Follow-Up Analysis of Genome-Wide Association Data Identifies Novel Loci for Type 1 Diabetes

Struan F.A. Grant,[1,2,3] Hui-Qi Qu,[4] Jonathan P. Bradfield,[1] Luc Marchand,[4] Cecilia E. Kim,[1] Joseph T. Glessner,[1] Rosemarie Grabs,[4] Shayne P. Taback,[5] Edward C. Frackelton,[1] Andrew W. Eckert,[1] Kiran Annaiah,[1] Margaret L. Lawson,[6] F. George Otieno,[1] Erin Santa,[1] Julie L. Shaner,[1] Ryan M. Smith,[1] Robert Skraban,[1] Marcin Imielinski,[1] Rosetta M. Chiavacci,[1] Robert W. Grundmeier,[7,8] Charles A. Stanley,[9] Susan E. Kirsch,[10] Daryl Waggott,[11] Andrew D. Paterson,[12] Dimitri S. Monos,[3,13] the DCCT/EDIC Research Group,* Constantin Polychronakos,[4] and Hakon Hakonarson[1,2,3]

OBJECTIVE—Two recent genome-wide association (GWA) studies have revealed novel loci for type 1 diabetes, a common multifactorial disease with a strong genetic component. To fully utilize the GWA data that we had obtained by genotyping 563 type 1 diabetes probands and 1,146 control subjects, as well as 483 case subject–parent trios, using the Illumina HumanHap550 BeadChip, we designed a full stage 2 study to capture other possible association signals.

RESEARCH DESIGN AND METHODS—From our existing datasets, we selected 982 markers with $P < 0.05$ in both GWA cohorts. Genotyping these in an independent set of 636 nuclear families with 974 affected offspring revealed 75 markers that also had $P < 0.05$ in this third cohort. Among these, six single nucleotide polymorphisms in five novel loci also had $P < 0.05$ in the Wellcome Trust Case-Control Consortium dataset and were further tested in 1,303 type 1 diabetes probands from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) plus 1,673 control subjects.

RESULTS—Two markers (rs9976767 and rs3757247) remained significant after adjusting for the number of tests in this last cohort; they reside in *UBASH3A* (OR 1.16; combined $P = 2.33 \times 10^{-8}$) and *BACH2* (1.13; combined $P = 1.25 \times 10^{-6}$).

CONCLUSIONS—Evaluation of a large number of statistical GWA candidates in several independent cohorts has revealed additional loci that are associated with type 1 diabetes. The two genes at these respective loci, *UBASH3A* and *BACH2*, are both biologically relevant to autoimmunity. ***Diabetes*** **58:290–295, 2009**

Type 1 diabetes is a multifactorial disease with a strong genetic component that results from autoimmune destruction of the pancreatic β-cells. The major type 1 diabetes susceptibility locus, mapping to the *HLA* class II genes at 6p21 (1) and encoding highly polymorphic antigen-presenting proteins, accounts for almost 50% of the genetic risk for type 1 diabetes (2). Several other loci with more modest effects are known, but they do not account for the remaining portion of the risk.

The recent development of high-throughput single nucleotide polymorphism (SNP) genotyping array technologies has enabled us (3) and others (4) to perform high-density genome-wide association (GWA) studies in search of the remaining type 1 diabetes loci. We recently reported the outcome of our GWA for type 1 diabetes in a large pediatric type 1 diabetic cohort of European descent (3); in addition to confirming previously identified loci, we observed highly significant and replicated association with *KIAA0350* (now renamed *CLEC16A* [C-type lectin domain family 16 member A]). Subsequent follow-up of our data also revealed a locus on 12q13 (5). In parallel and independently, the Wellcome Trust Case Control Consortium (WTCCC) (4) also demonstrated replicated (6) association to the same linkage disequlibrium blocks at 16p13 and 12q13, along with two additional loci on 12q24 and 18p11.

The results that we have reported thus far were of loci that achieved statistical significance on the basis of the results of the GWA genotyping (stage 1) or replication in additional cohorts (stage 2) of only a small number of the most promising loci. Here, we describe the results of a full evaluation of all statistical candidates from the GWA phase.

From the [1]Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the [2]Division of Human Genetics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the [3]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; the [4]Departments of Pediatrics and Human Genetics, McGill University, Montreal, Quebec, Canada; the [5]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada; the [6]Division of Endocrinology, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada; the [7]Pediatric Research Consortium, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the [8]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; the [9]Division of Endocrinology, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the [10]Markham-Stouffville Hospital, Markham, Ontario, Canada; the [11]Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Prosserman Center for Health Research, Toronto, Ontario, Canada; the [12]Department of Public Health Sciences, Hospital for Sick Kids, University of Toronto, Toronto, Ontario; and the [13]Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.

Corresponding author: Hakon Hakonarson, hakonarson@chop.edu.

Received 28 July 2008 and accepted 25 September 2008.

Published ahead of print at http://diabetes.diabetesjournals.org on 7 October 2008. DOI: 10.2337/db08-1022.

*A full list of members of the DCCT/EDIC Research Group can be found in the following article: *N Engl J Med* 356:1842–1852, 2007.

S.F.A.G. and H.-Q.Q. contributed equally to this study.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health.

TABLE 1
Flow process

	Sample information	Genotyping	Statistics*
Stage 1 (GWA)			
Case-control cohort	563 type 1 diabetes case subjects (mean age at diagnosis 7.6 years [range 0.1–18]) and 1,146 control subjects	550,000 SNPs on the Illumina HumanHap550	982 SNPs outside the MHC region with $P <$ 0.05 in the same direction in both cohorts
Family cohort	483 type 1 diabetes case-parent trios (mean case subject age of diagnosis 8.1 years [range 0.5–18])	550,000 SNPs on the Illumina HumanHap550	
Stage 2			
Family cohort	636 nuclear type 1 diabetic families (mean age at diagnosis of type 1 diabetes case subjects 9.6 years [range 0.1–37])	982 SNPs on the Illumina GoldenGate	75 SNPs with $P < 0.05$ in the same direction as the GWA stage
Additional evidence: WTCCC dataset (ref. 4)	2,000 type 1 diabetes case and 3,000 control subjects (age of type 1 diabetes diagnosis <17 years and insulin dependence since diagnosis, with a minimum period of at least 6 months)	500,000 SNPs on the Affymetrix GeneChip	33 SNPs with $P < 0.05$ in the same direction of the above cohorts; 6 SNPs in five loci were from previously unreported loci
Validation: case-control cohort†	1,303 DCCT/EDIC type 1 diabetes case subjects (mean age at diagnosis 21 years [range 0–38]) and 1,673 control subjects	1 million SNPs on the Illumina 1M assay; 550,000 SNPs on the Illumina HumanHap550	Two of the six SNPs with $P \leq 0.01$ in the same direction of the above cohorts

*Two of the six SNPs, rs10758593 and rs10758594, are in tight linkage disequilibrium ($r^2 = 0.86$). Therefore, five independent hypothesis were tested, and the corrected significant level for multiple comparisons is $\alpha = 0.01$. All P values are two sided. †Six SNPs in five loci were tested in the validation cohort. MHC, major histocompatibility complex.

RESEARCH DESIGN AND METHODS

Study populations

Type 1 diabetes cohort from Canada. The Canadian cohort consisted of 1,120 nuclear family trios (one affected child and two parents) and 267 independent type 1 diabetes cases, collected in pediatric diabetes clinics in Montreal, Toronto, Ottawa, and Winnipeg. The median age at onset is 8 years with lower and upper quartiles at 4.6 and 11 years, respectively. All patients were diagnosed under the age of 18 years and treated with insulin since diagnosis, and none have stopped treatment for any reason since. Disease diagnosis was based on these clinical criteria rather than any laboratory tests. Ethnic backgrounds were of mixed European descent, with the largest single subset (409 families) being French Canadian. The Research Ethics Board of the Montreal Children's Hospital and other participating centers approved the study, and written informed consent was obtained from all subjects.

Type 1 Diabetes Genetics Consortium cohort. The Type 1 Diabetes Genetics Consortium cohort consisted of 549 families (2,350 individuals) with at least two children diagnosed with diabetes and both parents available as of the July 2005 data freeze. Criteria were age at diagnosis below 35 years and uninterrupted treatment with insulin within 6 months of diagnosis. For siblings of probands diagnosed under the age of 35 years, the age-at-diagnosis limit was extended to 45 years if they were lean and had positive islet cell antibodies and/or low C-peptide levels at diagnosis. The median age is 8 years with quartiles at 4 and 13 years. The samples were collected in Europe, North America, and Australia.

Type 1 diabetes cohort from Philadelphia. The type 1 diabetes cohort consisted of 103 children recruited at the Children's Hospital of Philadelphia (CHOP) since September 2006, as previously described (3).

Diabetes Control and Complications Trial/Epidemiology of Diabetes Complications and Interventions type 1 diabetes cohort. The Diabetes Control and Complications Trial was a multicenter randomized clinical trial to determine the effect of intensive insulin treatment with respect to reduced development and progression of retinopathy and nephropathy complications in patients with type 1 diabetes (7,8). A total of 1,441 subjects with type 1 diabetes were recruited from 29 centers across North America into the DCCT between 1983 and 1989; they were between 13 and 39 years of age, and 53% were male. They were recruited into two cohorts: the primary prevention cohort consisted of 726 subjects with no retinopathy, an albumin excretion rate <28 μgmin, and diabetes duration of 1–5 years and was studied to determine whether intensive therapy prevented the development of diabetic retinopathy in patients with no retinopathy. The secondary intervention cohort consisted of 715 subjects who had nonproliferative retinopathy, a urinary albumin excretion rate <140 μg/min, and diabetes duration of 1–15 years and was studied to determine whether intensive therapy would affect the progression of early retinopathy (7). Approval for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Complications and Interventions (DCCT/EDIC) genetics study was provided by the Research Ethics Board of the Hospital for Sick Children, Toronto.

The Illumina 1M assay was genotyped on all available probands. To detect and remove outliers due to population stratification from the majority of self-reported white probands, Eigenstrat (9) was used to select probands by sequential analysis. After exclusions of outliers, there were 1,303 DCCT/EDIC probands (695 male and 608 female), with mean ± SD age of type 1 diabetes diagnosis 21 ± 8 years (range 0–38).

Control subjects from Philadelphia. The control group used to match with the DCCT/EDIC cases included 2,024 children with self-reported Caucasian

TABLE 2
Cohort datasets leveraged in selection of candidate loci for further validation efforts (data sorted by chromosomal position)

			Type 1 diabetic families:Montreal and T1DGC			Case-control cohort			
Chr	SNP	Position	Trans: untrans	OR	TDT *P*	Aff allele frequency	Ctrl allele frequency	*P*	OR
1	**rs1983853**	**85,083,780**	**202:254**	**0.8**	**0.015**	**0.121**	**0.151**	**0.021**	**0.779**
1	rs1230661	113,987,113	456:331	1.378	8.36×10^{-6}	0.267	0.216	8.75×10^{-4}	1.324
1	rs4839335	114,035,394	482:354	1.362	9.56×10^{-6}	0.3	0.25	0.002	1.285
1	rs1217407	114,195,271	505:395	1.278	2.46×10^{-4}	0.298	0.244	7.42×10^{-4}	1.316
1	rs12566340	114,221,851	492:379	1.298	1.29×10^{-4}	0.288	0.237	0.0015	1.299
1	rs7529353	114,221,985	474:354	1.339	3.04×10^{-5}	0.294	0.242	9.75×10^{-4}	1.309
1	rs2358994	114,230,984	398:287	1.387	2.22×10^{-5}	0.232	0.175	7.11×10^{-5}	1.426
1	rs7520320	114,336,816	211:169	1.249	0.031	0.136	0.107	0.013	1.315
1	rs12029644	114,338,303	222:172	1.291	0.012	0.136	0.105	0.008	1.343
2	rs2111485	162,818,782	417:500	0.834	0.0061	0.393	0.433	0.027	0.849
2	rs1990760	162,832,297	422:518	0.815	0.0017	0.398	0.434	0.048	0.864
2	rs1024161	204,429,997	534:462	1.156	0.023	0.439	0.397	0.02	1.191
2	rs926169	204,430,997	521:459	1.135	0.048	0.426	0.38	0.0094	1.212
2	rs231726	204,449,111	502:427	1.176	0.014	0.358	0.321	0.028	1.184
6	**rs3757247**	**91,014,184**	**545:482**	**1.13**	**0.049**	**0.504**	**0.455**	**0.0075**	**1.216**
9	**rs10758593**	**4,282,083**	**539:462**	**1.17**	**0.015**	**0.492**	**0.426**	$\mathbf{2.97 \times 10^{-4}}$	**1.303**
9	**rs10758594**	**4,285,583**	**535:456**	**1.17**	**0.012**	**0.513**	**0.451**	$\mathbf{6.66 \times 10^{-4}}$	**1.282**
10	rs706779	6,138,830	424:506	0.838	0.0072	0.425	0.492	2.60×10^{-4}	0.764
10	rs3118470	6,141,719	504:431	1.169	0.017	0.365	0.306	4.62×10^{-4}	1.308
10	rs7072793	6,146,272	574:507	1.132	0.042	0.486	0.41	2.96×10^{-5}	1.358
10	rs7073236	6,146,558	566:464	1.22	0.0015	0.487	0.414	5.67×10^{-5}	1.343
11	rs1004446	2,126,719	378:514	0.735	5.27×10^{-6}	0.254	0.354	4.38×10^{-9}	0.622
12	rs11171710	54,654,345	464:574	0.808	6.40×10^{-4}	0.405	0.462	0.0016	0.792
12	rs10876864	54,687,352	631:528	1.195	0.0025	0.458	0.388	8.39×10^{-5}	1.336
12	rs1701704	54,698,754	549:425	1.292	7.09×10^{-5}	0.379	0.303	9.89×10^{-6}	1.402
15	**rs8035957**	**36,625,556**	**423:342**	**1.24**	**0.0034**	**0.304**	**0.263**	**0.011**	**1.225**
16	rs12931878	10,949,695	362:446	0.812	0.0031	0.16	0.225	1.01×10^{-5}	0.657
16	rs1035089	10,955,851	517:451	1.146	0.034	0.48	0.42	8.25×10^{-4}	1.277
16	rs13330041	10,996,309	291:345	0.844	0.032	0.172	0.246	1.01×10^{-6}	0.637
16	rs725613	11,077,184	397:465	0.854	0.021	0.3	0.39	3.24×10^{-7}	0.672
16	rs2041670	11,082,153	384:444	0.865	0.037	0.265	0.345	2.01×10^{-6}	0.682
16	rs17673553	11,149,407	319:387	0.824	0.01	0.202	0.279	1.30×10^{-6}	0.655
21	**rs9976767**	**42,709,459**	**571:504**	**1.13**	**0.041**	**0.474**	**0.437**	**0.038**	**1.164**

Continued on following page

ethnicity and mean age 8.82 years (50.83% male and 49.17% female) who did not have diabetes or a first-degree relative with type 1 diabetes. These individuals were recruited by CHOP's clinicians and nursing staff within the CHOP's health care network, including four primary care clinics and several group practices and outpatient practices that included routine check-up visits of healthy children. Of these 2,024 individuals, 1,673 were selected using population-stratification analysis from Eigenstrat similar to that described above for DCCT/EDIC probands (868 male, 801 female, and 4 with ambiguous gender). We removed 351 (17.3%) self-reported European individuals from the control group to address the population heterogeneity. The Research Ethics Board of CHOP approved the study, and written informed consent was obtained from all subjects.

Genotyping. Genotypes for this study were obtained using the Infinium and GoldenGate platforms from Illumina. We performed high-throughput genome-wide SNP genotyping using the Illumina Infinium II HumanHap550 BeadChip technology (Illumina, San Diego) (10,11) at the Center for Applied Genomics at CHOP. We used 750 ng genomic DNA to genotype each sample according to the manufacturer's guidelines. DCCT/EDIC samples were genotyped on the Illumina 1M chip at Illumina (San Diego, CA).

Statistics. All statistical tests for association were carried out using the software package PLINK (12). The single-marker analysis for the genome-wide data were carried out using a χ^2 test on allele-count differences between 563 case and 1,146 control subjects. Odds ratios (ORs) and corresponding 95% CIs were calculated for the association analysis. The transmission disequilibrium test was used to calculate *P* values on differences between transmitted and untransmitted allele counts in the type 1 diabetic trios and nuclear families. Counts of untransmitted and transmitted alleles from heterozygous parents to affected offspring were determined using the standard transmission disequilibrium test implemented in the Haploview software package (13). The *P* values from the case-control and family-based analyses in our three discovery cohorts were combined by weighted *z* scores to quantify the overall evidence for association.

RESULTS

The flow process of this study is shown in Table 1. Comparisons of the statistical power of each population cohort are shown in supplementary Fig. 1 (available in an online appendix at http://dx.doi.org/10.2337/db08-1022). Using our GWA data from 563 Caucasian type 1 diabetes probands and 1,146 control subjects plus 483 type 1 diabetes case-parent trios using the Illumina HumanHap550 BeadChip (3), we identified 982 SNPs outside the major histocompatibility complex region that were suggestive of a potential type 1 diabetes association in the same direction in both cohorts ($P < 0.05$). We then genotyped these SNPs using the Illumina GoldenGate platform in an independent cohort of 636 nuclear type 1 diabetic families from Canada and the Type 1 Diabetes Genetics Consortium. With the completion of genotyping the third cohort, the WTCCC summary data became available (http://www.wtccc.org.uk) (4). Consequently, we selected markers that met the $P < 0.05$ threshold both in this third cohort and in the WTCCC dataset (4). Imputation from the Affymetrix

TABLE 2
Continued

Type 1 diabetic family trios				WTCCC				
Trans:untrans	OR	TDT *P*	Combined *P*	Aff allele frequency	Ctrl allele frequency	*P*	OR	Gene
105:136	**0.772**	**0.046**	**0.00060**	**0.122**	**0.137**	**0.036**	**0.878**	***EDG7****
209:161	1.298	0.013	7.56×10^{-10}	0.276	0.217	3.34×10^{-11}	1.371	*PTPN22*
224:172	1.302	0.009	9.5×10^{-10}	0.299	0.241	2.81×10^{-10}	1.339	*PTPN22*
225:168	1.339	0.004	1.89×10^{-9}	0.298	0.240	1.70×10^{-10}	1.344	*PTPN22*
212:164	1.293	0.013	2.08×10^{-8}	0.287	0.226	1.09×10^{-11}	1.377	*PTPN22*
218:165	1.321	0.0068	7.88×10^{-10}	0.287	0.227	2.69×10^{-11}	1.368	*PTPN22*
181:127	1.425	0.0021	1.95×10^{-10}	0.246	0.179	4.05×10^{-16}	1.504	*PTPN22*
124:82	1.512	0.0034	1.55×10^{-5}	0.134	0.115	0.0042	1.195	*PTPN22*
120:80	1.5	0.0047	5.54×10^{-6}	0.130	0.109	0.0018	1.218	*PTPN22*
210:254	0.827	0.041	3.76×10^{-5}	0.359	0.395	0.00034	0.858	*IFIH1*
203:251	0.809	0.024	4.48×10^{-5}	0.350	0.389	8.73×10^{-5}	0.845	*IFIH1*
222:173	1.283	0.014	0.000175	0.441	0.403	0.00024	1.166	*CTLA4*
236:191	1.236	0.029	0.000349	0.440	0.402	0.00022	1.167	*CTLA4*
228:177	1.288	0.011	0.000105	0.372	0.332	5.10×10^{-5}	1.191	*CTLA4*
253:209	**1.211**	**0.041**	**0.00035**	**0.511**	**0.489**	**0.033**	**1.092**	***BACH2****
254:209	**1.215**	**0.037**	$\mathbf{7.14 \times 10^{-6}}$	**0.440**	**0.410**	**0.004**	**1.129**	***GLIS3****
253:209	**1.211**	**0.041**	$\mathbf{1.60 \times 10^{-5}}$	**0.456**	**0.427**	**0.004**	**1.127**	***GLIS3****
185:257	0.72	6.16×10^{-4}	9.33×10^{-8}	0.419	0.458	0.00012	0.852	*IL2RA*
240:181	1.326	0.004	1.27×10^{-6}	0.361	0.319	1.32×10^{-5}	1.208	*IL2RA*
268:200	1.34	0.0017	2.45×10^{-7}	0.455	0.409	6.24×10^{-6}	1.207	*IL2RA*
264:196	1.347	0.0015	5.89×10^{-9}	0.455	0.409	7.31×10^{-6}	1.205	*IL2RA*
160:228	0.7018	5.56×10^{-4}	2.61×10^{-16}	0.443	0.464	0.047	0.921	*INS*
197:244	0.807	0.025	3.20×10^{-7}	0.404	0.452	2.40×10^{-6}	0.821	*12q13*
265:188	1.41	2.97×10^{-4}	4.56×10^{-9}	0.475	0.414	2.04×10^{-9}	1.283	*12q13*
245:180	1.361	0.0016	2.78×10^{-10}	0.397	0.339	5.91×10^{-9}	1.282	*12q13*
204:162	**1.259**	**0.028**	$\mathbf{2.88 \times 10^{-5}}$	**0.292**	**0.268**	**0.01**	**1.126**	***RASGRP1****
128:162	0.79	0.046	1.51×10^{-8}	0.158	0.178	0.0088	0.865	*KIAA0350*
265:212	1.25	0.015	9.85×10^{-7}	0.482	0.439	2.59×10^{-5}	1.190	*KIAA0350*
145:183	0.792	0.036	1.45×10^{-8}	0.175	0.204	0.00028	0.825	*KIAA0350*
178:248	0.718	6.95×10^{-4}	1.67×10^{-11}	0.292	0.340	3.90×10^{-7}	0.797	*KIAA0350*
172:233	0.738	0.0024	7.08×10^{-10}	0.264	0.315	7.05×10^{-8}	0.781	*KIAA0350*
146:203	0.719	0.0023	3.82×10^{-10}	0.217	0.249	0.00031	0.838	*KIAA0350*
260:203	**1.281**	**0.008**	**0.00020**	**0.493**	**0.461**	**0.002**	**1.135**	***UBASH3A****

The six SNPs indicated in bold type represent novel associations deemed appropriate for further investigation. For the joint analysis of the three discovery cohorts, the two family cohorts were pooled for the TDT analysis. The TDT results were combined with those of the case-control cohort by weighted *z* scores. Combined *P* values for the three cohorts are shown, together with the gene in which the markers resides or to which they are nearest. *P* values are two sided in each instance. *Gene not previously implicated in type 1 diabetes. Aff allele freq, minor allele frequency in affected individuals; Chr, chromosome; Ctrl allele freq, minor allele frequency in unaffected individuals; T1DGC, Type 1 Diabetes Genetics Consortium; Trans:untrans, transmitted:untransmitted allele ratio.

data of the WTCCC set was near perfect in all cases (supplementary Table 1). As shown in Table 2, 33 markers met the $P < 0.05$ threshold across all four cohorts. Although the bulk of them mapped to known loci (*PTPN22* [14], 12q13, *KIAA0350* [3,6], *IL2RA* [15–17], *CTLA4* [18], and *IFIH1* [19]), six SNPs in five loci were completely novel. These were tested in an additional case-control cohort consisting of 1,303 type 1 diabetes probands from the DCCT/EDIC study and an independent dataset of 1,673 control subjects from Philadelphia who had been genotyped on the Illumina 1M and HumanHap550K BeadChips, respectively.

Two signals replicated in this fifth independent cohort (Table 3), and the *P* values were significant after correction for testing six markers (five independent loci). They map to *UBASH3A* (ubiquitin-associated and SH3 domain-containing protein A) and *BACH2* (broad complex-tramtrack-bric-a-brac [BTB] and cap 'n' collar [CNC] homology 2). Table 4 shows that rs9976767 is in fact significant at the genome-wide level when all five cohorts utilized were combined ($P = 2.33 \times 10^{-8}$).

DISCUSSION

Taken together, our full second-stage approach and combined meta-analysis have revealed additional loci associated with type 1 diabetes. Clearly the risks are relatively modest compared with previously described associations, and it was only with this sample size at our disposal that we could we detect and establish these signals as true positives through an independent validation effort.

UBASH3A is the only gene in its corresponding region of linkage disequilibrium. Mice lacking *Sts2* (the mouse homologue for *UBASH3A*) have been shown to be normal in all respects, including T-cell function (20). Mice lacking both *Sts1* and *Sts2* do have increased splenocyte numbers and are hyper-responsive to T-cell receptor stimulation. It has been suggested that STS1 and STS2 are critical regulators of the signaling pathways that control T-cell activation (20).

BACH2 is also the only gene at its corresponding region of linkage disequilibrium. The gene product is a member of the small Maf family, which consists of basic region

TABLE 3
Validation results for the six SNPs of interest selected from the discovery process in the DCCT/EDIC type 1 diabetes probands and CHOP control subjects

Chr	SNP	Position	Gene	Aff allele freq	Ctrl allele freq	OR (95% CI)	*P*
1	rs1983853	85,083,780	*EDG7*	0.132	0.153	0.842 (0.726–0.976)	0.022
6	**rs3757247**	**91,014,184**	*BACH2*	**0.497**	**0.463**	**1.144 (1.033–1.268)**	**0.010**
9	rs10758593	4,282,083	*GLIS3*	0.429	0.426	1.013 (0.913–1.124)	0.81
9	rs10758594	4,285,583	*GLIS3*	0.434	0.443	0.963 (0.869–1.068)	0.48
15	rs8035957	36,625,556	*RASGRP1*	0.270	0.261	1.047 (0.932–1.176)	0.44
21	**rs9976767**	**42,709,459**	*UBASH3A*	**0.474**	**0.436**	**1.165 (1.051–1.292)**	**0.0036**

The two SNPs that successfully replicated are presented in bold type. Minor allele frequencies, *P* values, and ORs are shown together with the gene in which the markers resides or to which they are nearest. *P* values are two-sided in each instance. Aff allele freq, allele frequency in affected individuals; Chr, chromosome; Ctrl allele freq, allele frequency in unaffected individuals.

leucine zipper proteins that function either as transcriptional activators or repressors depending on the proteins with which they heterodimerize. Muto et al. (21) found that *Bach2*$^{-/-}$ mice had relatively high levels of serum IgM but low levels of IgA and IgG subclasses. The *Bach2*$^{-/-}$ mice have also been reported to present with deficient T-cell–independent and T-cell–dependent IgG responses, leading the authors to conclude that BACH2 was a regulator of the antibody response.

It should also be noted that rs1983853 yielded a nominally significant association with type 1 diabetes in all of the cohorts but did not survive correction for multiple testing in the final validation attempt in the Toronto dataset. This SNP resides in endothelial differentiation gene 7 (*EDG7*; formerly *LPA3*), which has been implicated in mechanisms of embryo implantation (22). The SNPs on *GLIS3* and *RASGRP1* were not validated. They may have been false positives in the earlier stages; alternatively, lack of replication in DCCT/EDIC may be due to different and/or weaker genetic risk determinants in this cohort with late age of onset of type 1 diabetes. This question must be addressed in future studies. The GLI-similar 3 (*GLIS3*) gene plays important roles in the development of pancreatic β-cells. Mutations in this gene cause a rare syndrome with neonatal diabetes and congenital hypothyroidism (23). The RAS guanyl releasing protein 1 (*RASGRP1*) gene has important roles in immune regulation, and it has been suggested that it contributes to the autoimmunity of systemic lupus erythematosus (24).

In addition to our findings, what we failed to find deserves comment. In addition to the findings described above, our study confirmed another interesting locus, rs17696736 (*C12orf30*) at 12q24, reported in the WTCCC study (4,6). Our GWA family cohort suggested type 1 diabetes association with $P = 0.011$; however, limited by the sample size, our GWA case-control cohort did not show statistical significance ($P > 0.05$). To validate the type 1 diabetes association, we genotyped rs17696736 using the Sequenom iPLEX assay (Sequenom, Cambridge, MA) in the 1,120 Canadian families and the 549 Type 1 Diabetes Genetics Consortium families. The call rate of rs17696736 genotyping was 99.8%, and no Mendelian error was found. With the family-based association test (25), we confirmed the type 1 diabetes association with $P = 8.00 \times 10^{-7}$, minor G allele frequency 0.452, and OR 1.276. However, given the very thorough coverage of European genetic variation by the Hap550 and the power of our aggregate sample size, it is very unlikely that we missed more than a very small number of common variants with an effect size approaching that of the *INS* (minor allele frequency 0.2 and OR 0.5; each of our three discovery cohorts has >99.9% power to detect it at $\alpha = 0.05$ level) or *PTPN22* (minor allele frequency 0.1 and OR 1.8; each of our three discovery cohorts has >99.0% power to detect it at $\alpha = 0.05$ level) loci.

Undoubtedly, larger sample sizes and meta-analysis of all available GWA data will discover an increasing number of loci with decreasing effect sizes, which are unlikely to explain the remaining familial clustering of type 1 diabetes. Such explanation should be sought, it appears, in rare variants, the detection of which is now coming within reach with the use of high-throughput methods for sequencing and for detecting structural variation.

ACKNOWLEDGMENTS

We gratefully acknowledge the use of DNA samples from the Type 1 Diabetes Genetics Consortium, funded by National Institutes of Health Grant U01-DK62418. This work was funded in part by the Juvenile Diabetes Research Foundation (JDRF) International and Genome Canada through the Ontario Genomics Institute. H.Q.Q. is supported by a fellowship from the Canadian Institutes of Health Research. All genotyping and other aspects of the study were funded by an institutional development grant to the Center for Applied Genomics from CHOP. S.F.A.G. and H.H. are funded in part by a JDRF award and a development award from the Cotswold Foundation. This

TABLE 4
Meta-analysis of the five cohorts

Chr	SNP	Position	Gene	Allele	OR (95% CI)	*P*
1	rs1983853	85,083,780	*EDG7*	A	0.833 (0.773–0.898)	1.87×10^{-6}
6	rs3757247	91,014,184	*BACH2*	A	1.134 (1.078–1.193)	1.25×10^{-6}
9	rs10758593	4,282,083	*GLIS3*	A	1.131 (1.074–1.190)	2.64×10^{-6}
9	rs10758594	4,285,583	*GLIS3*	A	1.114 (1.058–1.172)	3.51×10^{-5}
15	rs8035957	36,625,556	*RASGRP1*	C	1.144 (1.080–1.211)	3.92×10^{-6}
21	rs9976767	42,709,459	*UBASH3A*	C	1.155 (1.098–1.215)	2.33×10^{-8}

P values and ORs are shown together with the relevant allele for each of the six SNPs.

work has received funding from the National Institute of Diabetes and Digestive and Kidney Diseases (N01-DK-6-2204 and R01-DK-077510). A.D.P. holds a Canada Research Chair in the Genetics of Common Diseases.

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

We thank all the patients, their parents, and the healthy control subjects for their participation in the study.

REFERENCES

1. Todd JA, Bell JI, McDevitt HO: HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 329:599–604, 1987
2. Risch N: Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 40:1–14, 1987
3. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448:591–594, 2007
4. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678, 2007
5. Hakonarson H, Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Eckert AW, Annaiah K, Lawson ML, Otieno FG, Santa E, Shaner JL, Smith RM, Onyiah CC, Skraban R, Chiavacci RM, Robinson LJ, Stanley CA, Kirsch SE, Devoto M, Monos DS, Grant SF, Polychronakos C: A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* 57:1143–1146, 2008
6. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JHM, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AAC, Ovington NR, Allen J, Adlem E, Leung H-T, Wallace C, Howson JMM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SCL, Dunger DB, Wicker LS, Clayton DG: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39:857–864, 2007
7. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
8. The DCCT Research Group: The Diabetes Control and Complications Trial (DCCT): design and methodologic considerations for the feasibility phase. *Diabetes* 35:530–545, 1986
9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909, 2006
10. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS: A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37:549–554, 2005
11. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL: Whole-genome genotyping with the single-base extension assay. *Nat Methods* 3:31–33, 2006
12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575, 2007
13. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of linkage disequilibrium and haplotype maps. *Bioinformatics* 21:263–265, 2005
14. Bottini N, Vang T, Cucca F, Mustelin T: Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 18:207–213, 2006
15. Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, Widmer B, Jones R, Ring SM, McArdle W, Pembrey ME, Strachan DP, Dunger DB, Twells RC, Clayton DG, Todd JA: Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76:773–779, 2005
16. Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C: Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes* 56:1174–1176, 2007
17. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS, Todd JA: Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* 39:1074–1082, 2007
18. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511, 2003
19. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Tirgoviste C, Widmer B, Dunger DB, Savage DA, Walker NM, Clayton DG, Todd JA: A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 38:617–619, 2006
20. Carpino N, Turner S, Mekala D, Takahashi Y, Zang H, Geiger TL, Doherty P, Ihle JN: Regulation of ZAP-70 activation and TCR signaling by two related proteins, Sts-1 and Sts-2. *Immunity* 20:37–46, 2004
21. Muto A, Tashiro S, Nakajima O, Hoshino H, Takahashi S, Sakoda E, Ikebe D, Yamamoto M, Igarashi K: The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature* 429:566–571, 2004
22. Ye X, Hama K, Contos JJ, Anliker B, Inoue A, Skinner MK, Suzuki H, Amano T, Kennedy G, Arai H, Aoki J, Chun J: LPA3-mediated lysophosphatidic acid signalling in embryo implantation and spacing. *Nature* 435:104–108, 2005
23. Senee V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, Bougneres P, Taha D, Julier C: Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* 38:682–687, 2006
24. Yasuda S, Stevens RL, Terada T, Takeda M, Hashimoto T, Fukae J, Horita T, Kataoka H, Atsumi T, Koike T: Defective expression of Ras guanyl nucleotide-releasing protein 1 in a subset of patients with systemic lupus erythematosus. *J Immunol* 179:4890–4900, 2007
25. Horvath S, Xu X, Laird NM: The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 9:301–306, 2001