

# A Novel Susceptibility Locus for Type 1 Diabetes on Chr12q13 Identified by a Genome-Wide Association Study

Hakon Hakonarson,<sup>1,2</sup> Hui-Qi Qu,<sup>3</sup> Jonathan P. Bradfield,<sup>1</sup> Luc Marchand,<sup>3</sup> Cecilia E. Kim,<sup>1</sup> Joseph T. Glessner,<sup>1</sup> Rosemarie Grabs,<sup>3</sup> Tracy Casalunovo,<sup>1</sup> Shayne P. Taback,<sup>4</sup> Edward C. Frackelton,<sup>1</sup> Andrew W. Eckert,<sup>1</sup> Kiran Annaiah,<sup>1</sup> Margaret L. Lawson,<sup>5</sup> F. George Otieno,<sup>1</sup> Erin Santa,<sup>1</sup> Julie L. Shaner,<sup>1</sup> Ryan M. Smith,<sup>1</sup> Chioma C. Onyiah,<sup>1</sup> Robert Skraban,<sup>1</sup> Rosetta M. Chiavacci,<sup>1</sup> Luke J. Robinson,<sup>1</sup> Charles A. Stanley,<sup>6</sup> Susan E. Kirsch,<sup>7</sup> Marcella Devoto,<sup>2,8</sup> Dimitri S. Monos,<sup>9,10</sup> Struan F.A. Grant,<sup>1,2</sup> and Constantin Polychronakos<sup>3</sup>

**OBJECTIVE**—In stage 1 of our genome-wide association (GWA) study for type 1 diabetes, one locus at 16p13 was detected ( $P = 1.03 \times 10^{-10}$ ) and confirmed in two additional cohorts. Here we describe the results of testing, in these additional cohorts, 23 loci that were next in rank of statistical significance.

**RESEARCH DESIGN AND METHODS**—Two independent cohorts were studied. The Type 1 Diabetes Genetics Consortium replication cohort consisted of 549 families with at least one child diagnosed with diabetes (946 total affected) and DNA from both parents. The Canadian replication cohort consisted of 364 nuclear family trios with one type 1 diabetes-affected offspring and two parents (1,092 individuals).

**RESULTS**—One locus at 12q13, with the highest statistical significance among the 23, was confirmed. It involves type 1 diabetes association with the minor allele of rs1701704 ( $P = 9.13 \times 10^{-10}$ , OR 1.25 [95% CI 1.12–1.40]).

**CONCLUSIONS**—We have discovered a type 1 diabetes locus at 12q13 that is replicated in an independent cohort of type 1 diabetic patients and confers a type 1 diabetes risk comparable with that of the 16p13 locus we recently reported. These two loci are identical to two loci identified by the whole-

genome association study of the Wellcome Trust Case-Control Consortium, a parallel independent discovery that adds further support to the validity of the GWA approach. *Diabetes* 57:1143–1146, 2008

**T**ype 1 diabetes, a multifactorial disease with a strong genetic component, is due to the autoimmune destruction of pancreatic  $\beta$ -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 less than half of genetic type 1 diabetes risk (2). The remainder is accounted for by many loci with more modest effects, of which only a few are known. These include 1) the insulin (*INS*) VNTR (variable number tandem repeats) (3) modulating thymic expression of and tolerance to insulin, a major type 1 diabetes autoantigen (4,5); 2) the Arg620Trp single nucleotide polymorphism (SNP) at *PTPN22*, which affects the function of a negative regulator of T-cell receptor signaling (6); 3) noncoding SNPs at *IL2RA* (7–9), which encodes the  $\alpha$ -chain of the IL2 receptor complex (CD25), an important modulator of immunity; and 4) variants in the *CTLA4* locus (10) whose protein product transmits inhibitory signals to attenuate T-cell activation. It is worth noting that all of these type 1 diabetes-associated genes are expressed in cells with immune function, and all except *INS* have been associated with other autoimmune disorders. Together, these loci explain over half of the genetic type 1 diabetes risk, the remaining of which is composed of loci whose number and effect size are still unknown. It is also worth remarking that these known loci represent the “low-hanging fruit,” involving candidate genes whose importance in autoimmunity was already known, thus limiting the value of genetic discovery to generate previously unsuspected physiological insights.

The recent development of high-throughput genotyping array technologies has enabled us (11) and others (12,13) to perform genome-wide association (GWA) studies in search of the remaining type 1 diabetes loci. The first successful use in type 1 diabetes involved the screening of 12,000 nonsynonymous SNPs, which found type 1 diabetes association with rs1990760, involving an Ala946Thr substitution on the *IFIH1* (interferon-induced with helicase C domain 1) gene (14). More recently, the Wellcome Trust Case-Control Consortium (WTCCC) tested 2,000 type 1 diabetic case and 3,000 control subjects for 500 k SNPs (Affymetrix GeneChip) (12), and four novel associations

From the <sup>1</sup>Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the <sup>2</sup>Department of Pediatrics and Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the <sup>3</sup>Departments of Pediatrics and Human Genetics, McGill University, Montreal, Quebec, Canada; the <sup>4</sup>Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada; the <sup>5</sup>Division of Endocrinology, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada; the <sup>6</sup>Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; <sup>7</sup>Markham-Stouffville Hospital, Markham, Ontario, Canada; the <sup>8</sup>Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania; the <sup>9</sup>Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; and the <sup>10</sup>Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.

Address correspondence and reprint requests to Hakon Hakonarson, MD, PhD, Director, Center for Applied Genomics, 1216E Abramson Research Center, 3615 Civic Center Blvd., Philadelphia, PA 19104-4318, E-mail: hakonarson@chop.edu; or to Constantin Polychronakos, MD, FRCP, McGill University Health Center (Children's Hospital), 2300 Tupper, Montréal, QC, Canada, H3H 1P3, E-mail: constantin.polychronakos@mcgill.ca.

Received for publication 13 September 2007 and accepted in revised form 10 January 2008.

Published ahead of print at <http://diabetes.diabetesjournals.org> on 15 January 2008. DOI: 10.2337/db07-1305.

Additional information for this article can be found in an online appendix at <http://dx.doi.org/10.2337/db07-1305>.

GWA, genome-wide association; LD, linkage disequilibrium; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case-Control Consortium.

© 2008 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.

TABLE 1  
The type 1 diabetes-associated 12q13 SNPs in two independent cohorts

SNP	Minor allele	Frequency	Informative families*	S†	E(S)‡	Var(S)§	Z	P
Consortium set								
rs773107	G	0.322	217	329	290	208	2.73	0.006400
rs10876864	G	0.417	226	395	350	229	2.98	0.002885
rs1701704	C	0.346	216	341	296	211	3.09	0.002019
Canadian set								
rs773107	G	0.339	200	180	160	75	2.37	0.017736
rs10876864	G	0.452	212	210	191	87	2.09	0.036293
rs1701704	C	0.357	216	189	169	81	2.22	0.026268
Combined								
rs773107	G	0.329	417	509	449	283	3.56	0.000374
rs10876864	G	0.432	438	605	540	316	3.64	0.000278
rs1701704	C	0.350	432	530	465	292	3.80	0.000148

\*Number of nuclear families informative for (with a non-zero contribution to) FBAT analysis. †Observed allele number in the affected offspring. ‡Expected allele number in the affected offspring. §Variance of allele distribution among the affected offspring.

(12q24, 12q13, 16p13, and 18p11) were solidly replicated in 4,000 case and 5,000 control subjects, plus an additional 2,997 type 1 diabetic family trios, by the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (13). Our stage 1 results, which have recently been published (11), report one novel type 1 diabetes locus at Chr16p that reached genome-wide significance in stage 1 and coincided with one of the novel WTCCC findings. We used the Illumina Hap550 array to type 550 k SNPs, selected to tag most of the haplotype structure of the human genome, per phase 2 of the International HapMap project (<http://www.hapmap.org>). A set of 483 subjects with type 1 diabetes and both parents of each subject in addition to 563 type 1 diabetic probands and 1,146 control subjects, all of European ancestry, were examined. For the case-control data, population stratification was corrected by principal component analysis in the EIGENSTRAT implementation (15).

Both studies independently mapped this association to a 300-kb linkage disequilibrium (LD) block on 16p encompassing two genes of unknown function, *KIAA0350* (now renamed C-type lectin 16A, symbol *CLEC16A*) and *LOC642451*.

Pending completion of a full stage 2, in which the 1,536 markers with the highest statistical significance will be tested in additional cohorts, we decided to fast track 24 SNPs (23 distinct loci) that, in rank of statistical significance, came next to the 16p locus. This number was arbitrarily determined by the multiplexing capacity of our genotyping platform and corresponds to a cut off of  $\sim 10^{-4}$  (supplementary Table 1 [available in an online appendix at <http://dx.doi.org/10.2337/db07-1305>]). The power of our GWA to detect a range of effect sizes at that level is given in supplementary Fig. 1.

## RESEARCH DESIGN AND METHODS

The sample sets used for GWA scanning have been previously described in detail (11). The Type 1 Diabetes Genetics Consortium replication cohort

TABLE 2  
The combined *P* values of type 1 diabetes association of the three 12q13 SNPs

SNP	Chromosome	Position	GWA case-control*	GWA families	Stage 2	Combined
rs773107	12	54,655,773	$2.89 \times 10^{-5}$	$7.81 \times 10^{-3}$	$3.74 \times 10^{-4}$	$2.48 \times 10^{-8}$
rs10876864	12	54,687,352	$8.39 \times 10^{-5}$	$2.97 \times 10^{-4}$	$2.78 \times 10^{-4}$	$2.47 \times 10^{-9}$
rs1701704	12	54,698,754	$9.89 \times 10^{-6}$	$1.62 \times 10^{-3}$	$1.48 \times 10^{-4}$	$9.13 \times 10^{-10}$

Data are *P* values. \*Corrected by the EIGENSTRAT method (ref. 15).

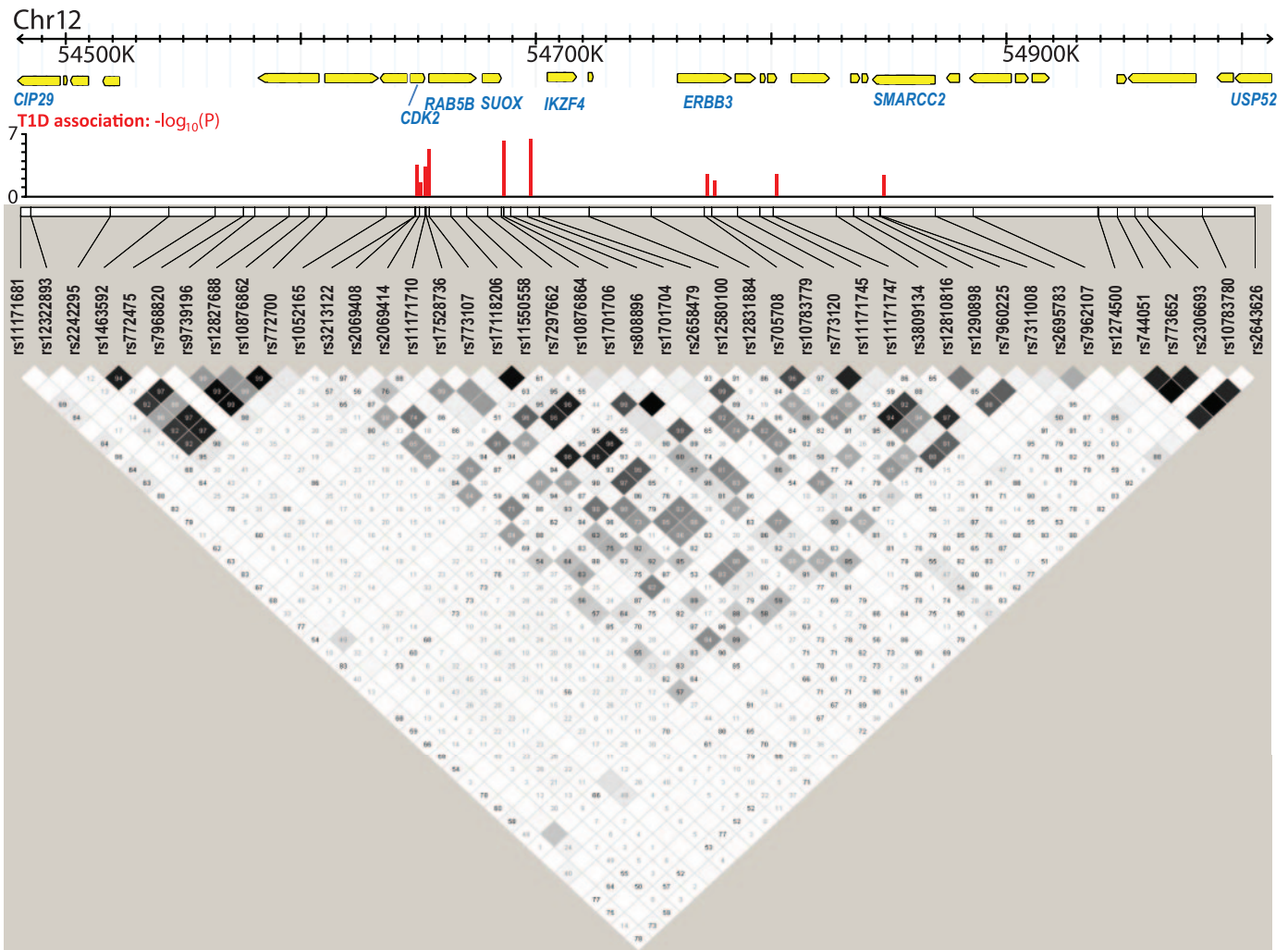
consisted of 549 families with at least one child diagnosed with diabetes (946 total affected) and DNA from both parents available as of the July 2005 data freeze (<https://www.t1dgc.org>). The samples were collected in Europe, North America, and Australia, and most subjects were of European ancestry. Criteria were age at diagnosis <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 antibodies and/or low C-peptide levels at diagnosis. The Canadian cohort consisted of 364 nuclear family trios with one type 1 diabetes-affected offspring and two parents (1,092 individuals). The samples were collected in pediatric diabetes clinics in Montreal, Toronto, Ottawa, and Winnipeg. All patients were diagnosed under the age of 18 years and were treated with insulin since diagnosis, and none has stopped treatment for any reason since. Ethnic backgrounds were of mixed European descent. 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.

**Genotyping.** Genotypes for this study were obtained using the Sequenom iPLEX assay (Sequenom, Cambridge, MA). The 90 CEUs (European-descent individuals genotyped in HapMap) were included as accuracy control subjects. The call rate of each of the three 12q13 SNPs was >98.2%, and no Mendelian error was found. The genotypes of each SNP were in Hardy-Weinberg equilibrium in the parents.

**Statistics.** Type 1 diabetes association was tested by the Family Based Association Test (FBAT) software (<http://www.biostat.harvard.edu/~fbat/fbat.htm>) (16), based on the transmission disequilibrium test method. Considering that most of the Type 1 Diabetes Genetics Consortium families have multiple siblings, the option of the empirical variance was used in the FBAT statistics to permit a robust but unbiased test of genetic association. The age-of-onset effect of the SNPs in type 1 diabetic patients was tested by the Kruskal-Wallis test.

## RESULTS AND DISCUSSION

As shown in Table 1, the type 1 diabetes association at the 12q13 locus was replicated in both sets, with similar odds ratios. In a combined analysis of the two GWA cohorts and a pool of the two replication cohorts by Fisher's method, all three SNPs attained *P* values ranging from  $2.48 \times 10^{-8}$  to  $9.13 \times 10^{-10}$  (Table 2). These data provide convincing evidence of a novel type 1 diabetes locus. The other loci fast tracked to stage 2 are shown in supplementary Table 1. While this report was in preparation the WTCCC results



**FIG. 1.** Type 1 diabetes association in the 12q13 region. The combined results of type 1 diabetes association tests of the case-control cohort and the family cohort is shown. The red bars, corresponding to  $-\log_{10}(P)$  values, represent the SNPs of significant type 1 diabetes association with  $P < 0.05$ . The LD map is based on our genotyping data of the family cohort.  $D'$  values (%) are shown in the boxes, and for the empty boxes  $D' = 1$ . The gray scale represents the  $r^2$  values. The type 1 diabetes-associated SNPs map to an LD region in the middle of the figure, while the strongest association is from the region around the *RAB5B* gene. The  $r^2$  values drastically drop beyond the two ends of the block containing the type 1 diabetes-associated SNPs. Relatively high  $D'$  values continue for an additional ~250 kb in the telomeric direction, but beyond the block shown there is no type 1 diabetes-associated SNP and no HapMap SNPs with substantial  $r^2$  values to those in the block shown (extended LD diagram in supplementary Fig. 3).

were published, reporting four novel type 1 diabetes associations, two of which coincide with our own two novel findings at 16p (13) and the one reported here, 12q13. We consider this simultaneous independent discovery of noncandidate loci a potent validation of the GWA approach.

The three SNPs are located in a 250-kb block of tight LD on Chr12q and are highly correlated ( $r^2 = 0.608-0.855$ ). Type 1 diabetes risk was conferred by the minor allele in all three. The OR (95% CI) of the haplotype of the three SNPs' minor alleles was 1.25 (1.12-1.40). The genotypic association of the SNPs best fits an additive model of susceptibility on a log scale (supplementary Table 2). There was a trend, with borderline statistical significance, of an effect on age-of-onset of type 1 diabetes (supplementary Table 3): homozygotes of the risk allele may have the disease onset 1 year earlier than the homozygotes of the protective allele. This is clearly the same locus as the one reported by Todd et al. (13), as the type 1 diabetes-associated rs2292239 is in tight LD with our markers (supplementary Fig. 2). In addition, imputed genotype

counts for our three significant markers in the WTCCC data showed highly significant type 1 diabetes association, as shown in a meta-analysis of the two studies for the 12q13 locus (supplementary Table 4). Significance was increased by several orders of magnitude. None of the other loci tested in this report (supplementary Table 1) achieved statistical significance as a result of this joint analysis.

The associated 250-kb LD block encompasses several genes that now become candidates for the type 1 diabetes association (Fig. 1 and supplementary Fig. 3). *RAB5B* (MIM:179514) encodes a member of the *RAS* oncogene family that may be involved in vesicular trafficking at the plasma membrane (17). *SUOX* (MIM:606887) encodes a liver-specific sulfite oxidase involved in the degradation of sulfur-containing amino acids. Neither gene has known function in immunity or pancreatic development. *IKZF4* (MIM:606239) encodes a better functional candidate, a zinc finger protein specifically expressed in lymphocytes and implicated in the control of lymphoid development. *ERBB3* (MIM:190151) encodes a member of the epidermal



growth factor receptor family of receptor tyrosine kinases involved in the regulation of cell proliferation or differentiation. Interestingly, it interacts with proteins with important immune functions, such as Lyn, Itk, and Fer (18). *CDK2* (MIM: 116953) encodes a member of the Ser/Thr protein kinase family whose activity is regulated by its protein phosphorylation; members of this gene family have been recently implicated in GWA studies of type 2 diabetes (19–22). Fine mapping and functional studies will be required to identify the causative variant and generate functional insights from this genetic finding.

A more immediate conclusion from this finding is the support it provides to the robustness of the GWA approach, through identification of the same loci in two independent studies. Further, it is worth noting that our power to include in this fast-track replication a locus of magnitude comparable with the one reported here was only 38%. This means that our full second stage, plus future stages and combined meta-analyses, are likely to reveal additional such loci.

#### 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 by the Children's Hospital of Philadelphia, the Juvenile Diabetes Research Foundation International and Genome Canada through the Ontario Genomics Institute. H.Q.Q. is supported by a fellowship from the Canadian Institutes of Health Research.

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

#### REFERENCES

- Todd JA, Bell JL, McDevitt HO: HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 329:599–604, 1987
- Risch N: Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet* 40:1–14, 1987
- Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA: Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes* 53:1884–1889, 2004
- Pugliese A, Zeller M, Fernandez A Jr, Zalcborg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297, 1997
- Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15:289–292, 1997
- 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
- 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
- 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
- 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
- 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, Allahabadi A, Nithiyanthan 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
- 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
- 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
- Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszkó JS, Hafler JP, Zeitels L, Yang JHM, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Mairuria 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
- 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
- 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
- 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
- Wilson DB, Wilson MP: Identification and subcellular localization of human rab5b, a new member of the ras-related superfamily of GTPases. *J Clin Invest* 89:996–1005, 1992
- Jones RB, Gordus A, Krall JA, MacBeath G: A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* 439:168–174, 2006
- Zeggini E, Weedon MN, Lindgren CM, et al.: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
- Saxena R, Voight BF, Lyssenko V, et al.: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336, 2007
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al.: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775, 2007
- Scott LJ, Mohlke KL, Bonnycastle LL, et al.: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345, 2007