

# A Search for Variants Associated With Young-Onset Type 2 Diabetes in American Indians in a 100K Genotyping Array

Robert L. Hanson,<sup>1</sup> Clifton Bogardus,<sup>1</sup> David Duggan,<sup>2</sup> Sayuko Kobes,<sup>1</sup> Michele Knowlton,<sup>2</sup> Aniello M. Infante,<sup>1</sup> Leslie Marovich,<sup>2</sup> Deb Benitez,<sup>2</sup> Leslie J. Baier,<sup>1</sup> and William C. Knowler<sup>1</sup>

**OBJECTIVE**—To identify genetic variants in linkage disequilibrium with those conferring diabetes susceptibility, a genome-wide association study for young-onset diabetes was conducted in an American-Indian population.

**RESEARCH DESIGN AND METHODS**—Data come from 300 case subjects with type 2 diabetes with age of onset <25 years and 334 nondiabetic control subjects aged ≥45 years. To provide for tests of within-family association, 121 nondiabetic siblings of case subjects were included along with 140 diabetic siblings of control subjects (172 sibships). Individuals were genotyped on the Affymetrix 100K array, resulting in 80,044 usable single nucleotide polymorphisms (SNPs). SNPs were analyzed for within-family association and for general association in case and control subjects, and these tests were combined by Fisher's method, with priority given to the within-family test.

**RESULTS**—There were more SNPs with low *P* values than expected theoretically under the global null hypothesis of no association, and 128 SNPs had evidence for association at *P* < 0.001. The association of these SNPs with diabetes was further investigated in 1,207 diabetic and 1,627 nondiabetic individuals from the population study who were not included in the genome-wide study. SNPs from 10 genomic regions showed evidence for replication at *P* < 0.05. These included SNPs on chromosome 3 near ZNF659, chromosome 11 near FANCF, chromosome 11 near ZBTB15, and chromosome 12 near SENP1.

**CONCLUSIONS**—These studies suggest several regions where marker alleles are potentially in linkage disequilibrium with variants that confer susceptibility to young-onset type 2 diabetes in American Indians. *Diabetes* 56:3045–3052, 2007

**T**ype 2 diabetes is substantially influenced by genetic factors, as indicated by studies of familial aggregation (1–3) and twins (4–6); however, the identity of most of the specific variants that influence diabetes susceptibility remains unknown. Con-

From the <sup>1</sup>Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, Arizona; and the <sup>2</sup>Translational Genomics Research Institute, Phoenix, Arizona.

Address correspondence and reprint requests to Robert L. Hanson, Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, 550 E. Indian School Rd., Phoenix, AZ 85014. E-mail: rhanson@phx.niddk.nih.gov.

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

Published ahead of print at <http://diabetes.diabetesjournals.org> on 10 September 2007. DOI: 10.2337/db07-0462.

Additional information for this article can be found in an online appendix at <http://dx.doi.org/10.2337/db07-0462>.

HRR, hazard rate ratio; PAR, population-attributable risk; SNP, single nucleotide polymorphism.

© 2007 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

sistent, albeit modest, associations have been observed with alleles at PPARG and KCNJ11 (7–8). Recently, Grant et al. (9) identified an association of type 2 diabetes with alleles in TCF7L2. This association, which is of greater magnitude than those for the other polymorphisms, has been widely replicated (10–13). These variants explain only a small fraction of the genetic contribution to type 2 diabetes, so it is likely that additional variants remain unidentified. A number of genome-wide linkage studies have been conducted (14–15), and, while these have revealed several putative susceptibility loci, the variants responsible have not been definitively identified.

Recent advances in technology have produced methods for large-scale genotyping of dense panels of single nucleotide polymorphisms (SNPs). Thus, genome-wide association studies are feasible, and these provide another, potentially powerful, approach for detection of novel variants that influence susceptibility to diabetes. The present study represents a genome-wide association study of type 2 diabetes in the Pima Indians, a population with a high prevalence of obesity and diabetes in whom, when diabetes occurs, it is overwhelmingly (if not exclusively) type 2, even in childhood (16). Analyses of the familial pattern of diabetes in this population show that young-onset diabetes is particularly familial and that genetic determinants are likely to influence age at onset of diabetes (3,17–18); therefore, the present study was designed to detect variants associated with young-onset diabetes.

## RESEARCH DESIGN AND METHODS

The present data come from participants in a longitudinal study conducted in the Gila River Indian Community in central Arizona (19). In this study, community residents aged ≥5 years are invited to a research examination every 2 years. These examinations include a 75-g oral glucose tolerance test. Diabetes is diagnosed if the fasting plasma glucose concentration is ≥7.0 mmol/l, the 2-h plasma glucose concentration is ≥11.1 mmol/l (20), or the diagnosis is made during routine clinical care (19). DNA has been extracted from blood leukocytes. To detect variants associated with young-onset diabetes, individuals were selected from the extremes of the age-of-onset distribution. Thus, 300 case subjects were selected who had developed type 2 diabetes before the age of 25 years; for comparison, 334 control subjects were selected who were nondiabetic and aged ≥45 years when last examined. All individuals were full-heritage American Indian, and all individuals fulfilling criteria were selected regardless of affection status of other family members.

Although the case-control approach is potentially powerful for detection of associated variants (21,22), it is liable to spurious results due to population stratification. Therefore, to allow for within-family association tests that are robust to such confounding, siblings of these case or control subjects who defined discordant sibling pairs were selected. Thus, 121 siblings of case subjects were included who were nondiabetic when last examined and whose age was older than that at which the youngest case of diabetes onset in the family occurred. Likewise, 140 diabetic siblings of control subjects with younger age of onset than that of the oldest control subject in the family were included. These individuals constituted 340 discordant sibling pairs in 172 potentially informative sibships.

TABLE 1  
 Characteristics of individuals in the genome-wide and population-based studies

	<i>n</i>	Percentage men	Age (years)*	BMI (kg/m <sup>2</sup> †)
Genome-wide association study				
Case-control study				
Diabetes	300	38	19.2 ± 4.5 (5.5–24.9)	38.9 ± 8.4
No diabetes	334	48	55.5 ± 9.8 (45.1–87.9)	35.4 ± 8.0
Siblings of case/control subjects				
Diabetes	140	34	41.0 ± 8.3 (25.0–67.4)	38.9 ± 9.3
No diabetes	121	47	27.8 ± 7.9 (12.3–42.1)	36.9 ± 8.9
Population study				
Not in genome-wide study				
Diabetes	1,207	38	39.7 ± 10.6 (10.7–78.0)	38.3 ± 8.2
No diabetes	1,627	46	27.7 ± 11.6 (5.5–79.7)	35.6 ± 8.2
All				
Diabetes	1,561	37	37.2 ± 12.1 (5.5–78.0)	38.5 ± 8.4
No diabetes	1,940	46	31.1 ± 14.5 (5.5–85.3)	35.7 ± 8.2

Data are means ± SD (range) or means ± SD unless otherwise indicated. \*Age at last examination for nondiabetic individuals and age at onset of diabetes for diabetic individuals. †Maximum BMI observed after age 15 years in the longitudinal study.

**Population-based association studies.** To examine the potential importance of associated markers on a population basis, a population-based sample of individuals from the longitudinal study was selected for genotyping with selected markers. This sample consisted of all participants from the longitudinal study with available DNA whose heritage was full Pima and/or Tohono O’odham (a closely related tribe); 1,561 of these individuals had diabetes, while 1,940 were nondiabetic. There were 2,834 individuals who were not included in the genome-wide association study, and analyses in these individuals were used to provide a replication of results from the genome-wide study. Analyses in all 3,501 individuals from the population study (who constituted 1,880 sibships) were used to determine population-based parameters. Because of differences in selection criteria, there were 228 individuals in the genome-wide study who were not in the population-based study. Characteristics of individuals in the various studies are shown in Table 1.

**Genotyping.** DNA was isolated using a proteinase K high-salt ethanol precipitation method. Before genotyping for the genome-wide association study, DNA was purified using Montage PCR plates (Millipore). Individuals were genotyped using the Affymetrix 100K Human Mapping array (Affymetrix, Santa Clara, CA), which contains 115,810 SNPs with known positions on the autosomal and X-chromosomes, according to the manufacturer’s protocol. Genotypes were generated using the BRLMM (Robust Linear Model with Mahalanobis distance classifier and the addition of a Bayesian step) algorithm (23). A total of 50 individuals were genotyped in duplicate, and 5,122 SNPs were excluded from analysis that had discrepancy rates of >2.7% or that produced valid genotypes in <85% of individuals. The mean genotypic concordance rate among the 50 pairs of duplicates was 99.5%. Since spurious associations may occur more frequently with very rare alleles, a further 28,215 SNPs with minor allele frequency <1% were excluded. Hardy-Weinberg equilibrium was tested among all genotyped individuals using a continuity correction to produce a better approximate statistic for rarer alleles (24). Since systematic genotyping errors may produce severe violations of Hardy-Weinberg equilibrium, a further 2,429 SNPs were excluded that deviated from Hardy-Weinberg equilibrium at *P* < 0.001. Thus, the present genome-wide association analyses include results for 80,044 SNPs. Family relationships (parents and siblings) were confirmed by comparison of the observed proportion of alleles identical by state for these markers with that expected. SNPplex (Applied Biosystems, Foster City, CA) was used to genotype individuals in the follow-up population-based studies. Physical positions are given according to NCBI Build 35.

**Statistical analysis.** Analyses were performed using the SAS package (SAS Institute, Cary, NC). The association between genotype and case-control status was assessed with logistic regression. Genotype was analyzed as a numeric variable representing the number (0, 1, or 2) of copies of a given allele. For X-chromosome markers outside the pseudoautosomal regions, men were coded as homozygous. To account for the inclusion of multiple individuals in the same sibship, data were analyzed using a modified regressive model in which, for each individual, the prevalence of case status among all of his or her siblings was included as a covariate (25). This produces an approximation to the regressive model of Bonney, in which covariates are used to model the residual phenotypic correlation among siblings (26). To provide for a specific test of within-family association in sibships discordant for diabetes, data were also analyzed with conditional logistic regression (27). All models included sex as a covariate, and the likelihood ratio test was used to assess statistical

significance. Odds ratios (ORs) were calculated from regression coefficients. In the event that one allele is absent from case or control subjects, parameter estimates will not reliably converge, but the likelihood ratio statistic is still approximately correct. In these cases, the estimated OR is infinite; therefore, we report an OR of infinity for these SNPs.

To summarize the results of case-control and within-family analyses, *P* values for the two tests were combined with Fisher’s method (28). To maintain robustness to population stratification, priority was given to the *P* value for the within-family test (*P*<sub>WI</sub>). This was accomplished by calculating a one-sided *P* value for the case-control test (*P*<sub>CC1</sub>) for an association in the same direction as that observed in the within-family analysis. The contribution of the case-control result (*P*<sub>CC\*</sub>) to the combined test statistic was taken as the maximum of *P*<sub>WI</sub> and 1 – (1 – *P*<sub>CC1</sub>)<sup>2</sup>, where *P*<sub>CC1</sub> employs a correction to partially account for the fact that the two tests have been performed in some of the same individuals. The combined test statistic was calculated as follows:

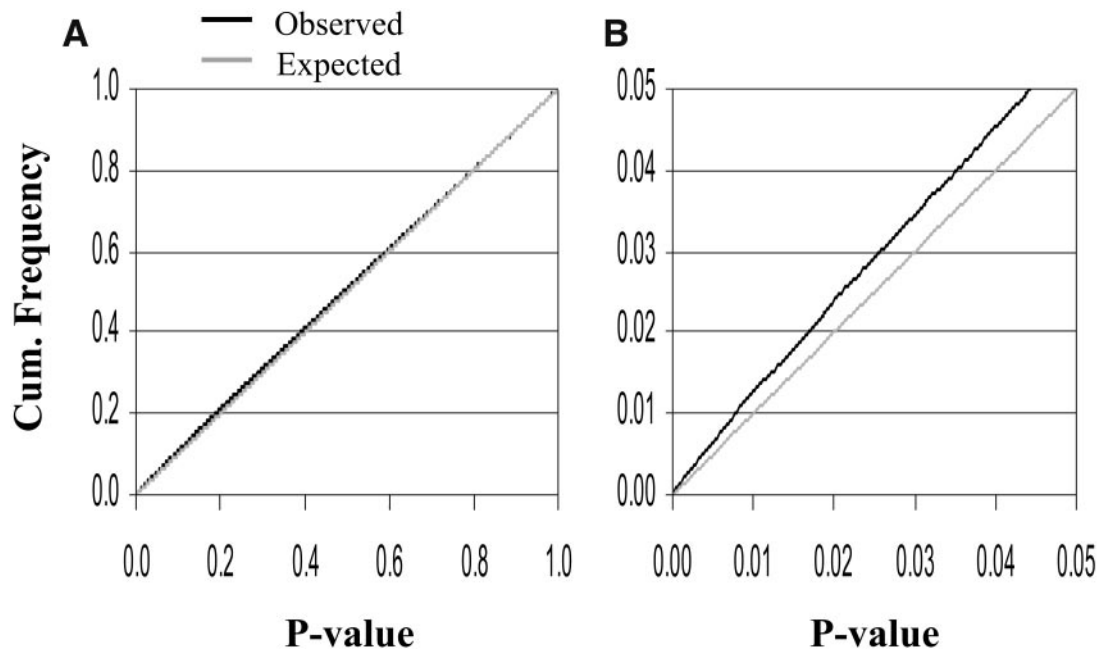
$$\chi^2_{CC-WI} = -2\log_e(P_{WI}) - 2\log_e(P_{CC*})$$

Simulation studies show that this method augments power of the within-family test by using information from the general test while maintaining robustness to population stratification (29). If *P*<sub>WI</sub> and *P*<sub>CC\*</sub> were independent, the *P* value associated with  $\chi^2_{CC-WI}$  on 4 d.f. would have a uniform distribution under the null hypothesis (28). However, because *P*<sub>CC\*</sub> is truncated to be  $\geq P_{WI}$  and because of correlation among the tests, this “nominal” *P* value for  $\chi^2_{CC-WI}$  does not have a uniform distribution. To generate a *P* value corrected for these distributional deviations, 50 million pairs of standardized test statistics, representing within-family and case-control logarithms of the OR divided by their SEs, were simulated from a bivariate distribution with correlation of 0.32 (the observed correlation among these statistics for 80,044 SNPs). In these simulated data,  $\chi^2_{CC-WI}$  was calculated for each pair of tests, and the *P* value for the value of  $\chi^2_{CC-WI}$  observed for a given SNP was taken as the proportion of replicates at which the test statistic for these simulated values exceeded the observed value.

**Permutations.** To assess experiment-wide statistical significance, affection status was permuted across individuals and analyses of the association of genotypes, with permuted affection status repeated genome wide. A total of 200 permutations were thus analyzed. To maintain the familial nature of the study, all data for a sibship were permuted in tandem (across sibships of the same size), and affection status of individuals, including covariates, was then further permuted within the sibship. The proportion of permutations in which the maximum of  $\chi^2_{CC-WI}$  exceeded the value observed for a given SNP was taken as the experiment-wide *P* value for that SNP. To compare the global distribution of test statistics with that expected under the null hypothesis of no association with any marker, the signed Kolmogorov-Smirnov deviation (*d*) was calculated for the observed distribution in comparison with the null distribution from each permutation (30). This statistic is the maximum deviation, over the full distribution of *P* values, of the observed cumulative distribution function (*f*) for any given observed *P* value [*f*(*P*)<sub>obs</sub>] from the value expected under the null [*f*(*P*)<sub>null</sub>]:

$$d = i \times \{\max[|f(P)_{obs} - f(P)_{null}|\},$$

where *i* is an indicator variable that is 1 if *f*(*P*)<sub>obs</sub>  $\geq$  *f*(*P*)<sub>null</sub> and –1 if *f*(*P*)<sub>obs</sub> < *f*(*P*)<sub>null</sub> at the point of maximum deviation. Under the global null hypothesis,



**FIG. 1.** Cumulative distribution of  $P$  values observed for all 80,044 SNPs in the genome-wide association study compared with the expected distribution under the null hypothesis of no association with any marker. *A:* Entire distribution. *B:* Distribution below  $P = 0.05$ .

the expected value of  $d$  is 0; thus, the test that the mean  $d = 0$  across all permutations provides an empirical test of the global null hypothesis. Values of  $d > 0$  indicate a shift toward lower  $P$  values in the observed compared with the null distribution. Since this analysis assumes exchangeability among men and women, it was restricted to the 78,568 autosomal SNPs.

**Analysis of population study.** SNPs that had the strongest associations in the genome-wide study were genotyped and analyzed in the population sample to examine the association on a population basis. To assess replication in a separate group of individuals, the association of genotype with prevalence of diabetes at the last available examination was analyzed among the 2,834 individuals not included in the genome-wide association study. These analyses were conducted using logistic regression models fit with generalized estimating equations to account for correlation among siblings (31). Within-family tests of association were also conducted using a modification of the method of Abecasis et al. (32), which partitions the association into between- and within-family components. In this method, the sibship mean of the numeric variable representing genotype is used to assess the between-family effects, and each individual's deviation from this mean is used to assess within-family effects. The  $P$  value for this within-family coefficient ( $P_{\text{wI}}$ ) was further combined with the general association result using the modification of Fisher's method described above to produce a summary test of these two effects, with the difference that  $P_{\text{wI}}$  was calculated in a one-sided fashion to ensure that claims of replication would only be made if the direction of association was the same as that observed in the genome-wide association study. The distribution-corrected  $P$  value was calculated as described above by simulation of 50 million pairs of test statistics from a bivariate distribution in which the correlation was 0.52.

Analyses of the potential importance of these SNPs in the population were conducted among the 3,501 individuals from the population study. The hazard rate ratio (HRR) for diabetes associated with each copy of the marker allele was estimated. In this model, individuals who developed diabetes were considered to have developed the disease at the age of onset observed in the longitudinal study, while nondiabetic individuals were considered to be at risk for diabetes until the age at last examination. To account for familial resemblance among siblings, these analyses were conducted with a generalized estimating equations model in which diabetes incidence rates were represented as a Poisson function. The HRR was used to calculate population-attributable risk (PAR) for each marker (33). Where  $P_{\text{LL}}$ ,  $P_{\text{LH}}$ , and  $P_{\text{HH}}$  represent, respectively, the proportion of individuals homozygous for the low-risk genotype and heterozygous and homozygous for the high-risk genotype, the PAR, under a multiplicative model, is as follows:

$$\text{PAR} = 1 - 1/(P_{\text{LL}} + P_{\text{LH}}\text{HRR} + P_{\text{HH}}\text{HRR}^2).$$

The PAR represents the proportion by which diabetes incidence would decrease if all individuals had the same risk as individuals with the low-risk genotype.

## RESULTS

The distribution of  $P$  values for all 80,044 SNPs is shown in Fig. 1, along with the distribution expected under the global null hypothesis. Overall, the observed distribution of  $P$  values was similar to that expected under the null hypothesis; however, there was a slight excess of low  $P$  values beyond that expected, and this is more apparent if the portion of the distribution at  $P < 0.05$  is examined (Fig. 1B). The average Kolmogorov-Smirnov deviation comparing the observed and expected distributions in 200 permutations was significantly greater than 0 (average  $d = 0.012$ , SE 0.00036, and  $P < 0.0001$ ). This indicates that the observed distribution contains a statistically significantly greater proportion of low  $P$  values than would be expected under the global null hypothesis of no association with any SNP. Similar results were obtained with the within-family test alone ( $d = 0.006$ , SE 0.00035, and  $P < 0.0001$ ). To explore genotyping artifact as a potential source of bias, analyses were repeated with more stringent thresholds for Hardy-Weinberg equilibrium, genotype success rate, and minor allele frequency with similar results. For the 25,915 SNPs with  $P > 0.05$  for Hardy-Weinberg equilibrium, genotype success rate  $> 0.99$ , and minor allele frequency  $> 0.1$ , results were very similar (e.g., for the within-family test,  $d = 0.008$ , SE 0.00058, and  $P < 0.0001$ ).

Results of tests for individual SNPs for association with young-onset type 2 diabetes are shown in Fig. 2. There were several regions where  $\geq 1$  SNP showed fairly strong evidence for association. The SNPs with the lowest  $P$  values were rs686989 ( $P = 2.7 \times 10^{-6}$ , which corresponds with an experiment-wide  $P$  value of 0.11) and rs672849 ( $P = 1.5 \times 10^{-5}$ , experiment-wide  $P = 0.55$ ), both on chromosome 11 at 113.54 Mb; rs2164000 on chromosome 9 at 18.75 Mb ( $P = 2.3 \times 10^{-5}$ , experiment-wide  $P = 0.69$ ); and rs10520926 on chromosome 5 at 25.36 Mb ( $P = 2.6 \times 10^{-5}$ , experiment-wide  $P = 0.73$ ). The two chromosome 11 SNPs were highly concordant ( $r^2 = 0.99$ ). There were 128 SNPs with  $P < 0.001$  ( $\sim 80$  expected under the null), and these SNPs were further genotyped in individuals from the



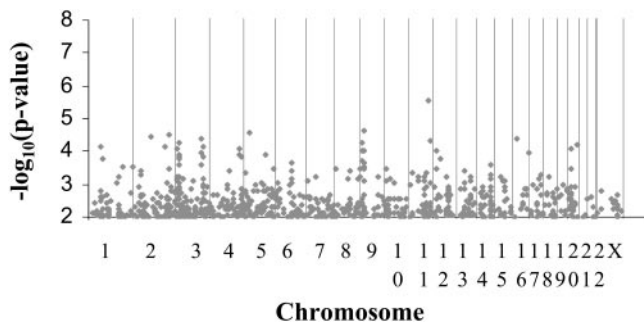


FIG. 2. *P* values for association with diabetes across all chromosomes. The *x*-axis represents the position of SNPs on each chromosome. The *y*-axis is the negative of the base 10 logarithm of the *P* value (higher values represent greater statistical significance). For ease of presentation, only SNPs with *P* < 0.01 are shown.

larger population. The 128 SNPs are listed in supplementary Table 1 (available in an online appendix at <http://dx.doi.org/10.2337/db07-0462>).

There were 11 of these SNPs (out of 119 successfully genotyped) that also showed evidence for association (*P* < 0.05) with diabetes in individuals from the population who were not included in the genome-wide study and for whom the direction of association was the same as that in the genome-wide study. One would expect approximately six such replications by chance at this level of significance. Results of analyses for these SNPs are shown in Table 2. The chromosome 11 SNPs at 113.54 Mb (rs686989 and rs672849) and the chromosome 5 SNP at 25.36 Mb (rs10520926) were among those replicated. In general, the HRR in the population study was lower than the OR from the original study; this is expected given the selection procedure and the fact that the OR only approximates the HRR under certain assumptions (e.g., where incident cases are sampled or the disease is rare [34]). The HRR for most alleles is modest, and those with high PAR are largely those for which the risk alleles are at high frequency.

To assess the extent to which SNPs identified as associated with type 2 diabetes in the present study replicated in other populations, data from three other studies conducted with the Affymetrix 100K array were analyzed for all 646 SNPs with *P* < 0.007 in the present study. These data were provided by investigators from the Amish Family Diabetes Study (35), the Framingham Heart Study (36),

and the Starr County Diabetes Study of Mexican-Americans (37). *P* values for an association in the same direction observed in the present study were combined across all three other studies by Fisher's method (28). Of the SNPs, 88.2% had valid data for all three other studies, 6.0% for two other studies, 5.6% for one other study, and 0.2% (one SNP) for none. There were 30 SNPs, shown in Table 3, that replicated the results of the present study at *P* < 0.05 (~32 expected under the null). None of these overlap with those shown in Table 2. The overall analysis strategy and the number of SNPs selected at various stages of analysis are summarized in Fig. 3.

DISCUSSION

Type 2 diabetes is partially genetically determined (1–6), but susceptibility variants conclusively identified to date explain only a small portion of the familial risk. It is therefore likely that there are additional susceptibility variants that have not yet been identified. Genome-wide association studies are a potentially powerful way to detect these variants. The present analysis represents a genome-wide association study in Pima Indians, a population with a high prevalence of type 2 diabetes and obesity (3,19). Diabetes in this population is highly familial, particularly when it occurs at young ages (3,17–18,38); therefore, the present study was designed to detect variants associated with young-onset diabetes. Although case-control and related association tests can be powerful, they are liable to confounding by population stratification (39,40). In contrast, within-family tests, though less powerful, are robust to such confounding (40). In the present study, both case-control and family-based designs were employed and results combined so that the case-control results augment the power of the within-family test while maintaining robustness to population stratification (29). Thus, markers identified with this approach are likely to reflect both association and linkage with diabetes susceptibility variants.

Many of the associations observed in the present study have *P* values that are quite strong by conventional criteria. However, the interpretation of *P* values in genetic association studies is problematic. Because the prior probability that any given variant is in linkage disequilibrium with a disease susceptibility allele is low, many statisticians recommend very stringent criteria for declaration of

TABLE 2 Association of SNPs detected in the genome-wide association study with evidence for replicated association in additional individuals

SNP	Chr.	Mb	Allele	Frequency*	Genome-wide association group		Replication group		Full population group	
					OR†	<i>P</i>	OR†	<i>P</i>	HRR‡	PAR§
rs10493685	1	81.33	C/T	0.98	∞	1.6×10 <sup>-4</sup>	3.65	0.0043	1.93	—
rs6994019	8	89.34	C/A	0.98	∞	3.8×10 <sup>-4</sup>	3.49	0.0189	1.68	—
rs1500415	3	20.80	A/G	0.98	∞	8.6×10 <sup>-4</sup>	3.69	0.0489	1.63	—
rs1859441	12	46.71	T/C	0.88	2.72	1.7×10 <sup>-4</sup>	1.86	0.0027	1.17	0.25
rs424695	3	21.47	G/A	0.88	2.33	2.6×10 <sup>-4</sup>	1.63	0.0149	1.15	0.23
rs10520926	5	25.36	T/A	0.55	2.17	2.6×10 <sup>-5</sup>	1.36	0.0335	1.07	0.07
rs10500938	11	22.60	A/G	0.20	2.14	4.6×10 <sup>-4</sup>	1.65	0.0011	1.14	0.05
rs672849	11	113.54	A/G	0.19	2.84	1.5×10 <sup>-5</sup>	1.34	0.0164	1.14	0.05
rs686989	11	113.54	A/G	0.19	3.26	2.7×10 <sup>-6</sup>	1.27	0.0333	1.13	0.05
rs9290075	3	162.50	A/C	0.25	2.47	7.0×10 <sup>-5</sup>	1.39	0.0134	1.10	0.04
rs10506855	12	80.88	T/G	0.02	6.70	5.7×10 <sup>-4</sup>	1.79	0.0335	1.36	0.02

Data are listed in order of descending PAR. \*Frequency of the allele listed first, which gave OR >1. †OR per copy of the allele listed first in the within-family analysis. ‡HRR per copy of the allele listed first. §Not shown for SNPs for which the number of low-risk homozygotes was too small for reliable estimation. Chr., chromosome.

TABLE 3

SNPs with evidence for replicated association with type 2 diabetes (combined  $P$  value  $<0.05$ ) in Amish, Framingham, and Starr County studies

SNP	Chr.	Mb	Allele	Present study		Amish		Framingham		StarrCounty		Combined
				OR*	$P$	OR*	$P$ †	HRR‡	$P$ §	OR*	$P$ †	$P$ ¶
rs1775368	1	72.70	T/A	5.15	$6.7 \times 10^{-5}$	0.84	0.9356	1.20	0.1804	2.85	0.0078	0.0392
rs1886004	1	165.06	A/G	$\infty$	$8.8 \times 10^{-4}$	2.16	0.0158	1.19	0.2644	1.22	0.4043	0.0468
rs10496191	2	73.59	G/A	1.88	$5.5 \times 10^{-3}$	1.13	0.1484	1.44	0.0329	1.34	0.1706	0.0277
rs10496193	2	73.69	C/T	1.90	$5.3 \times 10^{-3}$	1.14	0.1457	1.39	0.0461	1.44	0.1231	0.0275
rs10497681	2	189.38	A/G	12.97	$6.8 \times 10^{-5}$	0.65	0.9960	0.99	0.5260	4.39	0.0022	0.0358
rs2366760	3	21.15	A/G	1.82	$2.3 \times 10^{-3}$	1.07	0.2630	1.00	0.5020	2.05	0.0061	0.0271
rs10517403	4	37.05	G/A	1.57	$5.0 \times 10^{-3}$	1.33	0.0079	1.13	0.2104	1.36	0.1931	0.0132
rs1072389	4	75.36	T/C	2.02	$2.9 \times 10^{-3}$	1.17	0.0690	0.99	0.5238	1.66	0.035	0.0378
rs460610	4	103.51	A/G	2.02	$3.7 \times 10^{-4}$	1.22	0.0337	1.39	0.0314	1.15	0.3338	0.0143
rs543862	5	177.97	T/C	1.74	$3.2 \times 10^{-3}$	1.15	0.1088	1.27	0.0723	1.45	0.1084	0.0282
rs10488205	7	11.92	C/T	2.20	$2.3 \times 10^{-3}$	NA	NA	NA	NA	1.96	0.0378	0.0378
rs1486171	7	45.98	A/T	1.80	$5.5 \times 10^{-3}$	1.18	0.0633	1.00	0.5067	1.69	0.0354	0.0349
rs3847109	7	130.56	A/G	1.88	$2.2 \times 10^{-3}$	NA	NA	1.91	0.0146	1.35	0.2321	0.0227
rs3780825	9	132.19	A/G	4.92	$3.3 \times 10^{-3}$	0.99	0.5110	2.19	0.0007	1.08	0.5000	0.0084
rs2130154	11	86.50	C/T	2.34	$2.1 \times 10^{-3}$	NA	NA	1.38	0.0320	1.48	0.1078	0.0230
rs2048973	11	86.52	T/G	2.11	$3.8 \times 10^{-3}$	1.19	0.0507	1.59	0.0012	1.32	0.1739	0.0008
rs586421	11	94.50	C/T	1.65	$4.0 \times 10^{-3}$	1.28	0.0136	1.00	0.4907	1.39	0.1321	0.0290
rs1789369	11	109.50	T/C	1.91	$4.6 \times 10^{-3}$	0.94	0.7015	1.01	0.4874	2.51	0.0029	0.0312
rs516415	11	121.56	A/G	2.63	$5.0 \times 10^{-5}$	1.26	0.0179	0.90	0.7668	1.82	0.0205	0.0120
rs7301434	12	91.36	C/A	3.39	$1.5 \times 10^{-3}$	0.94	0.6263	1.35	0.1572	6.17	0.0014	0.0068
rs54939	13	54.45	C/A	1.64	$3.2 \times 10^{-3}$	1.21	0.0360	0.87	0.7497	1.69	0.0485	0.0388
rs1537779	13	81.92	C/T	1.73	$2.2 \times 10^{-3}$	0.94	0.7160	1.50	0.0019	1.53	0.0728	0.0053
rs9285325	13	81.95	C/G	1.74	$3.1 \times 10^{-3}$	0.98	0.5880	1.40	0.0101	1.42	0.1231	0.0251
rs1959225	14	40.11	A/T	1.95	$1.2 \times 10^{-3}$	0.98	0.5885	1.41	0.0238	1.85	0.0197	0.0118
rs10498547	14	80.19	C/A	2.57	$6.6 \times 10^{-4}$	1.17	0.0901	1.07	0.3591	1.86	0.0279	0.0294
rs958435	20	6.52	G/C	2.76	$5.7 \times 10^{-3}$	1.28	0.0341	1.44	0.0208	0.79	0.7598	0.0198
rs6140410	20	7.69	A/G	2.98	$6.0 \times 10^{-3}$	0.89	0.8319	2.33	0.0015	1.05	0.4663	0.0211
rs1223271	20	13.24	G/A	2.21	$5.3 \times 10^{-3}$	1.60	0.0039	1.20	0.2221	0.49	0.9546	0.0273
rs477184	20	13.25	A/G	2.32	$1.2 \times 10^{-3}$	1.00	0.5092	1.55	0.0045	1.00	0.5000	0.0351
rs6030447	20	40.79	C/T	2.53	$3.6 \times 10^{-3}$	1.16	0.0821	0.91	0.7231	2.00	0.0144	0.0282

\*OR per copy of the allele listed first (which gave OR  $>1$  in the present study); in the present study, this result is for the within-family analysis. †One-sided  $P$  for association in the direction observed in the present study. ‡HRR per copy of the allele listed first. § $P$  from the generalized estimating equation analysis in the Framingham Heart Study (ref. 36). ¶ $P$  combined from all three populations (except the present study) by Fisher's method. Chr., chromosome; NA, not analyzed.

statistical significance, e.g.,  $P$  values from  $10^{-5}$  to  $10^{-8}$  (41–43). The problem is compounded in genome-wide studies, where because of multiple tests, one would expect some strong associations to occur by chance. Classical methods of adjustment for multiple comparisons, such as Bonferroni, are too stringent for genome-wide association studies because they assume that tests are independent and ignore linkage disequilibrium among markers. Furthermore, while corrections for multiple testing proceed on the assumption that the global null hypothesis of no association with any marker is of interest (44), the corrections are applied on a marker-wise basis and, thus, may not efficiently utilize information from the full distribution of  $P$  values relevant to this global null.

In the present analysis, a permutation procedure was used to assess experiment-wide statistical significance. This procedure maintains the linkage disequilibrium structure among markers. In addition, the Kolmogorov-Smirnov test was used to compare the overall distribution of  $P$  values observed in the actual data with that observed in data permuted under the null hypothesis. Although no single marker achieved experiment-wide statistical significance, analysis of the distribution of  $P$  values resulted in strong rejection of the global null hypothesis of no association with any marker. This result is expected if there are multiple susceptibility variants because, in such a situa-

tion, a test of the full distribution can be more powerful than a test of a single variant. A systematic bias is an alternative possibility that is difficult to completely exclude, but the present techniques are robust to bias from population stratification. The present results were also unmodified when restricted to SNPs with increased stringency of Hardy-Weinberg equilibrium, allele frequency, and genotype success rate, and this suggests that they are not explained by bias due to genotypic artifacts that can be captured by these measures. The number of true functional susceptibility variants is difficult to determine, given that modest linkage disequilibrium may extend over long distances and that many associations may be due to chance. Thus, while the present results strongly suggest that some of the low  $P$  values observed reflect linkage disequilibrium with diabetes susceptibility loci, the distinction between true- and false-positive results is difficult.

Replication of the association in a separate group of individuals provides additional evidence that a marker is associated with diabetes. In the present study, SNPs with the strongest evidence for association in the genome-wide study were further evaluated in a separate set of individuals from the population. (However, some of these individuals were related to those in the genome-wide study.) Several SNPs showed nominal evidence for association ( $P < 0.05$ ) in this replication set. This analysis is limited in

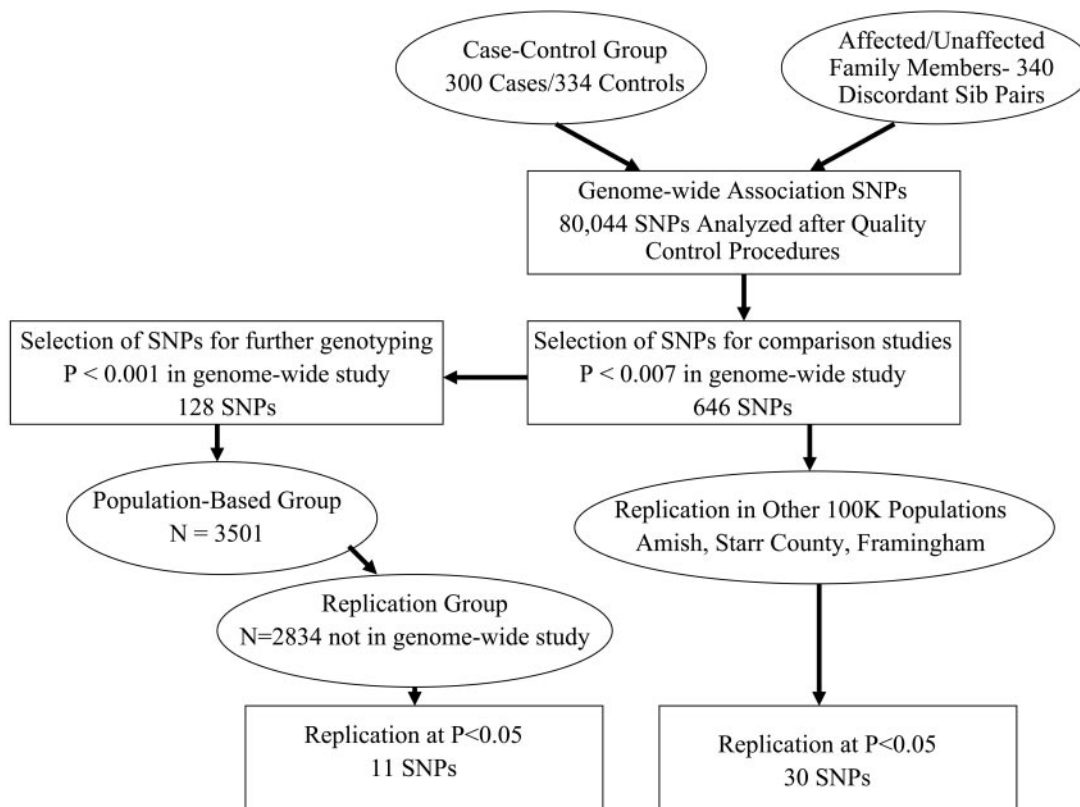


FIG. 3. Diagrammatic representation of the present study, along with molecular and analytic strategies to assess replication.

power because the vast majority of individuals from the extremes of the age-of-onset distribution, who provide much of the statistical power, were excluded due to participation in the genome-wide study. In addition, one would expect some of the markers to show association by chance. Thus, some false-positives are likely to remain among the SNPs that replicated and some true positives among those that did not. However, the probability that SNPs showing replication are in linkage disequilibrium with diabetes susceptibility variants is enhanced above that for the SNPs for which replication was not observed. The result for rs686989 is of particular interest because it is in a region identified as linked to diabetes and obesity in a previous genome-wide linkage study in this population (15); however, the association with rs686989 itself does not explain the linkage signal (data not shown).

The PAR, calculated in a group representative of the full population, provides a measure of the potential importance of each associated SNP, and this information can be used to prioritize regions for follow-up studies. The PAR is calculated on the assumption that the observed association is causal; therefore, it may be underestimated if the marker is not highly concordant with a functional allele or overestimated if confounded by population stratification. In the present analysis, the PAR was calculated from longitudinal data observed in the population study. Thus, in contrast to estimates derived from case-control studies based on prevalent cases, the present results do not depend on the questionable assumptions of a rare disease or of sampling only incident cases (34).

Studies of these SNPs in other populations may also be relevant in prioritization of regions for follow-up studies. The present results were further compared with those obtained from three other genome-wide association stud-

ies of type 2 diabetes on the Affymetrix 100K array (35–37). Comparison among studies is complicated by the fact that all had different study designs; the present study focused on young-onset diabetes and gave priority to within-family tests. (Characteristics of each study are presented in supplementary Tables 2 and 3.) Nonetheless, several of the SNPs identified in the present study had some evidence for association in the other studies (overall  $P < 0.05$ ). These regions are also high priority for follow-up studies. It is noteworthy that rs516415, which is also in the diabetes-linked region on chromosome 11q, showed replication, as did rs1886004, which is in a region on chromosome 1q linked to diabetes in Pima Indians and other populations (14–15). However, none of the SNPs with evidence for replication in the Pima Indians at  $P < 0.05$  also had a combined  $P < 0.05$  in these other studies. This may reflect low power across the studies to detect alleles of modest effect.

The power of genetic association studies depends on the frequency of the functional polymorphisms and on the magnitude of their effects (22). Given that case subjects represent the lower 10% of the age-of-onset distribution and control subjects the upper 45%, we estimate that the present sample size has ~75% power to detect a common (minor allele frequency  $>0.1$ ) functional allele at  $P < 0.001$  that explains 3% of the variance in age at onset of diabetes. Power also depends on the extent to which one of the typed markers is strongly concordant with an allele that influences disease susceptibility (21,22). Since the 100K array used in the present study does not exhaustively capture allelic variation, it is possible that additional regions that contain diabetes susceptibility variants were not identified. The pattern of linkage disequilibrium among SNPs identified in the HapMap project suggests that ~30%



TABLE 4

Strongest association results in present study for SNPs in regions identified as associated with type 2 diabetes in multiple high-density genome-wide association studies

Gene	Sentinel SNP*	Chr.	Mb	SNP in present study†	Mb	Allele	Frequency‡	OR§	P	SNPs (n)¶
PPARG	rs1801281	3	12.38	rs10510422	12.51	T/C	0.92	2.31	0.0384	19
IGF2BP2	rs4402960	3	187.99	rs6763887	187.97	G/A	0.28	1.34	0.0945	3
CDKAL1	rs10946398	6	20.77	rs7758851	20.65	C/T	0.94	1.82	0.2085	15
SLC30A8	rs13266634	8	118.25	rs10505309	118.31	C/T	0.14	2.18	0.0332	28
CDKN2A	rs1081161	9	22.12	rs2025798	22.27	G/A	0.97	1.93	0.0864	12
HHEX	rs1111875	10	94.45	rs2096177	94.50	T/G	0.79	1.63	0.0446	6
TCF7L2	rs7903146	10	114.75	rs10509970	114.91	G/T	0.53	1.30	0.0719	9
KCNJ11	rs5219	11	17.34	rs2190454	17.49	C/T	0.10	1.27	0.4600	4
FTO	rs8050136	16	52.37	rs10521308	52.42	T/C	0.05	1.27	0.3299	13

\*The sentinel SNP is that with the strongest association with diabetes across multiple genome-wide association studies (refs. 47–50). †The SNP with the smallest *P* value within a 200-kb window on either side of the sentinel SNP is reported. ‡Frequency of the allele listed first, which gave OR >1. §OR in the within-family analysis per copy of the allele listed first. ¶SNPs from the present study in the 400-kb window surrounding the sentinel SNP. Chr., chromosome.

of common variants have  $r^2 > 0.8$  with a marker on this array in non-African populations (45). American Indians are not included in the HapMap, and they tend to share relatively few common haplotypes with HapMap populations (46). This may be reflected in the fact that a larger proportion of markers were nearly monomorphic in the present study than in the other studies that used the 100K array (35–37), and this could result in lower power for this array to detect associations in American Indians. On the other hand, surveys of various populations suggest that linkage disequilibrium is higher among American Indians than in many other populations such that fewer variants are required to capture common haplotypes. (46). Thus, fixed marker sets, such as the 100K array, may capture common variation in American Indians to an extent similar to that in other non-African populations.

Recently, several high-density genome-wide association studies have been conducted in northern European populations, and these have identified six gene regions (apart from TCF7L2, PPARG and KCNJ11) that contain putative diabetes susceptibility variants (47–50). None of the SNPs most strongly associated in the present study was in any of these regions. Furthermore, none of the SNPs consistently associated across these other genome-wide studies is well-captured by the 100K array. However, as shown in Table 4, some SNPs in these regions had modest evidence for association in the present study ( $P < 0.05$ ). These putative diabetes susceptibility variants have been largely identified in northern Europeans, and it is not clear whether they play a significant role in diabetes in American Indians or other high-risk populations. A more exhaustive survey of variation in these regions is required to quantify their role in diabetes susceptibility in the Pima Indian population. The variants in TCF7L2 that are most strongly associated with type 2 diabetes in other populations (9–13) have been typed in the Pima Indians, in whom there is little evidence that they influence diabetes risk (51).

Genome-wide mapping studies are typically only an initial step in the elucidation of susceptibility variants. The present analyses have identified several regions that may harbor genetic variants that influence susceptibility to young-onset diabetes in American Indians. Several of the associated SNPs are in or near genes, including ZNF659 (chromosome 3, 21.47 Mb), FANCF (chromosome 11, 22.60 Mb), ZBTB15 (chromosome 11, 113.54 Mb), and

SENPI (chromosome 12, 46.71 Mb). Fine-mapping studies of these regions are needed to confidently localize the signals to specific genes. Confirmation of the role of genes in the regions identified in the present study will require replication studies in other populations and, ultimately, functional studies.

#### ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

We thank Jose Florez for review of the manuscript. We also thank the staff of the Diabetes Epidemiology and Clinical Research Section and the Diabetes Molecular Genetics Section (NIDDK, Phoenix, Arizona) for assistance. We further thank members of the Gila River Indian Community who participated in the study.

#### REFERENCES

- Rich SS: Mapping genes in diabetes: genetic epidemiological perspective. *Diabetes* 39:1315–1319, 1990
- Hanson RL, Knowler WC: Type 2 diabetes and maturity-onset diabetes of the young. In *Analysis of Multifactorial Disease*. Bishop T, Sham P, Eds. Oxford, U.K., BIOS Scientific Publishers, 2000, p. 131–147
- Knowler WC, Pettitt DJ, Saad MF, Bennett PH: Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. *Diabetes Metab Rev* 6:1–27, 1990
- Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD: Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia* 30:763–768, 1987
- Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengård J, Kesäniemi YA: Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35:1060–1067, 1992
- Matsuda A, Kuzuya T: Diabetic twins in Japan. *Diabetes Res Clin Pract* 24 (Suppl.):S63–S67, 1994
- Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPAR $\gamma$  Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80, 2000
- Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM: Large-scale association studies of variants in genes encoding the pancreatic  $\beta$ -cell  $K_{ATP}$  channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* 52:568–572, 2003
- Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A, Styrkars-

- dottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K: Variant of transcription factor 7-like 2 (*TFC7L2*) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323, 2006
10. Zeggini E, McCarthy MI: TCF7L2: the biggest story in diabetes genetics since HLA? *Diabetologia* 50:1–4, 2007
  11. Helgason A, Pálsson S, Thorleifsson G, Grant SFA, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, Benediktsson R, Hinney A, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Schäfer H, Faruque M, Doumatey A, Zhou J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Sigurdsson G, Hebebrand J, Pedersen O, Thorsteinsdottir U, Gulcher JR, Kong A, Rotimi C, Stefansson K: Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 39:218–225, 2007
  12. Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PIW, Shuldiner AR, Knowler WC, Nathan DM, Altschuler D, the Diabetes Prevention Program Research Group: *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250, 2006
  13. Chandak GR, Janipalli CS, Bhaskar S, Kulkarni SR, Mohankrishna P, Hattersley AT, Frayling TM, Yajnik CS: Common variants in the *TCF7L2* gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* 50:63–67, 2007
  14. McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159–167, 2003
  15. 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
  16. Dabelea D, Palmer JP, Bennett PH, Pettitt DJ, Knowler WC: Absence of glutamic acid decarboxylase antibodies in Pima Indian children with diabetes. *Diabetologia* 42:1265–1266, 1999
  17. Hanson RL, Knowler WC: Analytic strategies to detect linkage to a common disorder with genetically determined age of onset: diabetes mellitus in Pima Indians. *Genet Epidemiol* 15:299–315, 1998
  18. Hanson RL, Elston RC, Pettitt DJ, Bennett PH, Knowler WC: Segregation analysis of non-insulin-dependent diabetes mellitus in Pima Indians: evidence for a major-gene effect. *Am J Hum Genet* 57:160–170, 1995
  19. 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* 108:497–505, 1978
  20. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183–1197, 1997
  21. Zondervan KT, Cardon LR: The complex interplay among factors that influence allelic association. *Nat Rev Genet* 5:89–100, 2004
  22. Hanson RL, Looker HC, Ma L, Muller YL, Baier LJ, Knowler WC: Design and analysis of genetic association studies to finely map a locus identified by linkage analysis: sample size and power calculations. *Ann Intern Med* 70:332–349, 2006
  23. 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](http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). Accessed 25 March 2007
  24. Emigh TH: A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* 36:627–642, 1980
  25. Schnell AH, Karunaratne PM, Witte JS, Dawson DV, Elston RC: Modeling age of onset and residual familial correlations for the linkage analysis of bipolar disorder. *Genet Epidemiol* 14:675–680, 1997
  26. Bonney GE: Regressive logistic models for familial disease and other binary traits. *Biometrics* 42:611–625, 1986
  27. Witte JS, Gauderman WJ, Thomas DC: Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 149:693–705, 1999
  28. Elston RC: On Fisher's method of combining p-values. *Biometrical J* 33:339–345, 1991
  29. Hanson RL, Knowler WC: Design and analysis of genetic association studies to finely map a locus identified by linkage analysis: assessment of the extent to which an association can account for the linkage. *Ann Intern Med*. 12 July 2007 [Epub ahead of print]
  30. Massey FJ: The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 253:68–78, 1951
  31. Zeger SL, Liang KY: Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42:121–130, 1986
  32. Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292, 2000
  33. Kleinbaum DG, Kupper LL, Morgenstern H: Measures of potential impact and summary of the measures. In *Epidemiologic Research: Principles and Quantitative Methods*. Kleinbaum DG, Kupper LL, Morgenstern H, Eds. New York, Van Nostrand Reinhold Company, 1982, p. 159–180
  34. Walter WD: The estimation and interpretation of attributable risk in health research. *Biometrics* 32:829–849, 1976
  35. Rimpersaud E, Dancott DM, O'Connell J, McArdle P, Shen H, Fu M, Shelton J, Ying J, Shi X, Ott SH, Zhang L, Zhao Y, Mitchell BD, Shuldiner AR: Identification of novel candidate genes in the Old Order Amish with replication in independent genome-wide association scans (GWAS) of type 2 diabetes. *Diabetes* 56:3053–3062, 2007
  36. Florez JC, Manning MK, Dupuis J, McAteer J, Irenze K, Gianniny L, Mirel DB, Fox CS, Cupples LA, 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. *Diabetes* 56:3063–3074, 2007
  37. Hayes GM, Pluzhnikov A, Miyake K, Sun Y, Below JE, Ng MCY, 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. *Diabetes* 56:3033–3044, 2007
  38. Baier LJ, Hanson RL: Genetic studies of the etiology of type 2 diabetes in Pima Indians: hunting for pieces to a complicated puzzle. *Diabetes* 53:1181–1186, 2004
  39. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG: Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 43:520–526, 1988
  40. Cardon LR, Bell JI: Association study designs for complex diseases. *Nat Rev Genet* 2:91–99, 2001
  41. Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 273:1516–1517, 1996
  42. Colhoun HM, McKeigue PM, Davey Smith G: Problems of reporting genetic associations with complex outcomes. *Lancet* 361:865–872, 2003
  43. Manly KF: Reliability of statistical associations between genes and disease. *Immunogenetics* 57:549–558, 2005
  44. Rothman KJ: No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–46, 1990
  45. Pe'er I, de Bakker PIW, Maller J, Yelensky R, Altschuler D, Daly MJ: Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 38:663–667, 2006
  46. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK: A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38:1251–1260, 2006
  47. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885, 2007
  48. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, Chen H, Roix JR, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altschuler D: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1335, 2007
  49. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JRB, 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 ASF, Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
  50. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines CS, 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
  51. Guo T, Hanson RL, Traurig M, Muller YL, Ma L, Mack J, Kobes S, Knowler WC, Bogardus C, Baier LJ: TCF7L2 is not a major susceptibility gene for type 2 diabetes in Pima Indians: analysis of 3,501 individuals *Diabetes* 56:3075–3088, 2007