Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls

Mary D Fortune1, Hui Guo1,2, Oliver Burren1, Ellen Schofield1, Neil M Walker1, Maria Ban3, Stephen J Sawcer3, John Bowes4,5, Jane Worthington4,5, Anne Barton4,5, Steve Eyre4,5, John A Todd1 & Chris Wallace1,6

Determining whether potential causal variants for related diseases are shared can identify overlapping etiologies of multifactorial disorders. Colocalization methods disentangle shared and distinct causal variants. However, existing approaches require independent data sets. Here we extend two colocalization methods to allow for the shared-control design commonly used in comparison of genome-wide association study results across diseases. Our analysis of four autoimmune diseases—type 1 diabetes (T1D), rheumatoid arthritis, celiac disease and multiple sclerosis—identified 90 regions that were associated with at least one disease, 33 (37%) of which were associated with 2 or more disorders. Nevertheless, for 14 of these 33 shared regions, there was evidence that the causal variants differed. We identified new disease associations in 11 regions previously associated with one or more of the other 3 disorders. Four of eight T1D-specific regions contained known type 2 diabetes (T2D) candidate genes (COBL, GLIS3, RNLS and BCART1), suggesting a shared cellular etiology.

Overlaps of genetic association to different diseases have been widely observed and are thought to reflect shared etiology for the diseases1. However, showing that a variant is associated with two traits does not demonstrate that it is causal for both, as this effect may be due to distinct causal variants in linkage disequilibrium (LD)2. Colocalization analyses are used to study whether potential causal variants are shared by combining information across multiple SNPs in a region. The proportional approach3 tests a null hypothesis of proportionality under which, if causal variants are shared, the effects of any set of SNPs on the two diseases are expected to be proportional to each other. A weakness of this approach is its interpretation. Failure to reject the null hypothesis does not only imply colocalization but could also be caused by either disease not being associated or by insufficient power, owing to too few samples analyzed and/or an incomplete genetic map4 (Supplementary Fig. 1). There is no way of measuring how likely colocalization is. A strength of this approach is that no assumptions are made about the number of causal variants: the null hypothesis corresponds to complete sharing across all causal variants. An alternative to the proportional approach is to use a Bayesian framework5 to generate posterior probabilities for the competing hypotheses of colocalization and distinct causal variants. However, a weakness of this approach, as currently developed, is that it assumes only a single causal variant for each trait within any region.

Existing colocalization methods require that genetic association with the two traits of interest be tested in distinct samples. However, this requirement restricts the applicability of these approaches to related diseases, as each set of case samples must have a corresponding, distinct set of control samples, enabling a logistic binomial model to be used independently for each disease. In contrast, many studies use a common set of controls for different diseases to increase efficiency. Here we extend both colocalization methods to allow for the use of multinomial logistic regression, the natural model for shared controls. Previous studies have identified many regions associated with multiple autoimmune or autoinflammatory diseases, including T1D and celiac disease3,6. Such multi-disease association led to the development of the Immunochip7, a custom genotyping chip with 196,000 SNPs designed to densely cover 186 genomic regions known to associate with at least one immune-related disease on the basis of a genome-wide association study (GWAS) P value < 1 × 10−8. The Immunochip Consortium genotyped a common-control set, to which some disease research groups have added their own controls. We applied our extended methods to Immunochip raw genotyping data for a total of 36,030 samples, including one set of controls and 4 disease cohorts, to better understand the extent of shared genetic etiology in these diseases.

RESULTS

Overview of the method

The Bayesian method derives the posterior support for each of five hypotheses describing the possible associations of a given region

1JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. 2Centre for Biostatistics, Institute of Population Health, University of Manchester, Manchester, UK. 3University Neurology Unit, Addenbrooke’s Hospital, Cambridge, UK. 4Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. 5National Institute for Health Research Manchester Musculoskeletal Biomedical Research Unit, Central Manchester Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. 6Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK. Correspondence should be addressed to M.D.F. (mary.fortune@cirmm.cam.ac.uk) or C.W. (chris.wallace@cirmm.cam.ac.uk).

Received 24 October 2014; accepted 8 May 2015; published online 8 June 2015; corrected after print 8 July 2015; doi:10.1038/ng.3330
with two diseases. Of greatest interest are hypothesis H3, where both diseases are associated with the region, with different causal variants, and hypothesis H4, where both diseases are associated with the region and share a single causal variant. Association with both traits corresponds to H3 or H4; colocalization of association to a single variant corresponds to H4. This method requires the specification of prior probabilities for each hypothesis. We calibrated priors to match our expectation that about 50% of the regions associated with two immune-mediated diseases correspond to a shared causal variant (Supplementary Fig. 2), a proportion close to that found in a manually curated summary of association with six immune-mediated diseases (8) (58%). For rheumatoid arthritis and multiple sclerosis, which only the UK subsets of international cohorts were analyzed, we modified the priors for regions with published associations to reflect the additional information from the published papers. Where a region was annotated in ImmunoBase as associated with rheumatoid arthritis or multiple sclerosis, we shrink our priors for hypotheses corresponding to no association with the disease to close to 0 and increased our priors for the remaining hypotheses (Supplementary Note).

We analyzed 126 Immunochip regions assigned to at least one of the diseases (on the basis of knowledge when the chip was designed or association described in subsequent reports and curated in ImmunoBase; accessed 11 December 2013), using both approaches for all 6 pairwise comparisons of the 4 diseases. The sample and SNP quality control procedures are described in the Online Methods; we excluded low-frequency variants (minor allele frequency (MAF) < 1%) to reduce the number of models to be considered and because genotyping errors are more common among this group of SNPs and we did not have cluster plots available for all diseases. Although GWAS analyses typically have sufficient power to detect association only with more common SNPs, some rarer variants (for example, in TYK2) have been reported with these diseases, which will be missed in our analysis.

| Chromosome | Position (bp) | Disease association | Posterior probability of single association | Candidate causal gene(s) (gene(s) in region) |
|------------|---------------|---------------------|--------------------------------------------|--------------------------------------------|
| 1p22.1     | 92,023,171–93,311,800 | Multiple sclerosis | 1.00 | EVI5 |
| 1p21.2     | 100,982,239–101,455,699 | Multiple sclerosis | 0.57 | ETL2, VCAM1, SLC30A7 |
| 1q31.1     | 116,831,830–116,911,865 | Multiple sclerosis | 1.00 | CD58 |
| 3p24.1     | 28,015,774–28,105,476 | Multiple sclerosis | 0.99 | (CMC1) |
| 3q13.3     | 122,818,149–123,329,522 | Multiple sclerosis | 1.00 | ICB1, SLC15A2, CD86 |
| 5q21.1     | 102,062,861–102,777,130 | Rheumatoid arthritis | 0.58 | C5orf30 |
| 6q23.3     | 137,348,296–137,587,799 | Multiple sclerosis | 1.00 | IL2RA2 |
| 7p12.2     | 50,337,180–50,662,811 | T1D | 0.97 | 3' IKZF1 region |
| 7p12.2     | 50,866,661–51,640,000 | T1D | 1.00 | COBL |
| 8q21.12    | 79,575,897–79,914,680 | Multiple sclerosis | 1.00 | ZC2HC1A |
| 8q24.21    | 129,187,117–129,368,419 | Multiple sclerosis | 0.51 | PVT1, MIR1208 |
| 9p24.2     | 4,218,549–4,311,558 | T1D | 1.00 | GLIS3 |
| 10q23.31   | 89,998,026–90,268,360 | T1D | 0.87 | RNL5 |
| 11p15.5    | 2,024,999–2,264,880 | T1D | 1.00 | INS |
| 12q24.31   | 121,926,103–122,574,026 | Multiple sclerosis | 0.59 | PITPNM2 |
| 14q32.2    | 100,357,783–100,398,492 | T1D | 0.98 | DLK1 |
| 16q23.1    | 73,760,230–74,806,012 | T1D | 1.00 | BCR1 |
| 19p13.3    | 65,648,31–66,363,04 | Multiple sclerosis | 1.00 | (EPS15L1, CALR3, MED26, C19orf44, CHEPR, SLC35E1) |
| 19p13.11   | 163,004,97–166,122,40 | Multiple sclerosis | 1.00 | MPV17L2, IFI30 |
| 20p13      | 144,447,2–170,7590 | T1D | 0.99 | (SIRPD, SIRPBP1, SIRPG) |

The table lists the 21 regions that are most likely disease specific under our analysis and for which we know of no other immune-mediated diseases (from the 15 diseases curated in ImmunoBase) that have reported association to these regions (as curated in ImmunoBase, accessed 9 July 2014, and the US National Institutes of Health GWAS catalog, accessed 10 July 2014). To be included, regions had to have a posterior probability of association with a single disease of >0.5 in at least one pairwise analysis (SNP coverage varied between analyses) and a posterior probability of association to any other disorder of <0.2. Candidate causal genes are given. In the case where no candidate causal genes are known, we have given, in parentheses, the genes in and around the region.

*There are two Immunochip regions that overlap IKZF1 and are separated by a recombination hotspot. The region toward the 5' end has colocalizing associations with multiple sclerosis and T1D, whereas the region toward the 3' end appears to be specific to T1D (Supplementary Fig. 7). Note that we provide the coordinates of each region and not an index SNP as is conventional in GWAS because the method synthesizes information across the whole region and does not, in most cases, highlight a single SNP responsible for the association.
Table 2 Fourteen regions showing evidence of separate SNP effects (P(H3) > 0.5)

| Chromosome | Position (bp) | Associations | Evidence | Candidate causal genes(s) |
|------------|--------------|--------------|----------|---------------------------|
| 2p16.1     | 60,722,116–61,952,726 | C-M          | CM; H3 ~0.65 | REL |
| 2q32.2     | 191,412,527–191,739,472 | RC-M         | RM; H3 ~0.51 | STAT1, STAT4 |
| 2q33.1     | 202,920,548–204,528,303 | D-C-R        | DR; H3 ~0.98; RC; H3 ~0.91 | CD28, CTLA4, ICOS |
| 3p21.31    | 45,812,888–46,633,741 | D-C          | DC; H3 ~0.92 | CCR3, CCR1, CCR5 |
| 3q25.33    | 160,950,948–161,389,020 | C-M          | CM; H3 ~0.96 | IL12A |
| 4q27       | 123,121,079–124,497,235 | D-C          | DC; H3 ~1.00 | IL2 IL21 |
| 6q23.3     | 137,914,792–138,345,363 | DRC-M        | RM; H3 ~0.75; CM; H3 ~0.85 | TNFAIP3 |
| 10p15.1    | 6,068,495–6,237,542 | D-M          | DM; H3 ~1.00 | IL2RA |
| 11q23.3    | 117,805,448–118,403,529 | C-M          | CM; H3 ~0.82 | CXCRL5 |
| 13q32.3    | 98,723,782–99,034,738 | D-C          | DC; H3 ~0.67 | GPR183 |
| 16p13.13   | 10,831,557–11,019,034 | D-M          | DM; H3 ~1.00 | IL2RA |
| 18p11.21   | 12,407,903–12,919,721 | D-C          | DC; H3 ~0.58 | PTPN2 |
| 19p13.2    | 10,081,000–11,019,034 | DRM-C        | DC; H3 ~0.53 | ICAM1, ICAM3, TYK2 |
| 21q22.3    | 42,681,877–42,771,181 | D-R-C        | DR; H3 ~0.77; DC; H3 ~0.99; RC; H3 ~0.69 | UBASH3A |

Overview of the results

The Bayesian approach assumes a single causal variant per trait in any region. To allow for multiple causal variants within individual regions, we used a stepwise method. In the overwhelming majority of cases (740 of 756 pairwise comparisons, or 98%), the data were consistent with at least one most causal variant per trait in the 126 regions analyzed. In the remaining 16 pairwise comparisons from 8 regions, we used a stepwise method to allow for multiple causal variants. Ninety of the 126 regions (71%) showed association with at least one disease; in 33 regions, association was shared by at least 2 diseases (Fig. 1). Complete results are given in Supplementary Tables 1–3. For 57 regions, the greatest support was for association with precisely one of the 4 diseases; in 21 regions, we know of no other immune-mediated diseases for which association has been reported in these regions, and we therefore hypothesize that these may be disease-specific regions among autoimmune diseases (Table 1).

In the Bayesian approach, when the posterior probability of a hypothesis is close to 0.5, assignment cannot be made with confidence to any single hypothesis. However, in the 30 instances in which both diseases showed very strong evidence of association (P(H3 or H4) > 0.9), the Bayesian and proportional approaches gave consistent results. For these 30 cases, the proportional null hypothesis was rejected only for regions for which the Bayesian analysis favored H3 and was not rejected for regions where H4 was favored. Focusing on these cases, the data strongly support the idea that the same causal variants underlie all diseases in ten cases, whereas seven regions showed strong evidence for distinct variants, suggesting that just under half (42%) of overlapping association signals reflect distinct causal variants. In total, 14 regions showed evidence of separate SNP effects (P(H3) > 0.5; Table 2).

Figure 2 Distribution of \( \hat{\eta} \), the estimated proportionality coefficient, together with its 95% confidence interval. In the case of colocalization, \( \eta \) is the ratio of the effects the region exerts on the two traits. If \( |\hat{\eta}| > 1 \) corresponds to a stronger effect for trait 2 than for trait 1. We estimate \( \eta \) by \( \hat{\eta} \). Labels on the x axis give the traits and regions analyzed (D, T1D; R, rheumatoid arthritis; C, celiac disease; M, multiple sclerosis). Note that, in some regions, conditional analysis supports the existence of multiple associated variants. If none of these overlapped, then we considered the region to have separate SNP effects. Also note that we provide the coordinates of each region and not an index SNP as is conventional in GWAS because the method synthesizes information across the whole region and does not, in most cases, highlight a single SNP responsible for the association.

| Chromosome | Position (bp) | Associations | Evidence | Candidate causal genes(s) |
|------------|--------------|--------------|----------|---------------------------|
| 2q32.2     | 191,412,527–191,739,472 | RC-M         | RM; H3 ~0.51 | STAT1, STAT4 |
| 2q33.1     | 202,920,548–204,528,303 | D-C-R        | DR; H3 ~0.98; RC; H3 ~0.91 | CD28, CTLA4, ICOS |
| 3p21.31    | 45,812,888–46,633,741 | D-C          | DC; H3 ~0.92 | CCR3, CCR1, CCR5 |
| 3q25.33    | 160,950,948–161,389,020 | C-M          | CM; H3 ~0.96 | IL12A |
| 4q27       | 123,121,079–124,497,235 | D-C          | DC; H3 ~1.00 | IL2 IL21 |
| 6q23.3     | 137,914,792–138,345,363 | DRC-M        | RM; H3 ~0.75; CM; H3 ~0.85 | TNFAIP3 |
| 10p15.1    | 6,068,495–6,237,542 | D-M          | DM; H3 ~1.00 | IL2RA |
| 11q23.3    | 117,805,448–118,403,529 | C-M          | CM; H3 ~0.82 | CXCRL5 |
| 13q32.3    | 98,723,782–99,034,738 | D-C          | DC; H3 ~0.67 | GPR183 |
| 16p13.13   | 10,831,557–11,019,034 | D-M          | DM; H3 ~1.00 | IL2RA |
| 18p11.21   | 12,407,903–12,919,721 | D-C          | DC; H3 ~0.58 | PTPN2 |
| 19p13.2    | 10,081,000–11,019,034 | DRM-C        | DC; H3 ~0.53 | ICAM1, ICAM3, TYK2 |
| 21q22.3    | 42,681,877–42,771,181 | D-R-C        | DR; H3 ~0.77; DC; H3 ~0.99; RC; H3 ~0.69 | UBASH3A |
**Figure 3** The 2q33.1 region containing the candidate gene **CTLA4**. Three potential causal variants are partially shared by T1D, rheumatoid arthritis and celiac disease. (a) Manhattan plots of the region. Blue, index SNP rs231775; green, index SNP rs1980422; magenta, index SNP rs3087243; all other SNPs are colored according to their LD with these three SNPs. SNPs rs231775 and rs3087243 have \( r^2 = 0.50 \); all other pairwise \( r^2 \) values are \(<0.1\). T1D, type 1 diabetes; RA, rheumatoid arthritis; CEL, celiac disease. Gene positions and recombination rates for these regions are given in the bottom plot. (b) Each possible model involving these three SNPs was tested; the four models with the highest posterior probabilities, which together encompass over 90% of the total posterior probability, are shown. (c) Effect size estimates (including 95% confidence intervals) for each SNP on each disease in the most likely model. (d) Effect size estimates (including 95% confidence intervals) for each SNP on each disease in the second most likely model.

**Disentangling patterns of association**

For regions with colocalized associations, the directions of effect for the associations were generally consistent in the two diseases (Fig. 2), with the exception of the 6q25.3 region containing candidate gene **TAGAP**, which was associated in our analysis with celiac disease and multiple sclerosis only: the risk allele for celiac disease was protective for multiple sclerosis and vice versa (Supplementary Fig. 3). This opposing effect of **TAGAP** alleles has been described previously for T1D and celiac disease, although the region did not provide sufficient evidence for association with T1D in the data available to us. A similar effect for the 2q12.1 region containing candidate gene **IL18RAP** has also been reported. However, more recent data have not offered support for T1D association at 2q12.1, and in our analysis the posterior support was concentrated on celiac disease association alone.

Patterns of association with multiple diseases can be complex. In the 2q33 region containing established candidate gene **CTLA4**, as well as the equally strong functional candidate genes **CD28** and **ICOS**, three potential causal variants appear to be partially shared by T1D, rheumatoid arthritis and celiac disease. The strongest association for T1D was at rs3087243 (which has previously been called CT60), whereas the strongest association for celiac disease was with rs231775 (which alters the amino acid at position 17 of **CTLA4**, p.Ala17Thr, and has previously been called CT42). The two SNPs have LD (\( r^2 = 0.5 \), and haplotype analysis has previously suggested that CT60 and not CT42 is causal for Graves’ disease. For rheumatoid arthritis, the strongest single-SNP signal was at rs1980422, which is not in LD with either CT42 or CT60 (\( r^2 < 0.1 \)). We fitted the S12 possible standard multinomial models involving these 3 SNPs for the 3 diseases and computed approximate Bayes factors for each. Assuming each model to be equally likely a priori, the model with the highest posterior probability had rs1980422 and rs3087243 (CT60) signals for celiac disease and rs231775 (CT42) and rs1980422 signals for both T1D and rheumatoid arthritis; however, whereas rs231775 (CT42) was the SNP with the strongest effect for T1D, rs1980422 had a strongest effect on rheumatoid arthritis (Fig. 3). We note that our analysis is based on SNPs selected through a stepwise process and that, without fine-mapping analysis, we cannot claim that any one of these models correctly reflects the causal variants for any disease. These results do however clearly illustrate the different patterns of association for the three disorders and emphasize the potential complexity that can arise in regions with multiple association signals. They thus motivate the future extension of the colocalization approach developed here to allow model search strategies that do not require stepwise assumptions.

**Discovery of new associations**

Two regions were associated with all four diseases (Fig. 1). One was the 6q23.3 region containing candidate gene **TNFAIP3**, known to be associated with rheumatoid arthritis and celiac disease. There has been some published evidence that T1D is associated with this region, although not at genome-wide significant levels. Our results identify a T1D signal, colocalized with that for rheumatoid arthritis.
and celiac disease, suggesting a single shared causal variant affecting the three diseases. There is also evidence of association with multiple sclerosis, driven by a distinct causal variant (in celiac disease–multiple sclerosis analysis, $P(\mathbb{I}_3) = 0.83$; Fig. 4).

The second region associated with all four diseases was 19p13.2, known to be associated with T1D, rheumatoid arthritis and multiple sclerosis and containing the strong functional candidate gene TYK2, although the immune adhesion genes ICAM1 and ICAM3 are also good candidate genes. Our analysis provides support for these associations, with a posterior probability of colocalization approaching 1. We also found evidence for a previously unreported association with celiac disease. In each of the pairwise analyses involving celiac disease, the probability of both diseases being associated was approximately 0.88, although the new signal could be distinct: we found $P(\mathbb{I}_4|\mathbb{I}_3)$ to be about 0.5 (Supplementary Fig. 4). In total, 11 regions showed strong evidence of new association with $P(\mathbb{I}_4|\mathbb{I}_3) > 0.5$ (Table 3).

In regions with colocalization of previously unreported associations with known associations, the effect sizes tended to be smaller for the newly associated diseases (Fig. 2). This could indicate that the stronger effect is conferred by the previously known association or could result from winner’s curse, with the previously known associations displaying inflated effect size estimates. In general, for colocalized signals, the coefficient of proportionality was centered on 1.

One new association found was in the chromosome 1q24.3 region, known to be associated with celiac disease and containing the candidate gene FASLG. Pathway analysis also generated evidence for a T1D-associated variant here, although no SNP reached the genome-wide significance threshold. Our results provide support for a shared causal variant for T1D and celiac disease (posterior probability of 0.71). Our Bayesian approach also enables fine mapping when dense genotyping data are available, as was the case here. We identified a single likely causal variant lying in a region with strong evidence of predicted regulatory activity, rs78037977 (Supplementary Fig. 5), with a posterior probability of $P(\mathbb{I}_4|\mathbb{I}_3)$ of 0.97.

Table 3  Eleven regions showing strong evidence of new association ($P(\mathbb{I}_4|\mathbb{I}_3) > 0.5$) for an analysis involving a previously non-associated trait

| Chromosome Position | Previous associations | Associations found | Posterior probability of both diseases are associated with $\mathbb{I}_3$ or $\mathbb{I}_4$ | Posterior probability of shared causal variant given joint association $P(\mathbb{I}_4|\mathbb{I}_3$ or $\mathbb{I}_4)$ | Candidate causal gene(s) (gene(s) in region) |
|---------------------|----------------------|-------------------|-----------------------------------------------|-----------------------------------------------|---------------------------------------------|
| 1q24.3              | 170,882,016–171,208,336 | C DC              | DC: 0.75                                       | DC: 0.95                                       | FASLG                                       |
| 2p14                | 65,246,601–65,570,598  | R RM              | RM: 0.86                                       | RM: 0.72                                       | SPRED2                                      |
| 2q11.2              | 99,883,120–100,415,547 | DR DRC            | DC: 0.98; RC: 1.00                             | DC: 0.57; RC: 0.90                             | AFF3                                        |
| 2q37.1              | 230,758,228–230,962,304 | M CM              | CM: 0.94                                       | CM: 0.90                                       | SP140                                       |
| 5q11.2              | 55,450,712–55,492,884  | RM DRM            | DR: 0.71; DM: 0.71                             | DR: 1.00; DM: 1.00                             | ANKRD55                                     |
| 6q23.3              | 137,914,792–138,345,363 | RCM DRC-M         | DR: 0.80; DC: 0.77                             | DR: 0.94; DC: 0.93                             | TNAIP3                                      |
| 7p14.2              | 37,323,488–37,406,978  | CM RCM            | RC: 0.80; RM: 0.77                             | RC: 0.84; RM: 0.83                             | ELMO1                                       |
| 7p12.2              | 50,222,360–50,335,957  | M DM              | DM: 0.73                                       | DM: 0.70                                       | 5′ IKZF1 region\(^a\)                       |
| 13q32.3             | 98,723,872–99,034,738  | M D-C             | DC: 0.67                                       | DC: 0.00                                       | GPR183                                      |
| 15q25.1             | 76,773,859–77,050,416  | M DC              | DC: 0.82                                       | DC: 0.99                                       | CTSH                                        |
| 19p13.2             | 10,081,000–11,019,034  | DRM DRM-C          | DC: 0.87; CM: 0.88                             | DC: 0.40; RC: 0.46; CM: 0.57                  | ICAM1, ICAM3, TYK2                          |

| \(D\), T1D; R, rheumatoid arthritis; C, celiac disease; M, multiple sclerosis. New associations are denoted by bold font. Candidate causal genes are as associated across all curated diseases by ImmunoBase. Note that, in the case of TNAIP3, there is strong evidence that multiple sclerosis is caused by a distinct causal variant in comparison to the other traits. Distinct signals are separated by a hyphen. Because we only had a subset of the genotype data, not all of the pairs (previously published) associations are detected. Note that we provide the coordinates of each region and not an index SNP as is conventional in GWAS because the method synthesizes information across the whole region and does not, in most cases, highlight a single SNP responsible for the association. \(^a\)An association of T1D in a region 5′ of IKZF1, for which it is hypothesized that IKZF1 is the candidate causal gene, is already known. The new association we report here is in a region 5′ of IKZF1 and is independent of the established association.
probability of being causal among all the genotyped variants, given the colocalization hypothesis, of 0.99. Note that rs78037977 was removed from the celiac disease data in the original analysis owing to its failing a missingness check (the call rate of 99.942% was just below the 99.95% cutoff applied in quality control). Plots of the signal clouds for our samples at this SNP are given in Supplementary Figure 6. The clustering shown here was of good quality, implying that the rs78037977 genotype can be considered reliable.

Prior sensitivity

We tested prior sensitivity by varying $p_{12}$ (the probability that an arbitrary SNP is associated with both diseases) from $1 \times 10^{-5}$ to $1 \times 10^{-7}$, while keeping $p_1$ and $p_2$ (the probabilities that this SNP is associated with only trait 1 or only trait 2) constant at $1 \times 10^{-4}$ (Supplementary Table 4). Whether a region is disease specific is largely unaffected by choice of $p_{12}$, and, for the five regions discussed in detail in this paper (1q24.3 (FASLG), 2q33.1 (CTLA4), 6q23.3 (TNFAIP3), 6q25.3 (TAGAP) and 19p13.2 (TYK2)), the prior did not change which diseases were associated. However, the ratio of the posterior odds for $\mathbb{E}1$ and $\mathbb{E}3$ did vary with $p_{12}$. Under $p_{12} = 1 \times 10^{-4}$, neither 1q24.3 (FASLG) nor 6q23.3 (TNFAIP3) had strong posterior support as a new T1D-associated region, as the evidence for association in each region came about as a result of colocalization with the stronger previously known association. This dependence on prior belief is a strength of Bayesian methods, but these methods require that priors be carefully calibrated. Although our prior belief is that about 50% of the regions associated with two immune-mediated diseases are likely to correspond to a shared causal variant, others may disagree. The results given in Supplementary Table 2 can be used to calculate the posterior probability under any alternative $p_{12}$ using the formula given in the Supplementary Note.

DISCUSSION

Colocalization methods thus far have allowed for the simultaneous analysis of only two traits, which is a potential weakness when considering more than two diseases, as investigated here. The Bayesian approach could be extended to an arbitrarily large set of traits, at the cost of increased computational complexity and spreading the posterior probability over an exponentially increasing hypothesis space, potentially making it difficult to draw firm conclusions. Wen et al., in their description of an alternative method for partitioning the association of a single SNP among multiple related quantitative traits, suggest dealing with this complexity by considering only the extremes—where a SNP is associated with all traits, exactly one trait or no traits. Such reduction is impractical when analyzing regions, as it does not allow for overplotting but distinct signals. Although we have extended our software to consider 3 diseases simultaneously, we have chosen for practical reasons to focus on pairwise analyses, with manual curation of the 11 regions (9%) for which more than 2 diseases showed association.

Giambartolomei et al. showed that inference is consistent when the causal variant is directly genotyped or well imputed. The decision was taken when the Immunochip was designed not to thin by LD but to instead target all SNPs and small indels known at that time in 1000 Genomes Project European-ancestry samples, and it has since been shown that common variants can be very accurately imputed using the Immunochip. Therefore, we are likely very close to a scenario where causal variants are directly genotyped. The application of our method to the less complete coverage provided by genome-wide SNP arrays would require an imputation step to allow consistent inference to be made. The Bayesian colocalization analysis assumes a single causal variant per region, which could be restrictive, and we addressed this using a stepwise approach, attempting to colocalize the individual signals for each disease where there was evidence for more than one. The agreement between our results with this approach and those obtained using the proportional colocalization approach, which does not make this assumption, confirms the Appropriateness of the stepwise approach in the cases we consider.

We identified 21 regions that appeared to be associated with only one autoimmune disease. One challenge in interpretation when defining signals unique to a particular disease is exemplified by a region on chromosome 7p12.2, which contains the candidate causal gene IKZF1. This gene overlaps two Immunochip regions separated by a recombination hotspot, one 5′ of IKZF1 and one 3′ of IKZF1. The 5′ region contains colocalized signals for multiple sclerosis and T1D, whereas the 3′ region contains only a T1D signal (Supplementary Fig. 7). Our analysis has been based on regions, as defined in the design of the Immunochip, and on recombination hotspots. However, although the T1D signals in these regions are independent and the 3′ region of IKZF1 appears to be uniquely associated with T1D, it is plausible that the causal variants in both regions act through the same gene, IKZF1. Another challenge is to deal with the effects of power, given the established influence of sample size on power to detect associations.

Many of the regions listed in Table 1 contain genes linked to immune function, and we expect a number of apparently disease-specific regions to be associated with other diseases as sample sizes for each disease continue to increase. Indeed, the chromosome 19p13.11 region, associated only with multiple sclerosis in our analysis, has previously been associated with lymphocyte count, and there is high LD between the peak SNP for multiple sclerosis (rs1870071) and the one for lymphocyte count (rs11878602; $r^2 = 0.99$), suggesting that an immune mechanism underlies the associations.

However, in the case of T1D, three regions specific for this disease overlap known T2D-associated regions. Chromosome 9p24.2, containing the candidate gene GLIS3, has been associated with T2D and fasting glucose levels, and there is high LD between the peak SNP for T1D (rs10814914) and the SNPs associated with these other traits (rs7041847, $r^2 > 0.9$). GLIS3 and its causal allele affect disease risk by altering pancreatic $\beta$ cell function, probably by increasing $\beta$ cell apoptosis. Chromosome 16q23.1, containing the candidate gene BACR1, is associated with T1D in our analysis and with T2D, and the T2D alleles in this region have been associated with reduced $\beta$ cell function; again, there is high LD between the peak SNPs for T1D (rs8056814) and T2D (rs7202877; $r^2 = 0.81$). Inspecting the distribution of T2D GWAS $P$ values at the peak SNPs in our T1D-associated regions (Supplementary Fig. 8), we note that the peak SNP in the T1D region at 6q22.32, rs17754780, also shows association with T2D ($P = 7.9 \times 10^{-5}$) and is in tight LD with the peak T2D SNP in the region (rs9385400; $r^2 = 0.97$). This region has been reported to be associated with T2D at genome-wide significance in a larger study. Chromosome 6q22.3 is not uniquely associated with T1D in our analysis because it overlaps an established region for Crohn’s disease, but the lead SNP for Crohn’s disease (rs9491697) is not in LD with the T1D SNP ($r^2 = 0.03$). This region is thus likely to contain a third signal shared by T1D and T2D. The nearest genes are MIR588, about which little seems to be known, and CENPW (encoding centromere protein W), which has no obvious functional candidacy. This genetic overlap between T1D and T2D (Supplementary Table 5) emphasizes that T1D results from an interaction between the immune system and $\beta$ cells, and it is probable that some of our other apparently disease-specific regions will also prove to be specific to the targeting of autoimmune destruction in multiple sclerosis and rheumatoid arthritis.
By analyzing regions known to associate with one disease, we were able to link 11 to additional disorders: in most cases (8/11), the newly discovered disease association was clearly colocalized with a previously known signal, whereas in one case, GPR183, the evidence supported a distinct causal variant for the new association. In other cases (3/11), the evidence for colocalization was more equivocal, even with evidence for pairwise association.

In a standard GWAS analysis, a P-value significance threshold of $5 \times 10^{-8}$ is used in the absence of replication data owing to a desire to minimize the reporting of false positive results, although a relaxation of this threshold has been suggested\(^ {29} \). However, because autoimmune diseases are known to share etiology, when conditioning on an association for one autoimmune disease, a less stringent threshold should be required to believe that another association is significant. Indeed, although the question of whether the Immunochip significance threshold should be somewhat relaxed remains\(^ {8} \), examination of P values in the regions in which we observed new associations (Supplementary Fig. 9) suggests that a threshold between $1 \times 10^{-5}$ and $1 \times 10^{-6}$ for SNPs that are confirmed index SNPs for another disease might be more appropriate. We estimated that 42% of overlapping and genome-wide significant immune-mediated disease signals correspond to distinct causal variants. In these regions, therefore, there appear to be distinct causal variants for two or more autoimmune diseases that are physically proximal but in low LD. We suggest that physical proximity to a known associated variant for a related disease and not only LD with it may prove to be an appropriate criterion with which to alter the interpretation of a small but not genome-wide significance threshold. Variants meeting such thresholds might be prioritized for genotyping in replication samples. We also note that the four diseases we studied are all characterized by the presence of autoantibodies. Had we included autoantibody-negative diseases, we might have found a higher proportion of discordant associations, as reported in a previous manual curation of Immunochip studies\(^ {8} \), given that there remains considerable overlap in the locations of association signals. Although a careful and detailed manual curation of several studies has been conducted\(^ {8} \), the ability of colocalization methods to distinguish shared from distinct causal variants allows clearer interpretation of genetic results.

In summary, we have developed a methodology for examining shared genetic etiology for diseases in the context of common-control data sets, extending previous work\(^ {2-3} \). This approach enables the discovery of new disease associations and the exploration of complex association patterns. Although this method has been presented here in the analysis of autoimmune diseases, the prior is user defined and the method could be used to analyze any pair of related diseases.

**ACKNOWLEDGMENTS**

M.D.F. is funded by the Wellcome Trust (099772). C.W. and H.G. are funded by the Wellcome Trust (089989).

This work was funded by the JDRF (9-2011-253), the Wellcome Trust (091157) and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre. The Cambridge Institute for Medical Research (CIMR) is in receipt of a Wellcome Trust Strategic Award (100140). ImmunoBase is supported by Eli Lilly and Company.

We thank the UK Medical Research Council (MRC) and Wellcome Trust for funding the collection of DNA for the British 1958 Birth Cohort (MRC, grant G0000934 and Wellcome Trust grant 068545/Z/02). Control DNA samples were prepared and provided by S. Ring, R. Jones, M. Pembrey, W. McArdf, D. Strahan and P. Burton.

This research uses 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), the National Institute of Allergy and Infectious Diseases (NIAID), the National Human Genome Research Institute (NHGRI), the National Institute of Child Health and Human Development (NICHD) and the JDRF and supported by grant U01DK062418 from the NIDDK.

Collection of the rheumatoid arthritis data was funded by the Arthritis Foundation and the US National Institutes of Health.

We are grateful to G. Trynka and D. van Heel for use of the celiac disease data. Funding was provided by the Wellcome Trust, by grants from the Celiac Disease Consortium and an Innovative Cluster approved by the Netherlands Genomics Initiative, by the Dutch government (BSIK03009 to C.W.) and the Netherlands Organisation for Scientific Research (NWO; grant 918.66.620), and by US National Institutes of Health grant 1R01CA141743 and Fondo de Investigación Sanitaria grants FIS08/1676 and FIS07/0533.

Multiple sclerosis data were provided by S. Sawcer and the International Multiple Sclerosis Genetics Consortium. Funding was provided by the US National Institutes of Health, the Wellcome Trust, the UK Multiple Sclerosis Society, the UK MRC, the US National Multiple Sclerosis Society, the Cambridge NIHR Biomedical Research Centre, DeNDRon, the Bibbi and Niels Eens Foundation, the Swedish Brain Foundation, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Foundation for Strategic Research, the Stockholm County Council, Karolinska Institutet, INSERM, Fondation d’Aide à la Recherche sur la Sclérose en Plaques, Association Française contre les Myopathies, Infrastructures en Biologie Santé et Agronomie (GIS-IBISA), the German Ministry for Education and Research, the German Competence Network Multiple Sclerosis, Deutsche Forschungsgemeinschaft, Munich Bioteq Cluster M4, the Fidelity Biosciences Research Initiative, Research Foundation Flanders, Research Fund KU Leuven, the Belgian Charcot Foundation, Gemeinnützige Hertie Stiftung, University Zurich, the Danish Multiple Sclerosis Society, the Danish Council for Strategic Research, the Academy of Finland, the Sigrid Juselius Foundation, Helsinki University, the Italian Multiple Sclerosis Foundation, Fondazione Cariplo, the Italian Ministry of University and Research, the Torino Savings Bank Foundation, the Italian Ministry of Health, the Italian Institute of Experimental Neurology, the Multiple Sclerosis Association of Oslo, the Norwegian Research Council, the South-Eastern Norwegian Health Authorities, the Australian National Health and Medical Research Council, the Dutch Multiple Sclerosis Foundation and Kaiser Permanente.

M. Evangelou is thanked for motivating the investigation of the FASLG association.

**AUTHOR CONTRIBUTIONS**

M.D.F. conceived and designed experiments, performed statistical analyses, analyzed data and wrote the manuscript. H.G. conceived and designed experiments. O.B. analyzed data, prepared data and maintained ImmunoBase. E.S. prepared data and maintained ImmunoBase. N.M.W. prepared data.

M.B. and S.J.S. contributed multiple sclerosis data and interpreted results. J.B., J.W., A.B. and S.E. contributed rheumatoid arthritis data and interpreted results. J.A.T. analyzed data, contributed T1D data and wrote the manuscript.

C.W. conceived and designed experiments, analyzed the data and wrote the manuscript. All authors reviewed and contributed to the final manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**URLs.** ImmunoBase, http://www.immunobase.org/; mlogitBMA R package, http://cran.r-project.org/web/packages/mlogitBMA/index.html; National Institute for Health Research, http://www.nihr.ac.uk/; GWAS catalog, http://www.genome.gov/admin/gwascatalog.txt; DIAGRAM, http://diagram-consortium.org/; code available in the colocCommonControl R package, https://github.com/mdfortune colocCommonControl.
1. Cotopasas, C. et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 7, e1002254 (2011).
2. Plagnol, V., Smyth, D.J., Todd, J.A. & Clayton, D.G. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics 10, 327–334 (2009).
3. Wallace, C. et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. Hum. Mol. Genet. 21, 2815–2824 (2012).
4. Wallace, C. Statistical testing of shared genetic control for potentially related traits. Genet. Epidemiol. 37, 802–813 (2013).
5. Giambartolomei, C. et al. Bayesian test for colocalization between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).
6. Smyth, D.J. et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N. Engl. J. Med. 359, 2767–2777 (2008).
7. Cortés, A. & Brown, M.A. Promise and pitfalls of the Immunochip. Arthritis Res. Ther. 13, 101 (2011).
8. Parkes, M., Cortés, A., van Heel, D.A. & Brown, M.A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat. Rev. Genet. 14, 661–673 (2013).
9. Eyre, S. et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. 44, 1336–1340 (2012).
10. Beecham, A.H. et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat. Genet. 45, 1353–1360 (2013).
11. Mero, J.J. et al. A rare variant of the TYR2 gene is confirmed to be associated with multiple sclerosis. Eur. J. Hum. Genet. 18, 502–504 (2010).
12. Barrett, J.C. et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. 41, 703–707 (2009).
13. Ueda, H. et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 423, 506–511 (2003).
14. Fung, E.Y.M.G. et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAP3 as a susceptibility locus. Genes Immun. 10, 18–21 (2009).
15. Ioannidis, J.P.A. Why most discovered true associations are inflated. Epidemiology 19, 640-648 (2008).
16. Evangelou, E. et al. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. Genet. Epidemiol. 38, 661–670 (2014).
17. Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 43, 1193–1201 (2011).
18. Flute, T., Wen, X., Pritchard, J. & Stephens, M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 9, e1003486 (2013).
19. Deelen, P. et al. Improved imputation quality of low-frequency and rare variants in European samples using the Genome of The Netherlands. Eur. J. Hum. Genet. 22, 1321–1326 (2014).
20. Swafford, A.D.E. et al. An allele of IKZF1 (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes. Diabetes 60, 1041–1044 (2011).
21. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24 (2012).
22. Nalls, M.A. et al. Multiple loci are associated with white blood cell phenotypes. PLoS Genet. 7, e1002113 (2011).
23. Morris, A.P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981–990 (2012).
24. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. 42, 105–116 (2010).
25. Nogueira, T.C. et al. GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic β cell apoptosis via regulation of a splice variant of the BHS-only protein Bim. PLoS Genet. 9, e1003532 (2013).
26. Harder, M.N. et al. Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased β-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. J. Clin. Endocrinol. Metab. 98, E801–E806 (2013).
27. Scott, R.A. et al. Genome-wide association study imputed to 1000 Genomes reveals 18 novel associations with type 2 diabetes. (American Society of Human Genetics, 2014).
28. Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. 46, 234–244 (2014).
29. Panagiotou, O.A. & Ioannidis, J.P.A. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int. J. Epidemiol. 41, 273–286 (2012).
Online Methods

Samples. All samples included in this analysis were gathered in the UK and have reported or self-declared European ancestry. Informed consent was obtained from all subjects after approval from the ethics committee or institutional review board of all participating institutions. Detailed summaries of the sample cohorts are given in the Immunochip reports for celiac disease\(^6\), rheumatoid arthritis\(^9\), multiple sclerosis\(^10\) and T1D\(^30\). For rheumatoid arthritis and multiple sclerosis, we used the subset of cases from the UK. Sample exclusions were applied as described in each paper; in total, 6,691 T1D, 3,870 rheumatoid arthritis, 7,987 celiac disease, 5,112 multiple sclerosis and 12,370 control samples were analyzed. SNPs were filtered to meet the following criteria: call rate > 0.99; MAF > 0.01; and Hardy-Weinberg |\(|\) < 5. SNPs that passed these thresholds in controls and any specific pair of cases were used for that pairwise analysis.

Using only UK cases and controls means that we expect any effects of population stratification to be very limited, as evidenced by the low genomic inflation factors in published UK Immunochip analyses\(^31\), and we did not take any further specific actions to limit effects from population stratification.

Selection of regions for analysis. We considered all regions annotated in Immunochip\(^1\) (accessed 11 December 2013) to be associated with at least one of the four diseases we studied. Where regions overlapped, we formed the union. Regions containing fewer than 10 SNPs or with a SNP density of <1 SNP/kb were excluded. The major histocompatibility complex (MHC) region (chr. 6: 29,797,978–33,606,563, hg18) was removed from the analysis because this region is known to have complex multi-SNP effects. A full list of the 126 regions analyzed, together with our resulting associations, can be found in Supplementary Table 1.

Colocalization analysis. Two colocalization methods were applied to each of the 126 regions (Supplementary Fig. 1).

Bayesian approach. The first approach is based on a Bayesian method proposed by Giambartolomei et al.\(^5\). All models in which each trait is caused by at least one variant are considered, and approximate Bayes factors are computed for each. Our extension follows the same framework, but a multinomial model most one variant are considered, and approximate Bayes factors are computed for each. Phenotypes are modeled using multinomial logistic regression, producing maximum-likelihood estimates \(\hat{\beta}_1\) and \(\hat{\beta}_2\) of the regression coefficients \(\beta_1\) and \(\beta_2\). Because sample sizes can be large, the asymptotic normality of maximum-likelihood estimators is used to approximate

\[
\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}\right) \sim N\left(\frac{\beta_1}{\beta_2}, \frac{1}{V_{11}V_{22} / (V_{12}V_{21})}\right)
\]

for some variance-covariance matrix \(V\).

The proportional approach\(^3\,\,4\) assumes that \(\hat{\beta}_1\) and \(\hat{\beta}_2\) are independent (that is, \(V_{12} = V_{21} = 0\)). However, in the extension for a common-control data set, we cannot assume this and proceed with a fully unknown \(V\).

The null hypothesis corresponds to the existence of a constant \(\eta\) such that:

\[
\beta_1 = \frac{1}{\eta_1} \beta_2
\]

Under this hypothesis and given \(\eta\),

\[
\chi^2 = \left(\frac{1}{\eta} - 1\right) \sum_{i=1}^{2} \hat{\beta}_{2i} - \frac{1}{\eta} \sum_{i=1}^{2} \hat{\beta}_{1i} + \frac{1}{\eta^2} \left(\hat{\beta}_{11} + \hat{\beta}_{22}\right) - \chi^2_p
\]

\(\chi^2\) is used as our test statistic. However, because the value of \(\eta\) was unknown, a posterior predictive \(P\) value is generated instead by integrating the \(P\) values associated with the test statistic over the posterior distribution of \(\eta\). To avoid bias in regression coefficients due to the selection of SNPs on the basis of their strength of association, Bayesian model averaging was used to average inference over all plausible two-SNP models.

Further details of the colocalization methods can be found in the Supplementary Note.

Identification of disease-specific regions. To examine evidence for GWAS association with other traits, we took the index SNP with the smallest \(P\) values in a region and then identified proxy SNPs on the basis of LD (\(r^2 > 0.9\)) using 1000 Genomes Project EUR (European) data. We used these SNPs as a query set to examine associations annotated in the US National Institutes of Health GWAS catalog (accessed 10 July 2014).

We defined disease-specific regions as those for which (i) the posterior probability of single-SNP association was >0.5; (ii) the posterior probability of association with any other disease was <0.2; (iii) the region was not annotated as associated with any other autoimmune disease in Immunochip; and (iv) no proxies for the index SNP were associated with any other autoimmune disease in the US National Institutes of Health GWAS catalog.

Type 2 diabetes data. Summary data from a T2D GWAS meta-analysis\(^2\) were downloaded from the DIAGRAM website (accessed 20 October 2014).

Code availability. The code is used in the coloCommonControl R package, which can be found at https://github.com/mdfortune/coloCommonControl.

30. Ongenut-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, 381–386 (2015).
31. Cooper, J.D. et al. Seven newly identified loci for autoimmune thyroid disease. Hum. Mol. Genet. 21, 5202–5208 (2012).
32. Raftery, A.E. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. Biometrika 83, 291–266 (1996).
Corrigendum: Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls

Mary D Fortune, Hui Guo, Oliver Burren, Ellen Schofield, Neil M Walker, Maria Ban, Stephen J Sawcer, John Bowes, Jane Worthington, Anne Barton, Steve Eyre, John A Todd & Chris Wallace

Nat. Genet. 47, 839–846 (2015); published online 8 June 2015; corrected after print 8 July 2015

In the version of this article initially published, the two panels in Figure 2 were presented in the incorrect order. The error has been corrected in the HTML and PDF versions of the article.