A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci

Jonathan P. Bradfield1, Hui-Qi Qu2*, Kai Wang1, Haitao Zhang1, Patrick M. Sleiman1, Cecilia E. Kim1, Frank D. Mentch1, Haijun Qiu1, Joseph T. Glessner1, Kelly A. Thomas1, Edward C. Frackelton1, Rosetta M. Chiavacci1, Marcin Imieliński1, Dimitri S. Monos2,4, Rahul Pandey1, Marina Bakay1, Struan F. A. Grant1,3,5, Constantin Polychronakos2*, Hakon Hakonarson1,3,5*.

1The Center for Applied Genomics, The Children’s Hospital Philadelphia, Philadelphia, Pennsylvania, United States of America, 2Departments of Pediatrics and Human Genetics, McGill University, Montreal, Canada, 3Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, United States of America, 4Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, 5Division of Human Genetics, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America.

Abstract

Diabetes impacts approximately 200 million people worldwide, of whom approximately 10% are affected by type 1 diabetes (T1D). The application of genome-wide association studies (GWAS) has robustly revealed dozens of genetic contributors to the pathogenesis of T1D, with the most recent meta-analysis identifying in excess of 40 loci. To identify additional genetic loci for T1D susceptibility, we examined associations in the largest meta-analysis to date between the disease and ~2.54 million SNPs in a combined cohort of 9,934 cases and 16,956 controls. Targeted follow-up of 53 SNPs in 1,120 affected trios uncovered three new loci associated with T1D that reached genome-wide significance. The most significantly associated SNP (rs539514, \( P = 5.66 \times 10^{-11} \)) resides in an intronic region of the LMO7 (LIM domain only 7) gene on 13q22. The second most significantly associated SNP (rs478222, \( P = 3.50 \times 10^{-9} \)) resides in an intronic region of the EFR3B (protein EFR3 homolog B) gene on 2p23; however, the region of linkage disequilibrium is approximately 800 kb and harbors additional multiple genes, including NCOA1, C2orf79, CENPO, ADCY3, DNAJC27, POMC, and DNM1T3A. The third most significantly associated SNP (rs924043, \( P = 8.06 \times 10^{-7} \)) lies in an intergenic region on 6q27, where the region of association is approximately 900 kb and harbors multiple genes including WDR27, C6orf120, PHF10, TCL3E, C6orf208, LOC154449, DLI1, FAM120B, PSM81, TBP, and PCD2. These latest associated regions add to the growing repertoire of gene networks predisposing to T1D.

Citation: Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, et al. (2011) A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. PLoS Genet 7(9): e1002293. doi:10.1371/journal.pgen.1002293

Editor: Mark I. McCarthy, University of Oxford, United Kingdom

Received December 20, 2010; Accepted July 13, 2011; Published September 29, 2011

Copyright: © 2011 Bradfield et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: All genotyping and other aspects of the study were funded by an Institutional Development to the Center for Applied Genomics from CHOP. SFAG, CP, and HH are funded in part by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) award (DP3 DK085708) and SFAG and HH by a Developmental Award from the Crick Foundation. This research was financially supported by the Children’s Hospital of Philadelphia, Genome Canada through the Ontario Genomics Institute, and the Juvenile Diabetes Research Foundation. This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. The data from the T1DGC were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with investigators of the T1DGC study, except for those listed as authors on the current manuscript, and does not necessarily reflect the opinions or views of the T1DGC study, the NIDDK Central Repositories, or the NIDDK. The Diabetes Control and Complications Trial (DCCT) and its follow-up the Epidemiology of Diabetes Interventions and Complications (EDIC) study were conducted by the DCCT/EDIC Research Group and supported by National Institute of Health grants and contracts and by the General Clinical Research Center Program, NCRR. The data from the DCCT/EDIC study were supplied by the NIDDK Central Repositories. This manuscript was not prepared under the auspices of the DCCT/EDIC study and does not represent analyses or conclusions of the DCCT/EDIC study group, the NIDDK Central Repositories, or the NIH. The Genetics of Kidneys in Diabetes (GoKinD) Study was conducted by the GoKinD Investigators and supported by the Juvenile Diabetes Research Foundation, the CDC, and the Special Statutory Funding Program for Type 1 Diabetes Research administered by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data from the GoKinD study were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with investigators of the GoKinD study and does not necessarily reflect the opinions or views of the GoKinD study, the NIDDK Central Repositories, or the NIDDK. This manuscript makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hakonarson@chop.edu (HH); constantin.polychronakos@mcgill.ca (CP)

Introduction

Diabetes impacts approximately 200 million people worldwide [1], with microvascular and cardiovascular disease being the primary complications. Approximately 10% of cases are type 1 diabetes (T1D) sufferers, with ~3% increase in the incidence of T1D globally per year [2]. It is expected that the incidence is 40% higher in 2010 than in 1990 [3].

T1D is a clear example of a complex trait that results from the interplay between environmental and genetic factors. There are
Author Summary

Despite the fact that there is clearly a large genetic component to type 1 diabetes (T1D), uncovering the genes contributing to this disease has proven challenging. However, in the past three years there has been relatively major progress in this regard, with advances in genetic screening technologies allowing investigators to scan the genome for variants conferring risk for disease without prior hypotheses. Such genome-wide association studies have revealed multiple regions of the genome to be robustly and consistently associated with T1D. More recent findings have been a consequence of combining of multiple datasets from independent investigators in meta-analyses, which have more power to pick up additional variants contributing to the trait. In the current study, we describe the largest meta-analysis of T1D genome-wide genotyped datasets to date, which combines six large studies. As a consequence, we have uncovered three new signals residing at the chromosomal locations 13q22, 2p23, and 6q27, which went on to be replicated in independent sample sets. These latest associated regions add to the growing repertoire of gene networks predisposing to T1D.

many lines of evidence that there is a strong genetic component to T1D, primarily due to the fact that T1D has high concordance among monozygotic twins [4] and runs strongly in families, together with a high sibling risk [5].

Prior to the era of GWAS, only five loci had been fully established to be associated with T1D. However, the majority of the other reported associations in the pre-GWAS era [6–8] remain highly doubtful, where an initial report of association does not hold up in subsequent replication attempts by other investigative groups. This previous hazy picture of the genetics of T1D can be put down to the use of the only methodologies that were available at the time and which were much more limited than GWAS i.e. the candidate gene approach (where genomic regions were studied based on biological reasoning) and family-based linkage methodologies. Inconsistent findings can also be attributed to small sample sizes i.e. when power is low the false discovery rate tends to be high; GWAS per se has not improved consistency, rather it has leveraged large, well powered sample sizes combined with sound statistical analyses.

It has been long established that approximately half of the genetic risk for T1D is conferred by the genomic region harboring the HLA class II genes (primarily \textit{HLA-DRB1}, \textit{-DQA1} and \textit{-DQB1} genes), which encode the highly polymorphic antigen-presenting proteins. Other established loci prior to the application of GWAS are the genes encoding insulin (\textit{INS}) [9–12], cytotoxic T-lymphocyte-associated protein 4 (\textit{CTLA4}) [13–16], protein tyrosine phosphatase, non-receptor type 22 (\textit{PTPN22}) gene [17,18], interleukin 2 receptor alpha (\textit{IL2RA}) [19–21] and ubiquitin-associated and SH3 domain-containing protein A (\textit{UBASH3A}) [22].

The application of genome wide association studies (GWAS) has robustly revealed dozens of genetic contributors to T1D [23–29], the results of which have largely been independently replicated [30–36]. The most recently reported meta-analysis of this trait identified in excess of forty loci [29], including 18 novel regions plus confirmation of a number of loci uncovered through cross-disease comparisons [34–36]. As such, the risks conferred by these additional loci are relatively modest compared to the ‘low-hanging fruit’ described in the first studies and could only be ultimately uncovered when larger sample sizes were utilized.

We sought to expand further on this mode of analysis by combining our cohort with all publicly released genome wide SNP datasets to identify additional loci contributing to the etiology of T1D. Unfortunately, there is a relative paucity of control genotype data in these publicly available sources. To circumvent this problem, we combined individual level data from each available cohort and we then compared the cases with controls from two sources. We next separated all the individual level data into two groups, characterized by the type of genotyping platform that was used to genotype the samples, which would later be recombined using inverse-variance meta-analysis. The 6,523 cases genotyped on an Illumina BeadChip included subjects from McGill University, The Children’s Hospital of Philadelphia (CHOP), The Diabetes Control and Complications Trial – Epidemiology of Diabetes Interventions and Complications (DCCT-EDIC) cohort, and the Type 1 Diabetes Genetics Consortium (T1DGC), which in turn were compared with 6,648 similarly genotyped controls recruited at CHOP. The 3,411 cases genotyped on Affymetrix arrays included subjects from the Genetics of Kids in Diabetes Study (GoKinD) and the Wellcome Trust Cases Control Consortium (WTCCC) that were then compared with 10,308 similarly genotyped controls, including being derived from non-autoimmune disease related cases from the WTCCC, as well as from the British 1958 Birth Cohort and the UK National Blood Service [24].

Results

We compared the power of our meta-analysis to that of the previous largest meta-analysis to date. We have more than double the power of the Barrett et al. meta-analysis to find variants with a relative risk of 1.2 and approximately three times the power to detect variants with a relative risk of 1.1 [29] (Figure S1).

We used principal components analysis (PCA) [37] in order to minimize the potential impact of population stratification in our case/control sample sets. Eigenstrat 3.0 was employed to remove outliers and to subsequently calculate the principal components in the Illumina and Affymetrix assigned groups separately. The

Figure 1. QQ-plot of all previously unassociated regions in the combined meta-analysis discovery cohort.
doi:10.1371/journal.pgen.1002293.g001
| SNP            | Chr Position | Gene/Region   | Minor Allele | Frequency | Minor Allele Frequency | Minor Allele Frequency | OR     | 95% CI | P     |
|---------------|--------------|---------------|--------------|-----------|------------------------|------------------------|--------|-------|-------|
| rs539514      | 13           | LMO7          | A            | 0.499     | 0.499                  | 0.499                  | 2.91   | 2.19-10^-4 | 0.7948 |
| rs478222      | 2            | EFRA8         | T            | 0.412     | 0.412                  | 0.412                  | 3.97   | 3.79-10^-4 | 0.8732 |
| rs1094043     | 6            | 170200560     | T             | 0.46      | 0.46                   | 0.46                   | 4.49   | 4.44-10^-2 | 0.7222 |
| rs504948      | 7            | 2819567       | G             | 0.408     | 0.408                  | 0.408                  | 3.12   | 3.11-10^-4 | 0.8199 |
| rs13497837    | 8            | 172904618     | G             | 0.396     | 0.396                  | 0.396                  | 4.76   | 4.76-10^-4 | 0.8683 |
| rs5547853     | 2            | 170200560     | A             | 0.465     | 0.465                  | 0.465                  | 4.69   | 4.69-10^-4 | 0.8683 |

All values are two-sided. Allele frequencies are shown with respect to the minor allele. Combined P-values were computed with Fisher's combined P-value technique implemented in Haploview. Loci reported for the first time in this current study are annotated. Positions shown are based on Build 36 of the human genome. Minor allele frequencies are shown for the controls in the discovery cohort.

---

**Discussion**

We have carried out the largest meta-analysis of genome-wide genotyped datasets for T1D to date. The replication of three loci using the stratification-free TDT with minimal Mendelian error clearly indicates that they are not false positives due to artifacts such as uncorrected systematic error from stratification or genotyping bias.

The most significantly associated SNP (rs539514, P = 5.66 × 10^-11) resides in an intronic region of the *LMO7* (LIM domain only 7) gene on 13q22. We investigated the associated region using LocusZoom [41] and determined that it is the only gene residing within the block of linkage disequilibrium harboring the signal (Figure S3). Regional plots showing P-values, linkage disequilibrium, and recombination rate for all SNPs in Table 1 are outlined in the Figures S2, S3, S4, S5, S6, S7. *LMO7* encodes a
protein that contains multiple domains, including a calponin homology domain, a PDZ domain and a LIM domain. There are multiple LMO7 isoforms already known but their full nature and the actual extent of different isoforms remains unclear [42]. Mice with homozygous deletions of \textit{LMO7} display retinal, muscular, and growth retardation [43]. Although the function of LMO7 doesn’t clearly relate to the etiology of T1D, LMO7 is expressed in pancreatic islets and thus is a possible biological candidate at this locus [44]; however it should be noted that the retinal, muscular development and islet patterns are a key element in Emery-Dreifuss Muscular Dystrophy, caused by mutations in \textit{LMO7} [45], but bears very little similarity to T1D.

The second most significantly associated SNP (rs478222, \(P = 3.50 \times 10^{-6}\)) resides in an intronic region of the EFR3B (protein EFR3 homolog B) gene on 2p23; however the region of linkage disequilibrium is approximately 800 kb and harbors additional multiple genes, including 3MCOA1, C2orf79, CENPO, ADC13, DNAJC27, POMC, and DNMT3A. (Figure S2). A previous meta-analysis of a subset of the data used in this current study found suggestive association with T1D in the same LD block with the independent SNP, rs2165738 (\(r^2 = 0.115\)) [27]; however, we only found modest evidence of association with rs2165738 (\(P = 4.78 \times 10^{-5}\)) in our discovery cohort. There has also been association to inflammatory bowel disease [46] height [47,48] and BMI [49] reported at this locus, where in both cases the risk allele for increased height or BMI was protective for T1D risk.

The third most significantly associated SNP (rs924043, \(P = 8.06 \times 10^{-5}\)) lies in an intergenic region on 6q27, where the region of association is approximately 900 kb and harbors multiple genes including WDR27, C6orf120, PHF10, TCTE3, C6orf208, LOC154449, DLI1, FAM120B, PSMB1, TBP and PCD2 (Figure S3). In addition, despite not reaching the bar for genome wide significance, we did observe evidence for association at three additional loci (Table 1) containing the candidate genes LOC100128081, TNFRSF11B and FOSL2. Of these, it is notable that TNFRSF11B is a strongly associated locus with bone mineral density, also as a consequence of GWAS [50,51]. In addition, the locus harboring LOC100128081 has also been reported in the context of a GWAS of SLE [52]. Further work will be required to fully validate the role of these particular loci in the pathogenesis of T1D.

The Barrett \textit{et al.} meta-analysis was able to use British controls with British cases and American controls with American cases [29]. We did not have the same control data to be able to make the same comparisons. In the case of the Affymetrix analysis, some American cases were analyzed with purely British controls and, in the case of the Illumina analysis, some British cases with purely American controls. As such, we were forced to make our corrections using eigenvectors as covariates in our analysis; this will have the effect of modestly weakening the level of significance for associations that vary in allele frequency between the cases and controls, as now the case and controls will both vary with the eigenvectors to some degree. This in effect will make our analysis overly conservative with estimating the true effect of a SNP, and in fact every SNP that had a \(P\)-value less than 0.05 in the replication set did indeed have a greater effect than that which was estimated from the discovery set.

In summary, we provide convincing evidence for the existence of three additional loci associated with the T1D, adding to the repertoire of over 50 loci already demonstrated to be associated with the disease.

**Materials and Methods**

**Ethical statement**

The study was approved by the institutional review board and the ethics committee of each institution. Written informed consent...
| SNP     | CHR | Position   | Gene/Region | Effect Allele | P-Value  | OR   | References |
|---------|-----|------------|-------------|---------------|----------|------|------------|
| rs2476601 | 1   | 114179091  | PTPN22      | A             | 5.93E-80 | 1.96 | [23,25,29] |
| rs28136316 | 1   | 190803436  | RG51        | C             | 8.52E-04 | 0.89 | [34]       |
| rs3024505  | 1   | 205006527  | IL10        | A             | 2.09E-08 | 0.82 | [29]       |
| rs9633442  | 2   | 100191799  | AFF3        | C             | 5.89E-04 | 1.09 | [25]       |
| rs1990760  | 2   | 162832297  | IFHI1       | C             | 2.21E-08 | 0.87 | [25,29]    |
| rs7574865  | 2   | 191672878  | STAT4       | T             | 0.0544   | 1.06 | [35]       |
| rs3087243  | 2   | 204447164  | CTLA4       | A             | 1.42E-13 | 0.83 | [27,29]    |
| rs11711054 | 3   | 46320603   | CCR5        | G             | 0.0399   | 1.06 | [34]       |
| rs10517086 | 4   | 25694069   | 4p15.2      | NA            | NA       | NA   | [29]       |
| rs3024505  | 4   | 205006527  | IL10        | A             | 2.09E-08 | 0.82 | [29]       |
| rs12251307 | 4   | 6163501    | IL2RA       | T             | 1.22E-08 | 0.79 | [27,29]    |
| rs11258747 | 5   | 6512897    | PRKCB       | T             | 2.24E-05 | 1.13 | [27,29]    |
| rs10509540 | 5   | 90013013   | 10q23.31    | C             | 2.83E-06 | 0.88 | [29]       |
| rs3741208  | 6   | 2126350    | INS         | A             | 6.33E-08 | 1.16 | [23,25,29] |
| rs4763879  | 6   | 9801431    | 7p12.1      | NA            | NA       | NA   | [29]       |
| rs1075893  | 7   | 4280830    | GUS3        | A             | 1.18E-08 | 1.15 | [28,29]    |
| rs12251307 | 7   | 6163501    | IL2RA       | T             | 1.22E-08 | 0.79 | [27,29]    |
| rs11258747 | 8   | 6512897    | PRKCB       | T             | 2.24E-05 | 1.13 | [27,29]    |
| rs11258747 | 9   | 110368991  | SH2B3       | C             | 1.77E-21 | 0.79 | [29]       |
| rs9850566  | 9   | 98879767   | GPR183      | C             | 1.27E-03 | 1.09 | [55]       |
| rs1465788  | 10  | 68333525   | 14q24.1     | T             | 1.79E-06 | 0.87 | [29]       |
| rs1701704  | 10  | 54698754   | 12q13.2     | G             | 1.08E-30 | 1.35 | [24-27,29] |
| rs10877012 | 10  | 56448352   | CYP27B1     | NA            | NA       | NA   | [54]       |
| rs3184504  | 10  | 113068991  | SH2B3       | C             | 1.77E-21 | 0.79 | [29]       |
| rs1465788  | 11  | 2126350    | INS         | A             | 6.33E-08 | 1.16 | [23,25,29] |
| rs1701704  | 11  | 54698754   | 12q13.2     | G             | 1.08E-30 | 1.35 | [24-27,29] |
| rs10877012 | 11  | 56448352   | CYP27B1     | NA            | NA       | NA   | [54]       |
| rs9850566  | 13  | 98879767   | GPR183      | C             | 1.27E-03 | 1.09 | [55]       |
| rs1465788  | 14  | 68333525   | 14q24.1     | T             | 1.79E-06 | 0.87 | [29]       |
| rs941576   | 14  | 97568704   | 14q32.2     | G             | 0.0972   | 1.05 | [29]       |
| rs17574546 | 14  | 36689768   | RASGRFP1    | C             | 3.19E-03 | 1.09 | [57]       |
| rs3825932  | 14  | 77025301   | 15q25.1     | T             | 5.15E-05 | 0.90 | [27,29]    |
| rs2903692  | 15  | 11146284   | 16p13.13    | A             | 4.21E-15 | 0.81 | [23-25,27,29] |
| rs4788084  | 15  | 28447349   | IL27        | T             | 7.55E-04 | 0.92 | [29]       |
| rs7202877  | 15  | 73804746   | 16q23.1     | G             | 1.84E-05 | 1.19 | [29]       |
| rs2290400  | 15  | 35319766   | ORMDL3      | T             | 3.55E-03 | 0.93 | [29]       |
| rs7221109  | 15  | 36023812   | 17q21.2     | T             | 6.46E-04 | 0.92 | [29]       |
| rs478582   | 16  | 12825976   | PTPN2       | C             | 7.72E-04 | 0.92 | [24,25,27,29] |
| rs763361   | 16  | 65682622   | CD226       | T             | 1.17E-04 | 1.10 | [25]       |
| rs2304256  | 16  | 10336652   | TYK2        | NA            | NA       | NA   | [56]       |
| rs425105   | 17  | 51903321   | 19q13.32    | C             | 5.51E-06 | 0.85 | [29]       |
| rs2281808  | 17  | 1558551    | 20p13       | T             | 2.06E-06 | 0.88 | [29]       |
| rs9976767  | 18  | 42709459   | UBA5H2A     | G             | 1.69E-05 | 1.11 | [28,29]    |
| rs5753037  | 18  | 28911722   | 22q12.2     | T             | 0.0164   | 1.06 | [29]       |
| rs229514   | 18  | 35921264   | IL2R8       | A             | 3.67E-06 | 1.12 | [27,29]    |

The list of known SNPs was collected from references cited in the references column and shown below. One SNP from each locus was chosen when multiple SNPs from the same locus are known. NA in the effect allele, P-value, and OR column refers to SNPs that were not imputed in the discovery cohort. Positions shown are based on Build 36 of the human genome.

doi:10.1371/journal.pgen.1002293.t002
was obtained from each participant in accordance with institutional requirements and the Declaration of Helsinki Principles.

Samples
Cases in the discovery set were obtained from four publicly available resources and combined with those from a previous publication for the meta-analysis. Samples descriptions are available on dbGaP (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap) for the T1DGC [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000180.v1.p1], GoKinD [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000088.v1.p1], and DCCT-EDIC [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000066.v2.p1] patients. The WTCCC sample information is available from [24]. Samples from the T1D segment of the WTCCC were used as cases, while controls were derived from the 1958 Birth Cohort, UK Blood Service, Bipolar disorder, Coronary heart disease, Hypertension, and Type 2 Diabetes segments. The remaining cases used in the meta-analysis were previously described [23].

The total number of individuals used in the meta-analysis discovery set was 26,890 (9,934 cases/16,956 controls). The replication set consisted of 1,120 case-parent trios from the T1DGC and those identified through pediatric diabetes clinics in Canada. The replication set was identical to that used in Hakonarson et al, with an extension of patients identified through pediatric diabetes clinics in Montreal, Toronto, Ottawa, and Winnipeg. All individuals were of Caucasian ancestry. A comparison of the number of samples used in each discovery cohort from the current meta-analysis and those used in the previously reported meta-analysis [29] is shown in Table 3.

Table 3. A comparison of the numbers of samples used in each discovery cohort from the current meta-analysis and those used in the previously reported meta-analysis [29].

|                      | WTCCC | GoKinD/NIMH | DCCT-EDIC | T1DGC | CHOP-McGill | Totals |
|----------------------|-------|-------------|-----------|-------|-------------|--------|
| **Cases in Barrett et al meta-analysis** | 1,930 | 1,601       | 0         | 3,983 | 0           | 7,514  |
| **Controls in Barrett et al meta-analysis** | 3,342 | 1,704       | 0         | 3,999 | 0           | 9,045  |
| **Cases in current meta-analysis** | 1,920 | 1,491       | 1,363     | 4,029 | 1,131       | 9,934  |
| **Controls in current meta-analysis** | 10,308 | 0           | 0         | 6,648 | 16,956      |        |

doi:10.1371/journal.pgen.1002293.t003

Power analysis
Power analysis was performed with the genetic analysis calculator which can be found at (http://pngu.mgh.harvard.edu/~purcell/gpc/) [53]. Various assumption were made included perfect LD between the causative variant and the markers that were genotyped, an additive genetic model, a disease prevalence of 0.0033 and an alpha of 1.00001.

Genotyping, quality control, and imputation
Discovery samples from Philadelphia, Canada, T1DGC, and DCCT-EDIC were genotyped on a mixture of the Illumina HumanHap 550v1, 2, and 3, whereas samples from GoKinD and WTCCC were genotyped on the Affymetrix 500 K Chip. Sequenom iPLEX was used to replicate the findings of the meta-analysis in 1,120 affected offspring trios from the T1DGC and from Canada.

All individuals needed an individual genotyping call rate greater than 0.98 to be included in the analysis pre-imputation and individuals were removed that showed evidence of cryptic relatedness and duplication within and across cohorts using identity-by-state. SNP quality control was performed on all samples pre-imputation. SNPs were excluded from the analysis if the minor allele was below 1%, the genotyping call rate was less than 95%, or the Hardy Weinberg equilibrium P-value was less than 0.00001.

To control for population stratification, Eigenstrat 3.0 was used to compute the top 10 principal components of the individuals genotyped on the Illumina SNP chips and the Affymetrix SNP chips separately [37]. Individuals were removed from the analysis if they were 6 standard deviations away from the mean of one of the top 10 principal components. After controlling for population stratification, the estimated lambda in the Affymetrix data was 1.11 and 1.17 in the Illumina data.

Mach 1.0 was used to impute ~2.54 millions SNPs from the HapMap CEU panel for all individuals [39]. SNPs were excluded after imputation if they had a minor allele frequency less than 0.01 and an r² value less than 0.3.

Genome-wide association and meta-analysis
PLINK [38] was used to perform a logistic regression using the 10 principal components as covariates, T1D status as the outcome, and in the case of the Affymetrix cohort, an extra dummy covariate specifying WTCCC or GoKinD cohort membership. Results from the logistic regression of 2,436,110 SNPs from the Affymetrix samples and 2,062,307 SNPs from the Illumina samples separately were combined using inverse-variance meta-analysis in PLINK. A fixed effects meta-analysis was performed and 53 SNPs were chosen for replication who had a fixed effects P-value <0.00001, a Cochran’s Q statistic P-value greater than 0.05 and were not previously known to be associated with type 1 diabetes. However one of the SNPs consistently failed during the replication effort.

Supporting Information
Figure S1 Comparison of plot of power for previous and current meta-analyses. a: Plot of power (y-axis) for variants from the previously reported meta-analysis [29] with various allele frequencies (x-axis) and relative risks. Plots assume disease prevalence of 0.0033, an additive genetic model, perfect LD between causative variant and marker, and are shown for an alpha of 1.00001. b: Plot of power (y-axis) in the current meta-analysis for variants with various allele frequencies (x-axis) and relative risks. Plots assume...
disease prevalence of 0.0033, an additive genetic model, perfect LD between causative variant and marker, and are shown for an alpha of 1 x 10^{-5}.

**(DOC)**

**Figure S2** Regional plot of the EFRB associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**Figure S3** Regional plot of the LMO7 associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**Figure S4** Regional plot of the LOC100128081 associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**Figure S5** Regional plot of the Chromosome 6 associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**Figure S6** Regional plot of the Chromosome 8 associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**Figure S7** Regional plot of the Chromosome 2 FOSL2 associated region. –log10(P-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.

**(DOC)**

**References**

1. Steyn NP, Lambert EV, Tahana H (2009) Conference on "Multidisciplinary approaches to nutritional problems". Symposium on "Diabetes and health". Nutrition interventions for the prevention of type 2 diabetes. Proc Nutr Soc 68: 53–70.

2. EURODIAB ACE Study Group (2000) Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. Lancet 355: 873–876.

3. Onakomo P, Vaananen S, Karvonen M, Tuomilehto J (1999) Worldwide increase in incidence of Type I diabetes-the analysis of the data on published incidence trends. Diabetologia 42: 1395–1403.

4. Redondo MJ, Yu L, Hava M, Mackenzie T, Pyke DA, et al. (2001) Heterogeneity of type 1 diabetes: analysis of monozygotic twins in Great Britain and the United States. Diabetologia 44: 354–362.

5. Clayton DG (2009) Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 5: e1000540. doi:10.1371/journal.pgen.1000540.

6. Guo D, Li M, Zhang Y, Yang P, Eckerndode S, et al. (2004) A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. Nat Genet 36: 837–841.

7. Mirel DB, Valdes AM, Laxerrozzi LC, Reynolds KL, Erlich HA, et al. (2002) Association of IL4R haplotypes with type 1 diabetes. Diabetes 51: 3336–3341.

8. Biaso-Lauber A, Boehm B, Lang-Muritano M, Gauthier BR, Brun T, et al. (2003) Association of childhood type 1 diabetes mellitus with a variant of PAX4: possible link to beta cell regenerative capacity. Diabetologia 46: 900–905.

9. Bell GI, Horita S, Kacam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33: 176–183.

10. Bennett ST, Lacausen AM, Gough SC, Powell EE, Unfliden DE, et al. (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9: 294–292.

11. Valladis P, Bennett ST, Todd JA, Nadreau J, Grabs R, et al. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 15: 289–292.

12. Baranz BJ, Payne F, Lowe CE, Herrmann R, Healy BC, et al. (2004) Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes 53: 1848–1849.

13. Kristiansen OP, Larsen ZM, Pociot F (2000) CTLA-4 in autoimmune diseases—a general susceptibility gene to autoimmunity? Genes Immun 1: 170–184.

14. Ueda H, Howson JM, Epoitsoi L, Fawcett J, Smolic R, et al. (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423: 506–511.

15. Anjou SM, Tessler MG, Polychronakos C (2004) Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. J Clin Endocrinol Metab 89: 6257–6265.

16. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, et al. (2005) Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. J Clin Endocrinol Metab 89: 6257–6265.

17. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, et al. (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum Mol Genet 5: 1075–1080.

18. Bottini N, Masiuici I, Alonso A, Rahmonni S, Nika K, et al. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 36: 337–338.

19. Snyth D, Cooper JD, Collins JE, Heward JM, Snook H, et al. (2004) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423: 506–511.

20. Bennett ST, Lacausen AM, Gough SC, Powell EE, Unfliden DE, et al. (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 9: 294–292.

21. Valladis P, Bennett ST, Todd JA, Nadreau J, Grabs R, et al. (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 15: 289–292.

22. Baranz BJ, Payne F, Lowe CE, Herrmann R, Healy BC, et al. (2004) Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes 53: 1848–1849.

23. Kristiansen OP, Larsen ZM, Pociot F (2000) CTLA-4 in autoimmune diseases—a general susceptibility gene to autoimmunity? Genes Immun 1: 170–184.

24. Ueda H, Howson JM, Epoitsoi L, Fawcett J, Smolic R, et al. (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423: 506–511.

25. Anjou SM, Tessler MG, Polychronakos C (2004) Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. J Clin Endocrinol Metab 89: 6257–6265.

26. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, et al. (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum Mol Genet 5: 1075–1080.

27. Bottini N, Masiuici I, Alonso A, Rahmonni S, Nika K, et al. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 36: 337–338.

28. Snyth D, Cooper JD, Collins JE, Heward JM, Snook H, et al. (2004) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423: 506–511.
GWAS Meta-Analysis of Six Type 1 Diabetes Cohorts