SHORT COMMUNICATION

Confirmation of novel type 1 diabetes risk loci in families

J. D. Cooper · J. M. M. Howson · D. Smyth ·
N. M. Walker · H. Stevens · J. H. M. Yang · J.-X. She ·
G. S. Eisenbarth · M. Rewers · J. A. Todd · B. Akolkar ·
P. Concannon · H. A. Erlich · C. Julier · G. Morahan ·
J. Nerup · C. Nierras · F. Pociot · S. S. Rich ·
the Type 1 Diabetes Genetics Consortium

Abstract
Aims/hypothesis Over 50 regions of the genome have been associated with type 1 diabetes risk, mainly using large case/control collections. In a recent genome-wide association (GWA) study, 18 novel susceptibility loci were identified and replicated, including replication evidence from 2,319 families. Here, we, the Type 1 Diabetes Genetics Consortium (T1DGC), aimed to exclude the possibility that any of the 18 loci were false-positives due to population stratification by significantly increasing the statistical power of our family study.

Methods We genotyped the most disease-predicting single-nucleotide polymorphisms at the 18 susceptibility loci in 3,108 families and used existing genotype data for 2,319 families from the original study, providing 7,013 parent–child trios for analysis. We tested for association using the transmission disequilibrium test.

Results Seventeen of the 18 susceptibility loci reached nominal levels of significance ($p < 0.05$) in the expanded family collection, with 14q24.1 just falling short ($p = 0.055$). When we allowed for multiple testing, ten of the 17 nominally

J. D. Cooper (✉) · J. M. M. Howson · D. Smyth · N. M. Walker ·
H. Stevens · J. H. M. Yang · J. A. Todd
Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes
and Inflammation Laboratory, Department of Medical Genetics,
Cambridge Institute for Medical Research,
University of Cambridge, Addenbrooke's Hospital,
Cambridge CB2 0YX, UK
e-mail: jason.cooper@cimr.cam.ac.uk

J.-X. She
Medical College of Georgia,
Center for Biotechnology and Genomic Medicine,
Augusta, GA, USA

G. S. Eisenbarth · M. Rewers
Barbara Davis Center for Childhood Diabetes,
University of Colorado Health Sciences Center,
Denver, CO, USA

B. Akolkar
Division of Diabetes, Endocrinology, and Metabolic Diseases,
The National Institute of Diabetes and Digestive and Kidney
Diseases (NIDDK), National Institutes of Health,
Bethesda, MD, USA

P. Concannon
Department of Biochemistry and Molecular Genetics,
University of Virginia,
Charlottesville, VA, USA

P. Concannon · S. S. Rich
Center for Public Health Genomics, University of Virginia,
Charlottesville, VA, USA

H. A. Erlich
Roche Molecular Systems,
Pleasanton, CA, USA

C. Julier
Inserm UMR-S958, Faculté de Médecine Denis-Diderot,
Paris and Centre National de Génotypage,
Evry, France

C. Julier
Université Paris 7 Denis-Diderot,
Paris, France

significant loci reached the required level of significance ($p < 2.8 \times 10^{-3}$). All susceptibility loci had consistent direction of effects with the original study.

Conclusions/interpretation The results for the novel GWA study-identified loci are genuine and not due to population stratification. The next step, namely correlation of the most disease-associated genotypes with phenotypes, such as RNA and protein expression analyses for the candidate genes within or near each of the susceptibility regions, can now proceed.

Abbreviations

GWA	Genome-wide association
HWE	Hardy–Weinberg equilibrium
SNP	Single nucleotide polymorphism
T1DGC	Type 1 Diabetes Genetics Consortium

Introduction

The publication of the first type 1 diabetes locus found by a genome-wide association (GWA) study in 2006 (*IFIH1*) [1] heralded a new era in susceptibility locus discovery in this common autoimmune disease. Over 50 susceptibility loci have now been identified (www.t1dbase.org). Eighteen of these were identified by Barrett et al. [2] in a GWA meta-analysis of 7,514 cases and 9,045 controls (meta-analysis $p < 1 \times 10^{-6}$) and confirmed in 4,267 cases, 4,670 controls and 2,319 affected sib-pair families (providing 4,342 parent–child trios; replication $p < 0.01$; discovery and replication $p < 5 \times 10^{-8}$) [2].

G. Morahan
Centre for Diabetes Research, The Western Australian Institute for
Medical Research, and Centre for Medical Research,
University of Western Australia,
Perth, WA, Australia

J. Nerup
Steno Diabetes Center and Hagedorn Research Institute,
Gentofte, Denmark

C. Nierras
Juvenile Diabetes Research Foundation,
New York, NY, USA

F. Pociot
Glostrup Research Institute, Glostrup Hospital,
Glostrup, Denmark

S. S. Rich
Department of Public Health Sciences, Division of Biostatistics
and Epidemiology, University of Virginia,
Charlottesville, VA, USA

However, in the family component of the replication samples, eight of the confirmed 18 susceptibility loci failed to reach nominal levels of significance ($p < 0.05$; inferred from the reported 95% confidence intervals for the relative risks and assuming two-sided significance tests). Although replication was based on the combined evidence from case/control and family collections, and no evidence of population stratification in the case/control collection had been found previously [2, 3], family-based evidence, if possible, remains important in order to demonstrate that these associations did not arise through population stratification bias [4]. Such a bias can occur when a single nucleotide polymorphism (SNP) differs in allele frequency across subgroups of the population and risk of disease differs between these subgroups.

Based on the number of case/control and parent–child trio replication samples used in Barrett et al. [2], if we assume that the parent–child trios equate to an equal number of cases and controls, the power of the case/control and family replication sets would have been similar and the potential impact of winner's curse (the upward bias of the effect size of the initial finding) on replication would not differ between the replication sample sets. However, in type 1 diabetes, the effects (as measured by relative risk) of non-HLA loci tend to be smaller in affected sib-pair families [2, 5], which are enriched for type 1 diabetes with a higher frequency of high-risk HLA genotypes. Consequently, when the family component of the replication samples used in Barrett et al. [2] is considered in isolation, the 2,319 affected sib-pair families are likely to have been underpowered (too few samples analysed) to replicate the initial associations. Therefore, in the present study, we genotyped the best disease-predicting SNPs at the 18 susceptibility loci [2] in an additional 3,108 families (providing 2,801 parent–child trios to the analysis) from the Type 1 Diabetes Genetics Consortium (T1DGC) and the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory. The analyses of these additional families, combined with the original 2,319 families [2], provided protection from population stratification bias, and increased power to provide further replication support for the associations of these 18 susceptibility loci [2].

Methods

Subjects After the additional genotyping of 3,108 families (2,322 families of white European ancestry and providing at least one parent–child trio; electronic supplementary material [ESM] Table 1), we had a collection of 5,427 families (including 2,319 families previously genotyped [2]). All families were collected with appropriate informed consent. We analysed 4,429 families of white European ancestry and providing one or more parent–child trios (ESM Table 1).

Genotyping The best disease-predicting SNPs at the 18 susceptibility loci [2] were genotyped in the additional family samples using the TaqMan 5′ nuclease assay (Applied Biosystems, Warrington, UK) according to the manufacturer's protocol. Genotyping was performed blind to disease status and double scored to minimise error. Genotype frequencies were tested for deviation from Hardy–Weinberg equilibrium (HWE), and genotype checks were conducted for SNPs that deviated from HWE. We note that disease association can result in deviation from HWE in affected offspring and parents of affected offspring, who are not representative of the general population. The same genotyping technology and protocols had been applied in Barrett et al. [2] for the replication samples.

Statistical analysis All statistical analyses were performed in either Stata (www.stata.com) or R (www.r-project.org). In R, we used the snpStats package available from the Bioconductor project (www.bioconductor.org), and, in Stata, we used some additional routines available from www-gene. cimr.cam.ac.uk/clayton/software.

The family-based power to replicate the 18 type 1 diabetes susceptibility loci [2] is reported in ESM Table 2. Based on the odds ratios from the case/control component of the replication samples in Barrett et al. [2], which are not subject to winner's curse, the expanded family collection is well powered, except for 17q21.2/*CCR7* (53.4% power at $\alpha=0.05$; 17.2% power at $\alpha=2.8\times10^{-3}$, which corresponds to the Bonferroni adjustment of the 0.05 significance level for the 18 independent tests; ESM Table 2). We have greater than 90% power at $\alpha=0.05$ for 17/18 loci, and greater than 80% power at $\alpha=2.8\times10^{-3}$ for 14/18 loci (17/18 have greater than 60% power at $\alpha=2.8\times10^{-3}$).

The best disease-predicting SNPs at the 18 susceptibility loci were analysed using the transmission disequilibrium test, except for the chromosome X locus, rs2664170 Xq28/*GAB3*, which was analysed using the method proposed by Clayton [6]. As we were attempting to replicate the associations reported in the case/control component of the replication samples analysed in Barrett et al. [2], we performed one-sided significance tests. We tested for population heterogeneity in SNP genotype frequencies across unaffected parents using Kruskal-Wallis one-way analysis of variance. We tested for population heterogeneity in disease association, after generating pseudo-controls [7], by testing the addition of the genotype–population interaction term to the conditional logistic regression model of disease status on genotype and population. Parent-of-origin and imprinting effects were tested using the Wallace et al. extension of the Weinberg method [8, 9].

Results

As no *p* values have been reported previously for the 18 novel susceptibility loci in the family component of the

replication samples [2], we reanalysed the original data. We excluded 312 families because of either non-white European ancestry based on updated sample information or not providing at least one parent–child trio. Seven of the 18 loci failed to reach $p<0.05$ in these 2,107 families (providing 4,212 parent–child trios; Table 1). In other words, 11 of the 18 loci reach at least nominal levels of significance. If we applied a Bonferroni adjustment for the 18 independent tests, 15 loci failed to reach $p<2.8\times10^{-3}$.

The inclusion of the additional 2,322 families (providing 2,801 trios; 786 families excluded) increased the number of susceptibility loci replicated at $p<0.05$ from 11 to 17 of the 18 loci. Only *ZFP36L1, C14orf181*/14q24.1 ($p=0.055$) failed to reach $p<0.05$ (Table 1). The number of susceptibility loci replicated at $p<2.8\times10^{-3}$ increased from three to ten (Table 1). Importantly, all of the susceptibility loci had consistent direction of effects with the case/control and family replication samples reported in Barrett et al. [2], and there was no evidence of heterogeneity in the disease associations across family collections, despite there being significant SNP genotype frequency differences (ESM Table 3). The difference in SNP genotype frequencies across family collections was not surprising given that Europe is a large and diverse collection of countries. For example, we have a large number of families from Finland, a genetically isolated population, which exhibits many and large differences in common SNP allele frequencies.

We tested the 17 autosomal loci for parent-of-origin and imprinting effects; only *COBL*/7p12.1 showed any evidence of biased maternal transmission, $p=1.1\times10^{-3}$ (ESM Table 4). However, this needs to be replicated in an independent dataset.

Discussion

In the expanded family collection, only one of the previously confirmed susceptibility loci failed to reach nominal levels of significance, *ZFP36L1, C14orf181*/14q24.1, as the *p* value was just above 0.05. All of the susceptibility loci had consistent direction of effects with the case/control component of the replication samples reported in Barrett et al. [2] (ESM Table 5), and even with our over-conservative threshold for multiple testing, given the very strong prior information that these were true effects [2], ten loci remained significant after the adjustment for multiple testing. This study clearly demonstrates that additional replication families were required for the 18 susceptibility loci to reach nominal levels of significance and consequently that the previously reported associations (discovery and replication $p<5\times10^{-8}$) with odds ratios often less than 1.15 ([2]; ESM Table 5) did not arise through population stratification bias, thereby further validating the case/control collection (results).

Table 1 A summary of the analysis of the 18 novel type 1 diabetes susceptibility loci discovered in Barrett et al. [2]

Region/candidate gene rs number	MAF	Reanalysis of Barrett et al. (2,107 families)		Additional 2,322 families		Combined	
		RR (95% CI)	One-tailed p value	RR (95% CI)	One-tailed p value	RR (95% CI)	One-tailed p value
1q32.1/*IL10* rs3024505 C>T	0.149	0.94 (0.86, 1.03)	0.10	0.83 (0.74, 0.93)	4.4×10^{-3}	0.89 (0.83, 0.96)	9.5×10^{-4}
4p15.2 rs10517086 G>A	0.317	1.07 (1.00, 1.15)	0.020	1.05 (0.96, 1.14)	0.14	1.07 (1.02, 1.13)	8.5×10^{-3}
6q22.32/*C6orf173* rs9388489 A>G	0.471	1.04 (0.98, 1.11)	0.12	1.09 (1.01, 1.18)	0.015	1.06 (1.01, 1.12)	0.010
7p12.1/*COBL* rs4948088 C>A	0.0427	0.96 (0.81, 1.14)	0.33	0.77 (0.60, 0.99)	0.019	0.87 (0.76, 1.01)	0.046
7p15.2/*SKAP2* rs7804356 T>C	0.231	0.99 (0.92, 1.06)	0.36	0.86 (0.78, 0.95)	1.2×10^{-3}	0.94 (0.89, 1.00)	0.021
9p24.2/*GLIS3* rs7020673 G>C	0.490	0.96 (0.90, 1.02)	0.12	0.92 (0.85, 1.00)	0.023	0.95 (0.90, 1.00)	0.017
10q23.31/*RNLS* rs10509540 T>C	0.251	0.82 (0.76, 0.88)	2.3×10^{-8}	0.76 (0.69, 0.84)	2.1×10^{-8}	0.80 (0.75, 0.84)	9.0×10^{-15}
12p13.31/*CD69* rs4763879 G>A	0.364	1.11 (1.04, 1.19)	6.0×10^{-4}	1.10 (1.01, 1.20)	0.014	1.11 (1.05, 1.17)	5.0×10^{-5}
14q24.1/*ZFP36L1,C14orf181* rs1465788 G>A	0.282	0.95 (0.88, 1.02)	0.075	0.96 (0.88, 1.05)	0.19	0.95 (0.90, 1.01)	0.055
14q32.2/*C14orf64* rs4900384 A>G	0.281	1.09 (1.02, 1.17)	6.5×10^{-3}	1.08 (0.99, 1.17)	0.034	1.08 (1.02, 1.14)	1.4×10^{-3}
16p11.2/*IL27* rs4788084 G>A	0.425	0.92 (0.87, 0.98)	6.5×10^{-3}	0.91 (0.84, 0.99)	0.012	0.92 (0.88, 0.97)	4.6×10^{-4}
16q23.1/*CTRB2* rs7202877 T>G	0.112	1.05 (0.95, 1.16)	0.18	1.20 (1.06, 1.35)	1.6×10^{-3}	1.09 (1.01, 1.18)	6.5×10^{-3}
17q12/*GSDMB,ORMDL3* rs2290400 G>A	0.479	0.92 (0.87, 0.98)	6.0×10^{-3}	0.84 (0.77, 0.90)	4.3×10^{-6}	0.89 (0.85, 0.94)	1.8×10^{-6}
17q21.2/*CCR7* rs7221109 C>T	0.342	0.93 (0.87, 0.99)	0.016	0.88 (0.81, 0.96)	1.2×10^{-3}	0.91 (0.86, 0.96)	1.8×10^{-4}
19q13.32/*PRKD2* rs425105 A>G	0.150	0.88 (0.80, 0.97)	4.9×10^{-3}	0.89 (0.79, 0.99)	0.020	0.88 (0.82, 0.95)	4.0×10^{-4}
20p13/*SIRPG* rs2281808 C>T	0.363	0.91 (0.85, 0.97)	2.2×10^{-3}	0.88 (0.81, 0.96)	2.3×10^{-3}	0.89 (0.85, 0.94)	2.6×10^{-5}
22q12.2/*HORMAD2* rs5753037 C>T	0.400	1.09 (1.02, 1.17)	4.0×10^{-3}	1.11 (1.02, 1.20)	7.5×10^{-3}	1.10 (1.05, 1.16)	7.0×10^{-5}
Xq28/*GAB3* rs2664170 A>G	0.331	1.06 (1.00, 1.12)	0.019	1.03 (0.95, 1.10)	0.26	1.04 (1.00, 1.09)	0.025

Note that we excluded families with non-white European ancestry from the families analysed by Barrett et al. [2] based on updated sample information and that the major and minor alleles were derived from unaffected parents from Warren 1 Diabetes UK (DUK) collection.

MAF, minor allele frequency in unaffected parents from DUK collection and for Xq28/*GAB3*, in unaffected mothers from DUK collection

After unequivocal replication of type 1 diabetes loci, the next steps involve dense SNP mapping in even larger sample sets and experiments analysing genotype–phenotype associations. For example, studying correlations between type 1 diabetes SNP risk alleles and haplotypes and expression of genes at the RNA and protein levels [10] can identify which genes in the associated regions are more likely to be causal. Consequently, genes with both positional and functional evidence for a role in disease aetiology can reveal the pathways and early precursors or biomarkers underlying the pathogenesis of type 1 diabetes.

Duality of interest   The authors declare that there is no duality of interest associated with this manuscript.

References

1. Smyth DJ, Cooper JD, Bailey R et al (2006) 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
2. Barrett JC, Clayton DG, Concannon P et al (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 41:703–707
3. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678
4. Qu HQ, Bradfield JP, Li Q et al (2010) In silico replication of the genome-wide association results of the type 1 Diabetes Genetics Consortium. Hum Mol Genet 19:2534–2538
5. Howson JM, Barratt BJ, Todd JA, Cordell HJ (2005) Comparison of population- and family-based methods for genetic association analysis in the presence of interacting loci. Genet Epidemiol 29:51–67
6. Clayton DG (2009) Sex chromosomes and genetic association studies. Genome Med 1:110
7. Cordell HJ, Clayton DG (2002) A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet 70:124–141
8. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG (2010) The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat Genet 42:68–71
9. Weinberg CR (1999) Methods for detection of parent-of-origin effects in genetic studies of case–parents triads. Am J Hum Genet 65:229–235
10. Dendrou CA, Plagnol V, Fung E et al (2009) Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. Nat Genet 41:1011–1015