and insulin resistance (29). Our 100K GWAS contained a total of twelve SNPs in *GRB10*, six of which were associated with T2D ($p < 0.05$) and were in partial LD with each other ($r^2 = 0.16 - 0.78$). Rs2237457, located in intron 4, provided the lowest p-value for association (OR 0.61 for the G vs. A allele, $p = 1.07 \times 10^{-5}$). This SNP was also strongly associated with OGTT glucose area under the curve in nondiabetic Amish individuals ($p = 0.001$). Rs2237457 was not associated with T2D in the other three populations in which this SNP was genotyped or in the 500K SNP Scandanavian T2D GWAS; however, three SNPs (rs2190496, rs2237478, and rs7805310) in *GRB10* that were genotyped in the Scandinavian cohort were associated with T2D ($p = 0.029$, $p = 0.01$, $p = 0.004$, respectively), and are in partial LD with rs2237457 (r^2 =0 .12-0.49 in HapMap CEU). Lack of replication could suggest a false positive or that variation in *GRB10* is a true positive specific to the Amish due to a founder effect or context dependent phenotypic expression of the variant due to genetic background or environmental influences. Alternatively, this variant could be in LD with a functional variant and extended LD in the Amish enabled a T2D association to be detected in this population and not the others.

In a recent report by Di Paola et al. (30), the A allele of rs4947710, a synonymous coding SNP in *GRB10,* was associated with decreased risk of T2D in a relatively homogeneous population of Italian Caucasians ($p = 0.0001$). This SNP was not part of the 100K SNP panel nor was our most highly T2D-associated SNP (rs2237457) genotyped in the Italian sample. We found that rs2237457 and rs4947710 are not in LD ($r^2 = 0$) in HapMap CEU samples. However, rs10486757, another *GRB10* SNP associated with T2D in the Amish ($p = 0.024$), is in LD with rs4947710 ($r^2 = 0.64$ in HapMap CEU). Further investigation of *GRB10* is currently underway.

Our GWAS and replication strategy has several limitations. First, the relatively small sample size limits our ability to detect gene variants of modest effect size. Second, we recognize that the definition of external replication of our top SNPs across three independent 100K studies of T2D might represent a skewed distribution of the overall results since replication was limited to our ~1000 most highly T2D-associated SNPs. This approach was used to facilitate comparisons across populations and also to limit the number of false positive replications due to multiple comparisons. To the extent that we attempted to pursue signals that

represent the 'lowest hanging fruit', we believe that the approach we have taken is reasonable. A formal meta-analysis of the entire set of data from all four 100K studies is currently underway. Third, our replication approach was focused at the level of the SNP in order to avoid additional multiple comparisions. However, it is possible that we did not identify significantly-associated SNPs in other populations that were in LD with our top SNPs. This is particularly relevant for our comparisons with the Scandinavian 500K GWAS, for which we only assessed 27% of the SNPs identified in the Amish with $p < 0.01$.

The likelihood that we missed common variants important to T2D is high due to the relatively sparse density of the 100K SNP panel (mean intermarker distance = 29 kb) compared with other denser GWAS SNP panels. For example, SNPs in well-replicated genes *SLC30A8, IGF2BP2, CKDAL1, CDKN2A/CKDN2B, HHEX/IDE* found in four recently published T2D GWAS studies (15-18), and previously known T2D-associated variants in *TCF7L2*, *KCNJ11*, *HNF4A*, or *CAPN10* (9) were not adequately covered on the 100K genotyping panel, i.e., $r^2 < 0.8$ between the SNP of interest and SNPs on the 100K panel). As a positive control, we previously demonstrated *TCF7L2* SNP rs7903146 and the *HNF4A* promoter SNP rs2425640, neither of which is present on the 100K panel, were associated with T2D and IGT in the Amish Family Diabetes Study ($p = 0.008$ OR = 1.57 and $p = 0.04$; OR = 1.60, respectively) (12;28). Interestingly, rs10509645 in *HHEX* on the 100K panel ($r^2 = 0.7$ with rs7923837 found previously to be strongly associated with T2D in other GWAS studies), was significantly associated with T2D in the Amish (OR = 1.30 for the G allele; $p = 0.02$). Rs9300039 on chromosome 11, shown to be associated with T2D in the other GWAS studies (17) was present on the 100K panel, but was not significantly associated with T2D in the Amish (OR = 1.09 for the C allele, $p = 0.67$).

In summary, we presented results from our initial examination of a GWAS of T2D in the Amish. Although we did not identify any genes associated with T2D that reached genome-wide significance, we report a number of genes and loci that are worthy of further study based on replication in other studies or on QTL consistency. This report (and the three companion manuscripts) provides a valuable resource for other investigators to utilize in the search for the pathogenic variants for T2D.

ACKNOWLEDGMENTS

This work was supported by NIH research grants T32 HL072751 Training Grant in Cardiac and Vascular Cell Biology; R01 DK54261; the University of Maryland General Clinical Research Center (M01 RR16500); Hopkins Bayview General Clinical Research Center (M01 RR02719); the Maryland Clinical Nutrition Research Unit (P30 DK072488); and the Baltimore Veterans Administration Geriatric Research and Education Clinical Center. We extend our thanks to our collaborators in the Type 2 Diabetes 100K GWAS Consortium for sharing pre-publication results from the Starr County Health Studies, Framingham Heart Study, and the Pima Indians, and to the Broad-Lund-Novartis Diabetes Genomic Initiative for access to results from their GWAS. We also thank Soren Snitker for helpful comments and Adam Naj for help with formatting figures for this manuscript. Lastly, we gratefully acknowledge our Amish liaisons and fieldworkers and the extraordinary cooperation and support of the Amish community, without whom these studies would not be possible.

REFERENCE LIST

1. Elbein,SC, Hoffman,MD, Teng,K, Leppert,MF, Hasstedt,SJ: A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* 48:1175-1182, 1999
2. Wiltshire,S, Hattersley,AT, Hitman,GA, Walker,M, Levy,JC, Sampson,M, O'Rahilly,S, Frayling,TM, Bell,JI, Lathrop,GM, Bennett,A, Dhillon,R, Fletcher,C, Groves,CJ, Jones,E, Prestwich,P, Simecek,N, Rao,PV, Wishart,M, Foxon,R, Howell,S, Smedley,D, Cardon,LR, Menzel,S, McCarthy,MI: A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 69:553-569, 2001
3. Du,W, Sun,H, Wang,H, Qiang,B, Shen,Y, Yao,Z, Gu,J, Xiong,M, Huang,W, Chen,Z, Zuo,J, Hua,X, Gao,W, Sun,Q, Fang,F: Confirmation of susceptibility gene loci on chromosome 1 in northern China Han families with type 2 diabetes. *Chin Med J (Engl)* 114:876-878, 2001
4. Meigs,JB, Panhuysen,CI, Myers,RH, Wilson,PW, Cupples,LA: A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of Caucasian pedigrees: The Framingham Offspring Study. *Diabetes* 51:833-840, 2002
5. Hanson,RL, Ehm,MG, Pettitt,DJ, Prochazka,M, Thompson,DB, Timberlake,D, Foroud,T, Kobes,S, Baier,L, Burns,DK, Almasy,L, Blangero,J, Garvey,WT, Bennett,PH, Knowler,WC: An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63:1130-1138, 1998
6. Hsueh,WC, St,JP, Mitchell,BD, Pollin,TI, Knowler,WC, Ehm,MG, Bell,CJ, Sakul,H, Wagner,MJ, Burns,DK, Shuldiner,AR: Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* 52:550-557, 2003
7. Ng,MC, So,WY, Cox,NJ, Lam,VK, Cockram,CS, Critchley,JA, Bell,GI, Chan,JC: Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* 53:1609-1613, 2004
8. McCarthy,MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159-167, 2003
9. Owen,KR, McCarthy,MI: Genetics of type 2 diabetes. *Curr Opin Genet Dev* 17:239-244, 2007
10. Hsueh,W-C, Mitchell,BD, Aburomia,R, Pollin,T, Sakul,H, Ehm M.G., Michelsen,BK, Wagner,MJ, St.Jean,PL, Knowler,WC, Burns,DK, Bell,CJ, Shuldiner,AR: Diabetes in the old order Amish: characterization and heritability analysis of the Amish Family Diabetes Study . *Diabetes Care* 23:595-601, 2000
11. Damcott,CM, Pollin,TI, Reinhart,LJ, Ott,SH, Shen,H, Silver,KD, Mitchell,BD, Shuldiner,AR: Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* 55:2654-2659, 2006
12. Damcott,CM, Hoppman,N, Ott,SH, Reinhart,LJ, Wang,J, Pollin,TI, O'Connell,JR, Mitchell,BD, Shuldiner,AR: Polymorphisms in both promoters of hepatocyte nuclear factor 4-alpha are associated with type 2 diabetes in the Amish. *Diabetes* 53:3337-3341, 2004

13. Klein,RJ, Zeiss,C, Chew,EY, Tsai,JY, Sackler,RS, Haynes,C, Henning,AK, SanGiovanni,JP, Mane,SM, Mayne,ST, Bracken,MB, Ferris,FL, Ott,J, Barnstable,C, Hoh,J: Complement factor H polymorphism in age-related macular degeneration. *Science* 308:385-389, 2005
14. Gibbs,JR, Singleton,A: Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet* 2:e150, 2006
15. Zeggini,E, Weedon,MN, Lindgren,CM, Frayling,TM, Elliott,KS, Lango,H, Timpson,NJ, Perry,JR, Rayner,NW, Freathy,RM, Barrett,JC, Shields,B, Morris,AP, Ellard,S, Groves,CJ, Harries,LW, Marchini,JL, Owen,KR, Knight,B, Cardon,LR, Walker,M, Hitman,GA, Morris,AD, Doney,AS, Burton,PR, Clayton,DG, Craddock,N, Deloukas,P, Duncanson,A, Kwiatkowski,DP, Ouwehand,WH, Samani,NJ, Todd,JA, Donnelly,P, Davison,D, Easton,D, Evans,D, Leung,HT, Spencer,CC, Tobin,MD, Attwood,AP, Boorman,JP, Cant,B, Everson,U, Hussey,JM, Jolley,JD, Knight,AS, Koch,K, Meech,E, Nutland,S, Prowse,CV, Stevens,HE, Taylor,NC, Walters,GR, Walker,NM, Watkins,NA, Winzer,T, Jones,RW, McArdle,WL, Ring,SM, Strachan,DP, Pembrey,M, Breen,G, St,CD, Caesar,S, Gordon-Smith,K, Jones,L, Fraser,C, Green,EK, Grozeva,D, Hamshere,ML, Holmans,PA, Jones,IR, Kirov,G, Moskvina,V, Nikolov,I, O'donovan,MC, Owen,MJ, Collier,DA, Elkin,A, Farmer,A, Williamson,R, McGuffin,P, Young,AH, Ferrier,IN, Ball,SG, Balmforth,AJ, Barrett,JH, Bishop,DT, Iles,MM, Maqbool,A, Yuldasheva,N, Hall,AS, Braund,PS, Dixon,RJ, Mangino,M, Stevens,S, Thompson,JR, Bredin,F, Tremelling,M, Parkes,M, Drummond,H, Lees,CW, Nimmo,ER, Satsangi,J, Fisher,SA, Forbes,A, Lewis,CM, Onnie,CM, Prescott,NJ, Sanderson,J, Mathew,CG, Barbour,J, Mohiuddin,MK, Todhunter,CE, Mansfield,JC, Ahmad,T, Cummings,FR, Jewell,DP, Webster,J, Brown,MJ, Lathrop,GM, Connell,J, Dominiczak,A, Braga Marcano,CA, Burke,B, Dobson,R, Gungadoo,J, Lee,KL, Munroe,PB, Newhouse,SJ, Onipinla,A, Wallace,C, Xue,M, Caulfield,M, Farrall,M, Barton,A, Bruce,IN, Donovan,H, Eyre,S, Gilbert,PD, Hider,SL, Hinks,AM, John,SL, Potter,C, Silman,AJ, Symmons,DP, Thomson,W, Worthington,J, Dunger,DB, Widmer,B, Newport,M, Sirugo,G, Lyons,E, Vannberg,F, Hill,AV, Bradbury,LA, Farrar,C, Pointon,JJ, Wordsworth,P, Brown,MA, Franklyn,JA, Heward,JM, Simmonds,MJ, Gough,SC, Seal,S, Stratton,MR, Rahman,N, Ban,M, Goris,A, Sawcer,SJ, Compston,A, Conway,D, Jallow,M, Rockett,KA, Bumpstead,SJ, Chaney,A, Downes,K, Ghori,MJ, Gwilliam,R, Hunt,SE, Inouye,M, Keniry,A, King,E, McGinnis,R, Potter,S, Ravindrarajah,R, Whittaker,P, Widden,C, Withers,D, Cardin,NJ, Ferreira,T, Pereira-Gale,J, Hallgrimsdottir,IB, Howie,BN, Su,Z, Teo,YY, Vukcevic,D, Bentley,D, Compston,A, Ouwehand,NJ, Samani,MR, Isaacs,JD, Morgan,AW, Wilson,GD, rdern-Jones,A, Berg,J, Brady,A, Bradshaw,N, Brewer,C, Brice,G, Bullman,B, Campbell,J, Castle,B, Cetnarsryj,R, Chapman,C, Chu,C, Coates,N, Cole,T, Davidson,R, Donaldson,A, Dorkins,H, Douglas,F, Eccles,D, Eeles,R, Elmslie,F, Evans,DG, Goff,S, Goodman,S, Goudie,D, Gray,J, Greenhalgh,L, Gregory,H, Hodgson,SV, Homfray,T, Houlston,RS, Izatt,L, Jackson,L, Jeffers,L, Johnson-Roffey,V, Kavalier,F, Kirk,C, Lalloo,F, Langman,C, Locke,I, Longmuir,M, Mackay,J, Magee,A, Mansour,S, Miedzybrodzka,Z, Miller,J, Morrison,P, Murday,V, Paterson,J, Pichert,G: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336-1341, 2007
16. Steinthorsdottir,V, Thorleifsson,G, Reynisdottir,I, Benediktsson,R, Jonsdottir,T, Walters,GB, Styrkarsdottir,U, Gretarsdottir,S, Emilsson,V, Ghosh,S, Baker,A, Snorradottir,S, Bjarnason,H, Ng,MC, Hansen,T, Bagger,Y, Wilensky,RL, Reilly,MP,

Adeyemo,A, Chen,Y, Zhou,J, Gudnason,V, Chen,G, Huang,H, Lashley,K, Doumatey,A, So,WY, Ma,RC, Andersen,G, Borch-Johnsen,K, Jorgensen,T, van Vliet-Ostaptchouk,JV, Hofker,MH, Wijmenga,C, Christiansen,C, Rader,DJ, Rotimi,C, Gurney,M, Chan,JC, Pedersen,O, Sigurdsson,G, Gulcher,JR, Thorsteinsdottir,U, Kong,A, Stefansson,K: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770-775, 2007

17. Scott,LJ, Mohlke,KL, Bonnycastle,LL, Willer,CJ, Li,Y, Duren,WL, Erdos,MR, Stringham,HM, Chines,PS, Jackson,AU, Prokunina-Olsson,L, Ding,CJ, Swift,AJ, Narisu,N, Hu,T, Pruim,R, Xiao,R, Li,XY, Conneely,KN, Riebow,NL, Sprau,AG, Tong,M, White,PP, Hetrick,KN, Barnhart,MW, Bark,CW, Goldstein,JL, Watkins,L, Xiang,F, Saramies,J, Buchanan,TA, Watanabe,RM, Valle,TT, Kinnunen,L, Abecasis,GR, Pugh,EW, Doheny,KF, Bergman,RN, Tuomilehto,J, Collins,FS, Boehnke,M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341-1345, 2007
18. Saxena,R, Voight,BF, Lyssenko,V, Burtt,NP, de Bakker,PI, Chen,H, Roix,JJ, Kathiresan,S, Hirschhorn,JN, Daly,MJ, Hughes,TE, Groop,L, Altshuler,D, Almgren,P, Florez,JC, Meyer,J, Ardlie,K, Bengtsson,BK, Isomaa,B, Lettre,G, Lindblad,U, Lyon,HN, Melander,O, Newton-Cheh,C, Nilsson,P, Orho-Melander,M, Rastam,L, Speliotes,EK, Taskinen,MR, Tuomi,T, Guiducci,C, Berglund,A, Carlson,J, Gianniny,L, Hackett,R, Hall,L, Holmkvist,J, Laurila,E, Sjogren,M, Sterner,M, Surti,A, Svensson,M, Svensson,M, Tewhey,R, Blumenstiel,B, Parkin,M, Defelice,M, Barry,R, Brodeur,W, Camarata,J, Chia,N, Fava,M, Gibbons,J, Handsaker,B, Healy,C, Nguyen,K, Gates,C, Sougnez,C, Gage,D, Nizzari,M, Gabriel,SB, Chirn,GW, Ma,Q, Parikh,H, Richardson,D, Ricke,D, Purcell,S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331-1336, 2007
19. BRLMM: an improved genotype calling method for the GeneChip® Human Mapping 500K Array Set [article online], 2006, Available from http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf
20. Almasy,L, Blangero,J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198-1211, 1998
21. Boerwinkle,E, Chakraborty,R, Sing,CF: The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 50 (Pt 2):181-194, 1986
22. Purcell,S, Cherny,SS, Sham,PC: Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149-150, 2003
23. Florez, JC, Manning AK, Dupuis, J, McAteer, J, Irenze, K, Gianniny, L, Mirel, DB, Fox, CS, Cupples AL, Meigs JB.A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: Replication and integration with other genome-wide datasets.(submitted)
24. Hanson, RL, Bogardus, C, Duggan, D, Kobes, S, Infante, AM, Baier, LJ, Knowler, WC. A search for variants associated with young-onset type-2 diabetes in American Indians among 80,044 single nucleotide polymorphisms. (submitted)
25. Hayes, MG, Pluzhnikov, A, Miyake, K, Sun, Y, Below, JE, Ng M, Roe, CA, Bell, GI, Cox,NJ, Hanis CL. Identification and replication of novel type 2 diabetes genes in Mexican Americans through genome-wide association studies. (submitted)

26. Deng,Y, Bhattacharya,S, Swamy,OR, Tandon,R, Wang,Y, Janda,R, Riedel,H: Growth factor receptor-binding protein 10 (Grb10) as a partner of phosphatidylinositol 3-kinase in metabolic insulin action. *J Biol Chem* 278:39311-39322, 2003
27. Langlais,P, Dong,LQ, Ramos,FJ, Hu,D, Li,Y, Quon,MJ, Liu,F: Negative regulation of insulin-stimulated mitogen-activated protein kinase signaling by Grb10. *Mol Endocrinol* 18:350-358, 2004
28. Mounier,C, Lavoie,L, Dumas,V, Mohammad-Ali,K, Wu,J, Nantel,A, Bergeron,JJ, Thomas,DY, Posner,BI: Specific inhibition by hGRB10zeta of insulin-induced glycogen synthase activation: evidence for a novel signaling pathway. *Mol Cell Endocrinol* 173:15-27, 2001
29. Shiura,H, Miyoshi,N, Konishi,A, Wakisaka-Saito,N, Suzuki,R, Muguruma,K, Kohda,T, Wakana,S, Yokoyama,M, Ishino,F, Kaneko-Ishino,T: Meg1/Grb10 overexpression causes postnatal growth retardation and insulin resistance via negative modulation of the IGF1R and IR cascades. *Biochem Biophys Res Commun* 329:909-916, 2005
30. Di,PR, Ciociola,E, Boonyasrisawat,W, Nolan,D, Duffy,J, Miscio,G, Cisternino,C, Fini,G, Tassi,V, Doria,A, Trischitta,V: Association of hGrb10 genetic variations with type 2 diabetes in Caucasian subjects. *Diabetes Care* 29:1181-1183, 2006

Table 1. Description of sample characteristics for T2D GWAS in the Amish.

T2D case - NGT control dataset

Characteristic	T2D cases (n=124)	NGT controls (n=295)
Males (%)	33	52
Age (yrs)	51.3 ± 10.5	64.4 ± 12.9
BMI (kg/m^2)	29.3 ± 5.8	27.4 ± 4.7

Oral glucose tolerance test derived quantitative trait dataset

Characteristic	All (n=427)	Men (n=200)	Women (n=227)
Age (yrs)	51.9 ± 11.9	52.2 ± 11.9	51.7 ± 11.9
BMI (kg/m^2)	27.7 ± 5.0	26.4 ± 4.0	28.9 ± 5.5
Fasting glucose (mmol/l)	5.09 ± 0.45	5.12 ± 0.47	5.07 ± 0.43
Glucose AUC (mmol/l·hr)	379.2 ± 68.0	362.6 ± 65.3	393.9 ± 67.1
Insulin AUC (mU/l·hr)	137.9 ± 88.9	108.5 ± 60.6	164.5 ±101.3
HOMA-IR (mU·mmol/l^2)	2.6 ± 1.7	2.6 ± 2.2	2.6 ± 1.1
ISI (U/g)	0.9 ± 1.5	0.85 ± 1.9	0.93 ± 1.0

All subjects were nondiabetic. Values are mean ± SD. Abbreviations: AUC, area under oral glucose tolerance test (OGTT) curve; BMI, body mass index; HOMA-IR, homeostatic model assessment of insulin resistance; ISI, insulinogenic index

Table 2. Fifty SNPs most highly associated with T2D from Amish GWAS.

SNP	Chr	Position*	Gene**	Alleles 1/2	Strand	N Cases	N Controls	Allele 2 Cases	Allele 2 Controls	*T2D p*-value†	*OR***	*T2D p*-value‡ (BMI)
rs2237457	7	50693638	***GRB10***	**A/G**	-	124	293	0.53	0.68	**1.07 x 10-5**	0.61	1.46 x 10-5
rs980720	4	82272087	*PRKG2*	A/G	-	122	287	0.80	0.90	1.25 x 10-5	0.52	4.87 x 10-5
rs10509199	10	65295820		G/T	-	124	295	0.31	0.46	1.68 x 10-5	0.62	4.46 x 10-5
rs1373147	17	20147237	*SPECC1*	A/T	-	124	293	0.47	0.60	2.79 x 10-5	0.63	5.26 x 10-6
rs4082516	10	71374662	*COL13A1*	C/G	-	122	293	0.77	0.88	3.19 x 10-5	0.53	1.27 x 10-5
rs10509195	10	65193372		A/C	-	118	281	0.42	0.28	3.30 x 10-5	1.62	7.1E-05
rs1395931	2	123535443		A/G	+	124	293	0.68	0.58	6.36 x 10-5	1.57	2.90 x 10-5
rs3136279	2	47871272	*MSH6*	G/T	-	124	294	0.20	0.11	6.81 x 10-5	1.80	9.2E-05
rs1446732	2	134374210		G/T	-	121	295	0.51	0.37	7.38 x 10-5	1.55	0.0001
rs10485249	6	70161908		C/G	+	124	295	0.82	0.92	7.96 x 10-5	0.54	0.0004
rs2703813	17	20055907	*SPECC1*	A/G	-	123	285	0.51	0.64	8.12 x 10-5	0.64	2.36 x 10-5
rs1916412	10	65340801		C/T	-	124	293	0.62	0.49	8.35 x 10-5	1.54	0.0002
rs1916411	10	65340839		C/G	-	123	292	0.38	0.51	8.48 x 10-5	0.65	0.0002
rs3751797	16	85124993	*MTHFSD*	A/T	-	123	285	0.76	0.64	8.74 x 10-5	1.60	0.0005
rs930621	2	134418548		C/T	+	122	290	0.54	0.41	8.74 x 10-5	1.55	0.0001
rs10509201	10	65342248		A/G	-	123	293	0.62	0.49	8.80 x 10-5	1.53	0.0002
rs2158473	17	20079682	*SPECC1*	C/T	+	119	292	0.51	0.39	0.0001	1.54	3.00 x 10-5
rs2502497	6	75399083		G/T	+	116	279	0.05	0.10	0.0001	0.43	0.0002
rs430123	5	106109632		A/G	-	114	272	0.16	0.27	0.0002	0.61	0.0002
rs10504553	8	75038957	*TCEB1*	A/G	-	115	275	0.77	0.89	0.0002	0.56	0.0003
rs9287428	2	133912538	*FLJ34870*	C/T	+	124	287	0.83	0.71	0.0002	1.62	0.0002
rs10507601	13	55047379		A/G	+	123	294	0.04	0.12	0.0002	0.44	0.0006
rs994952	6	78319674		A/G	+	120	287	0.52	0.42	0.0002	1.52	0.0002
rs7604549	2	133923233	*FLJ34870*	A/G	-	123	294	0.26	0.40	0.0002	0.65	0.0002
rs2737245	8	116727757	*TRPS1*	A/C	-	113	276	0.58	0.71	0.0002	0.67	0.0003
rs1513287	3	114754898	*SIDT1*	A/G	+	120	289	0.58	0.44	0.0002	1.49	0.0003
rs7817780	8	119486147	*SAMD12*	C/T	-	120	294	0.91	0.80	0.0002	1.85	0.0002
rs297765	20	4435111		A/G	+	124	295	0.75	0.63	0.0002	1.55	0.0004