Divergent genetic effects for type 1 and type 2 diabetes at overlapping association signals

Jamie R.J. Inshaw[1], Carlo Sidore[2], Francesco Cucca[2], M. Irina Stefana[1], Daniel J. M. Crouch[1], Mark I. McCarthy[3,$], Anubha Mahajan[3$], John A. Todd[1*]*

[1] JDRF/ Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] Institute for Research in Genetics and Biomedicine (IRGB), Sardinia, Italy

[3] Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Current address: [$] Genentech, 1 DNA Way, South San Francisco, CA, 94080

*Authors contributed equally to the work

Co-corresponding authors: Jamie Inshaw and John Todd

JDRF/ Wellcome Diabetes and Inflammation Laboratory

Wellcome Centre for Human Genetics

Nuffield Department of Medicine

NIHR Biomedical Research Centre

University of Oxford

Roosevelt Drive

Oxford

United Kingdom

OX3 7BN

Word Count: 1,492

Abstract

Aims/hypothesis: Given the potential shared aetiology between type 1 and type 2 diabetes, we aimed to identify any genetic regions associated with both diseases. For associations where there is a shared signal and the allele that increases risk to one disease also increases risk to the other, inference about shared aetiology could be made, with the potential to develop therapeutic strategies to treat or prevent both diseases simultaneously. Alternatively, if a genetic signal colocalises with divergent effect directions, it could provide valuable biological insight into how the association affects the two diseases differently.

Methods: Using publicly available type 2 diabetes summary statistics from a meta-analysis of European ancestry individuals (74,124 cases and 824,006 controls) and type 1 diabetes summary statistics from a meta-analysis of studies on individuals from the UK and Sardinia (7,467 cases and 10,218 controls), we identified all regions of 0.5 Mb that contained variants associated with both diseases (false discovery rate<0.01). In each region, we performed forward stepwise logistic regression to identify independent association signals, then examined colocalisation of each type 1 diabetes signal with each type 2 diabetes signal using *coloc*. Any association with a colocalisation posterior probability of ≥0.9 was considered a genuine shared association with both diseases.

Results: Of the 81 association signals from 42 genetic regions that showed association with both type 1 and type 2 diabetes, four association signals colocalised between both diseases (posterior probability ≥0.9): (i) chromosome 16q23.1, near Chymotripsinogen B1 (*CTRB1*) / Breast Cancer Anti-Estrogen Resistance Protein 1 (*BCAR1*), which has been previously identified; (ii)

chromosome 11p15.5, near the Insulin (INS) gene; (iii) chromosome 4p16.3, near Transmembrane protein 129 (*TMEM129*), and (iv) chromosome 1p31.3, near Phosphoglucomutase 1 (*PGM1*). In each of these regions, the effect of genetic variants on type 1 diabetes was in the opposite direction to the effect on type 2 diabetes. Use of additional datasets also supported the previously identified colocalisation on chromosome 9p24.2, near the GLIS Family Zinc Finger Protein 3 (*GLIS3*) gene, in this case with a concordant direction of effect.

Conclusions/interpretation: That four of five association signals that colocalise between type 1 diabetes and type 2 diabetes are in opposite directions suggests that genetically identified targets to treat one disease may have limited efficacy in treating the other disease.

62 ## Research in Context

63 **What is already known about this subject?**

64 - Other than insulin, there are currently no treatments for both type 1 and
65 type 2 diabetes.
66 - Findings that genetic variants near the GLIS3 gene increase risk of both
67 type 1 and type 2 diabetes have indicated shared genetic mechanisms at
68 the level of the pancreatic β cell.

69 **What is the key question?**

70 - By examining chromosome regions associated with both diseases, are
71 there any more variants that affect risk of both diseases and could
72 support common mechanisms and repositioning of therapeutics between
73 the diseases?

74 **What are the new findings?**

75 - At current sample sizes, there is evidence that five genetic variants in
76 different chromosome regions impact risk of developing both diseases.
77 - However, four of these variants have the opposite direction of effect in
78 type 1 diabetes compared to type 2 diabetes, with only one, near *GLIS3*,
79 having a concordant direction of effect.

80 **How might this impact on clinical practise in the foreseeable future?**

81 - Genetic findings have furthered research in type 1 and type 2 diabetes
82 independently, and suggest therapeutic strategies. However, our current
83 investigation into their shared genetics suggests that repositioning of
84 current type 2 diabetes treatments into type 1 diabetes may not be
85 straightforward.

86 **Introduction**

87 There is a genetic component to both type 1 and type 2 diabetes, with
88 approximately 60 chromosome regions associated with type 1 diabetes [1] and
89 over 200 associated with type 2 diabetes [2] at genomewide significance.
90 Examination of regions associated with both diseases could lead to uncovering
91 signals that simultaneously alter disease risk for both diseases, termed
92 colocalisation. Uncovering colocalising signals could provide biological insights
93 into shared disease mechanisms, and, in the case of signals that increase the risk
94 of both type 1 and type 2 diabetes, potentially reveal therapeutic targets effective
95 for both diseases. A recent analysis suggested that the same genetic variant is
96 altering risk of both type 1 and type 2 diabetes in five regions, near Centromere
97 Protein W (*CENPW*), Chymotripsinogen B1 (*CTRB1*)/Breast Cancer Anti-
98 Estrogen Resistance Protein 1 (*BCAR1*), GLIS Family Zinc Finger Protein 3
99 (*GLIS3*), B-cell Lymphoma 11A (*BCL11A*) and Thyroid Adenoma-Associated
100 Protein (*THADA*) [3].
101 Here, we identified all regions across the genome that showed evidence of
102 association to both type 1 and type 2 diseases at false discovery rate (FDR)
103 <0.01, and assessed colocalisation between the two diseases in each of these
104 regions. Furthermore, to account for the possibility of multiple causal variants
105 within an associated region, we extended the analysis to investigate
106 conditionally-independent associations within each region, to assess whether
107 any of the associations with one disease colocalised with any associations in the
108 other.

109 **Methods**

110 Type 1 diabetes meta-analysis summary statistics were generated using genome
111 wide association study (GWAS) data from 3,983 cases and 3,994 controls from
112 the UK genotyped using the Illumina Infinium 550K platform, 1,926 cases and
113 3,342 controls from the UK genotyped using the Affymetrix GeneChip 500K
114 platform and 1,558 cases and 2,882 controls from Sardinia genotyped using the
115 Affymetrix 6.0 and Illumina Omni Express platforms, totalling 7,467 cases and
116 10,218 controls (**Supplementary Table 1**). Genotypes were imputed using the
117 Haplotype Reference Consortium (HRC) reference panel for the UK collections [4],
118 and a custom Sardinian reference panel of 3,514 Sardinians for the Sardinian
119 collection (**Electronic supplementary material (ESM)**).
120 Summary statistics for type 2 diabetes were from 74,124 cases and 824,006
121 controls of European ancestry, imputed up to the HRC reference panel [2].
122 Regions associated with both diseases were identified by selecting all variants
123 with type 1 diabetes and a type 2 diabetes association with FDR<0.01 (**ESM**). In
124 each such region, windows of approximately 0.5 Mb were taken to examine
125 colocalisation. Within these regions, forward stepwise logistic regressions were
126 carried out for both diseases, and conditional summary statistics were obtained
127 so each conditionally-independent signal from both diseases could be tested
128 against each other for colocalisation (**ESM**).
129 Colocalisation of signals was assessed using *coloc* [5], a Bayesian method that
130 enumerates the posterior probability that the association signals in a region are
131 shared between traits. The prior probability of association with either disease
132 was taken to be 1×10^{-4} and the prior probability that the association signal is
133 shared across traits was taken to be 5×10^{-6}, as recommended [6]. The threshold to

134 consider signals as colocalising was conservatively chosen at a posterior

135 probability ≥0.9.

136 **Results**

137 Including conditionally-independent association signals, 81 colocalisation
138 analyses were carried out across 42 chromosomal regions that showed
139 association to both diseases (**Supplementary Table 2**).
140 Four signals showed evidence of colocalisation (**Supplementary Table 3**), with
141 the first on chromosome 16q23.1, near *CTRB1* and *BCAR1*, having a posterior
142 probability of colocalisation (H4PP, hereafter) of 0.98 (**Supplementary Figure**
143 **1**). The minor A allele at the type 2 diabetes index variant, rs72802342 (C>A), is
144 protective for type 2 diabetes (OR=0.87, $p=4.00\times10^{-32}$) and susceptible for type 1
145 diabetes (OR=1.33, $p=5.81\times10^{-10}$).
146 The second was on chromosome 11p15.5, near *INS*, where the primary type 2
147 diabetes association colocalised with the secondary type 1 diabetes association
148 (H4PP=0.95, **Supplementary Figure 2**). The direction of effect was opposite,
149 with the minor A allele at the type 2 diabetes index variant, rs4929965 (G>A),
150 associated with susceptibility to type 2 diabetes (OR=1.07, $p=4.80\times10^{-25}$) and
151 protection from type 1 diabetes (OR=0.87, $p=1.89\times10^{-5}$).
152 Thirdly, a region on chromosome 4p16.3 colocalised (H4PP=0.97) (**Figure 1**),
153 near Transmembrane protein 129 (*TMEM129*). The minor T allele at the type 2
154 diabetes index variant, rs56337234 (C>T), was associated with decreased risk of
155 type 2 diabetes (OR=0.94, $p=1.4\times10^{-17}$) and increased risk of type 1 diabetes
156 (OR=1.12, $p=4.07\times10^{-6}$).
157 Finally, a region on chromosome 1p31.3, near Phosphoglucomutase 1 (*PGM1*),
158 colocalised (H4PP=0.91, **Supplementary Figure 3**), with the minor T allele at
159 the type 2 diabetes index variant rs2269247 (C>T) decreasing risk of type 2

160 diabetes (OR=0.96, $p=4.6\times10^{-7}$) and increasing risk of type 1 diabetes (OR=1.15,
161 $p=1.9\times10^{-6}$) (**Table 1**).
162 We did not replicate the finding that the chromosome regions near *CENPW,*
163 *GLIS3, BCL11A or THADA* colocalised between type 1 diabetes and type 2
164 diabetes (H4PP *CENPW*=0.12, *GLIS3*=0.29, *BCL11A*=0.28, *THADA* not examined as
165 no association (FDR=0.07) existed with type 1 diabetes in the region). To
166 investigate these discrepancies, we examined two other large type 2 diabetes
167 meta-analyses: a trans-ethnic study including 1,407,282 individuals [7] and a study
168 of 433,540 individuals of East Asian ancestry [8]. For the *CENPW* and *BCL11A*
169 regions, the type 2 diabetes signal is consistent with at least one of the other
170 GWAS studies (measured by linkage disequilibrium (LD) in Europeans to the
171 other study index variants, **Supplementary Table 4**); and, the type 1 diabetes
172 index variant is not in strong LD ($r^2<0.41$) with any of the index variants for type
173 2 diabetes across the three GWAS studies. However, at *GLIS3*, there appears to be
174 a distinct signal in the European study [2] compared to the trans-ethnic and East
175 Asian type 2 diabetes studies ($r^2=0.65$), and the index variants from these two
176 studies are in higher r^2 with the type 1 diabetes signal in our analysis ($r^2=0.68$),
177 and even higher r^2 with the index variant from a larger T1D genetic analysis [1]
178 ($r^2=0.99$), indicating that the signal near *GLIS3* does colocalise between type 1
179 and type 2 diabetes with concordant direction of effect, as previously identified [9].

180 **Discussion**

181 Using genetic association summary statistics from European populations, we
182 identified 42 regions that showed association with both type 1 and type 2
183 diabetes, with 81 conditionally-independent association signals across those
184 regions. Four signals (near *CTRB1/BCAR1*, *INS*, *TMEM129* and *PGM1*) colocalised
185 between the two diseases, including a signal at the complex *INS* region for the
186 first time, which was achieved by examining conditional summary statistics.
187 However, in all four cases, the allele increasing risk for one disease was
188 protective against the other. Examination of additional trans-ethnic and East
189 Asian type 2 diabetes genetic analyses, indicated that a fifth association, near
190 *GLIS3*, is likely to colocalise between diseases, with concordant direction of
191 effect.

192 It is perhaps unsurprising that, in the current analysis, no additional regions
193 were identified where the same variant and allele are associated with increased
194 risk to both diseases, given the distinct mechanisms underlying β-cell
195 dysfunction and cell death between the two diseases [10]. However, the type 1
196 diabetes GWAS was much smaller than the type 2 diabetes analysis, and
197 therefore had less statistical power to detect more subtle genetic effects. If a type
198 1 diabetes GWAS were to be performed with similar power to the type 2 diabetes
199 GWAS, more regions might colocalise between the two diseases, but the effects of
200 these additional regions on type 1 diabetes would be small compared to the
201 currently known associations.

202 That four of five colocalisation signals had opposite directions of effect implies a
203 complex genetic relationship between the two diseases. Whilst the directional
204 discordance offers little hope for effective treatments for both diseases

205 simultaneously at these particular targets, it can offer biological insight into the
206 disease pathways that these regions act upon.
207 We did not replicate the findings that the associations near *BCL11A*, *CENPW* and
208 *THADA* colocalise between the two diseases [3], despite overlapping samples and
209 similar numbers of cases and controls in the type 1 diabetes GWAS. This is likely
210 due to three reasons: i) the previous study [3] examined colocalisation using a
211 much weaker association signal, for example, the colocalisation near *THADA* was
212 based on a type 1 diabetes association p-value of 0.01; ii) we used a more
213 stringent prior for colocalisation between the two diseases, as recently
214 suggested [6] (5×10^{-6} vs. 1×10^{-5}); and iii) we used a more stringent posterior
215 probability threshold to declare colocalisation (0.9 vs. 0.5). Our increased
216 stringency compared to the previous analysis [3], whilst increasing the probability
217 that any identified shared signals will be true positives, may have decreased our
218 sensitivity to detect all colocalisations. For example, by examining other large
219 type 2 diabetes GWAS analyses and a larger type 1 diabetes genetic analysis, we
220 conclude that the association near *GLIS3* likely does colocalise between the two
221 diseases, and with concordant directions of effect.
222 In conclusion, with current GWAS sample sizes, just five associations appear to
223 colocalise between type 1 diabetes and type 2 diabetes, four with opposing
224 direction of effect. This indicates that the depth of genetically identified
225 therapeutic targets to treat or prevent both diseases simultaneously may be
226 limited.

227 ## Tables and Figures

228 **Table 1**: Regions with a colocalisation posterior probability of ≥0.9 between type 1 diabetes and type 2 diabetes. Summary statistics

229 given from the perspective of the index type 2 diabetes variant and with respect to the ALT allele. T2D= type 2 diabetes, T1D= type 1

230 diabetes. r^2 obtained from 1000 Genomes Project European population.

rsID	Proximal gene(s)	chr	pos (gr37)	REF	ALT	T2D conditional on	T2D OR (95% CI)	T2D p	r^2 to T1D index variant (T1D index variant)	T1D conditional on	T1D OR (95% CI)	T1D p
rs2269247	*PGM1*	1p31.3	64107284	C	T	-	0.96 (0.94, 0.97)	4.60×10^{-7}	0.86 (rs2269246)	-	1.14 (1.08, 1.22)	1.94×10^{-6}
rs56337234	*TMEM129*	4p16.3	1784403	C	T	-	0.94 (0.93, 0.96)	1.40×10^{-17}	0.97 (rs6829631)	-	1.12 (1.07, 1.18)	4.07×10^{-6}
rs4929965	*INS*	11p15.5	2197286	G	A	rs11042596, rs555759341, rs571342427, rs10838787	1.07 (1.06, 1.09)	4.80×10^{-25}	0.97 (rs7119275)	rs689	0.87 (0.81, 0.93)	1.89×10^{-5}
rs72802342	*CTRB1/ BCAR1*	16q23.1	75234872	C	A	rs3115960	0.87 (0.85, 0.89)	4.00×10^{-32}	0.89 (rs55993634)	-	1.33 (1.22, 1.46)	5.81×10^{-10}

231 ## Figure Legends

232 **Figure 1:** Manhattan plots showing $-\log_{10}$p-value of association for each variant
233 by position along chromosome 4 (genome build 37) in the *TMEM129* region for
234 type 2 diabetes (top panel) and type 1 diabetes (bottom panel), coloured by r^2 to
235 the type 2 diabetes index variant, rs56337234.

236 **Supplementary Figure 1:** Manhattan plots showing $-\log_{10}$p-value of
237 association for each variant by position along chromosome 16 (genome build 37)
238 in the *CTRB1/BCAR1* region for type 2 diabetes (top panel) and type 1 diabetes,
239 conditional on primary signal index variant rs689 (bottom panel), coloured by r^2
240 to the type 2 diabetes index variant, rs72802342.

241 **Supplementary Figure 2:** Manhattan plots showing $-\log_{10}$p-value of
242 association for each variant by position along chromosome 11 (genome build 37)
243 in the *INS* region for type 2 diabetes (top panel) and type 1 diabetes (bottom
244 panel), coloured by r^2 to the type 2 diabetes index variant, rs4929965.

245 **Supplementary Figure 3:** Manhattan plots showing $-\log_{10}$p-value of
246 association for each variant by position along chromosome 1 (genome build 37)
247 in the *PGM1* region for type 2 diabetes (top panel) and type 1 diabetes (bottom
248 panel), coloured by r^2 to the type 2 diabetes index variant, rs2269247.

249

Acknowledgments

We gratefully acknowledge all participants for allowing the analysis of their genotypic and phenotypic data.

253 **Data availability**

254 Type 1 diabetes summary statistics will be available through GWAS catalog

255 (https://www.ebi.ac.uk/gwas/). Type 2 diabetes summary statistics are already

256 publicly available.

Funding

This work was funded by the Juvenile Diabetes Research Foundation (JDRF) (9-2011-253, 5-SRA-2015-130-A-N) and Wellcome (091157, 107212) to the Diabetes and Inflammation Laboratory, University of Oxford.

Additional funding was obtained from the Wellcome (090532, 098381, 106130, 212259) and the National Institute of Diabetes and Digestive and Kidney diseases (U01-DK105535).

Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. Financial support was provided by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Authors' relationships and activities

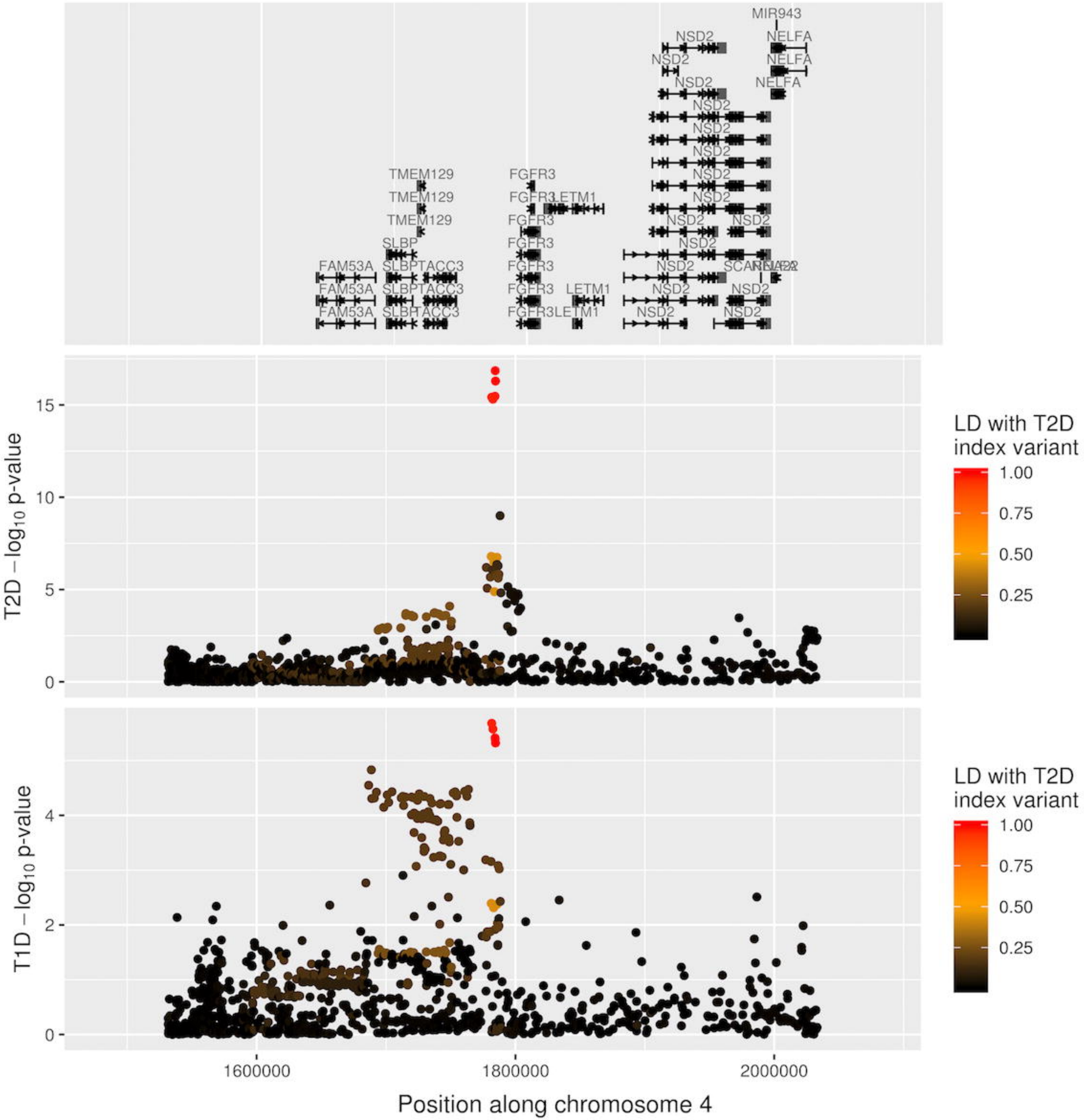
Mark McCarthy has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, NovoNordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, Mark McCarthy is an employee of Genentech, and a holder of Roche stock. Anubha Mahajan is an employee of Genentech since January 2020, and a holder of Roche stock.

Contribution statement

JRJI carried out the type 1 diabetes meta-analysis and the colocalisation analyses. AM carried out the type 2 diabetes meta-analysis and conditional analyses, as well as provided guidance throughout. CS and FC were involved in data collection in the Sardinia collection and carried out the association testing in this collection. DC and QL provided statistical advice and input. MIS provided biological insight. MM, AM and JAT oversaw the research. All authors carried out critical examination of the manuscript.

References

1. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–6 (2015).
2. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
3. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum. Mol. Genet.* (2018). doi:10.1093/hmg/ddy314
4. Consortium, T. H. R. A reference panel of 64 , 976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
5. Guo, H. *et al.* Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.* **24**, 3305–3313 (2015).
6. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *bioRxiv* doi:https://doi.org/10.1101/838946
7. Vujkovic, M. *et al.* Discovery of 318 novel loci for type-2 diabetes and related micro- and macrovascular outcomes among 1.4 million participants in a multi-ethnic meta-analysis. *MedRxiv* (2019). doi:10.1101/19012690
8. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433 , 540 East Asian individuals. *Nature* (2020). doi:10.1038/s41586-020-2263-3
9. Dooley, J. *et al.* Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nat. Genet.* **48**, 519–527 (2016).
10. Elzirik, D. L., Pasquali, L. & Cnop, M. Pancreatic beta-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. *Nat. Rev. Endocrinol.* (2020). doi:10.1038/ s41574-020-0355-7


MIR943
NSD2
NELFA
TMEM129
FGFR3
LETM1
SLBP
FAM53A
SLBPTACC3
SCARNA22
T2D −log10 p-value
T1D −log10 p-value
LD with T2D index variant
1.00
0.75
0.50
0.25
1600000
1800000
2000000
Position along chromosome 4