A catalog of GWAS fine-mapping efforts in autoimmune disease

Minal Caliskan,[1,*] Christopher D. Brown,[2] and Joseph C. Maranville[1]

Summary

Genome-wide association studies (GWASs) have enabled unbiased identification of genetic loci contributing to common complex diseases. Because GWAS loci often harbor many variants and genes, it remains a major challenge to move from GWASs' statistical associations to the identification of causal variants and genes that underlie these association signals. Researchers have applied many statistical and functional fine-mapping strategies to prioritize genetic variants and genes as potential candidates. There is no gold standard in fine-mapping approaches, but consistent results across different approaches can improve confidence in the fine-mapping findings. Here, we combined text mining with a systematic review and formed a catalog of 85 studies with evidence of fine mapping for at least one autoimmune GWAS locus. Across all fine-mapping studies, we compiled 230 GWAS loci with allelic heterogeneity estimates and predictions of causal variants and trait-relevant genes. These 230 loci included 455 combinations of locus-by-disease association signals with 15 autoimmune diseases. Using these estimates, we assessed the probability of mediating disease risk associations across genes in GWAS loci and identified robust signals of causal disease biology. We predict that this comprehensive catalog of GWAS fine-mapping efforts in autoimmune disease will greatly help distill the plethora of information in the field and inform therapeutic strategies.

Introduction

Over the past 15 years, genome-wide association studies (GWASs) have identified thousands of loci that are associated with complex traits and common diseases.[1] Because genotypes at nearby variants are usually correlated, a phenomenon referred to as linkage disequilibrium (LD), significant GWAS findings pinpoint genomic regions rather than individual variants (Figure 1A). The resolution at individual loci is influenced by many factors, including the effect size and minor-allele frequency (MAF) of causal variants and local recombination rates, as well as the sample size and ethnic composition of study cohorts (Figure 1B). Additionally, nearly 90% of significant GWAS variants reside in non-protein-coding regions of the genome,[1,2] making their functional interpretation challenging.

While GWASs continue to identify genomic loci associated with complex traits, tremendous efforts go into dissecting such loci to identify trait-relevant genes, regulatory regions, and variants.[6–9] This process of identifying trait-relevant genetic elements in an already constrained genetic locus is referred to as "fine mapping" and is crucial to understanding the underlying disease mechanisms and translating GWAS findings into potential therapeutics. Drug targets that are genetically supported have a significantly higher probability of being approved as successful treatments.[10,11] Many established autoimmune-disease drug targets, such as TNF blockers, IL23/IL12 antagonists, and CTLA4 agonists, have either prospectively or retrospectively been shown to reside in GWAS loci of autoimmune diseases.[1] With hundreds of loci having been associated with autoimmune diseases, comparative analyses of GWAS follow-up studies have great potential to reveal the most consistently dissected loci where a single gene was prioritized on the basis of multiple sources of functional and statistical evidence. Such approaches could then be integrated into the drug-discovery pipelines to allocate time and resources into the most probable targets.

Fine mapping is also critical to understanding the pleiotropic effects of genes across a range of phenotypes. For example, there is notable overlap in GWAS loci across diseases that involve autoimmunity.[12,13] However, the presence of a shared GWAS locus does not imply the same underlying causal variant or the same gene. There are also several reports of shared GWAS loci that harbor alleles with opposite effects on different autoimmune diseases.[14,15] Comparative analyses of GWAS fine-mapping studies could highlight both shared and disease-specific causal variants, genes, and pathways that are involved in autoimmunity. Insight gained from these analyses could aid development of therapeutic hypotheses and experimental directions that could be pursued within or across autoimmune diseases.

Characterized by an inappropriate and excessive inflammatory response against self-antigens, common autoimmune diseases include Crohn disease (CD; MIM: 266600), ulcerative colitis (UC; MIM: 191390), rheumatoid arthritis (RA; MIM: 180300), juvenile idiopathic arthritis (JIA; MIM: 618795), psoriasis (PSO; MIM: 177900), type 1 diabetes (T1D; MIM: 222100), multiple sclerosis (MS; MIM: 126200), vitiligo (VIT; MIM: 193200), systemic lupus erythematosus (SLE; MIM: 152700), and systemic sclerosis

[1]Department of Informatics and Predictive Sciences, Bristol Myers Squibb, Princeton, NJ 08540, USA; [2]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
*Correspondence: minal.caliskan@bms.com

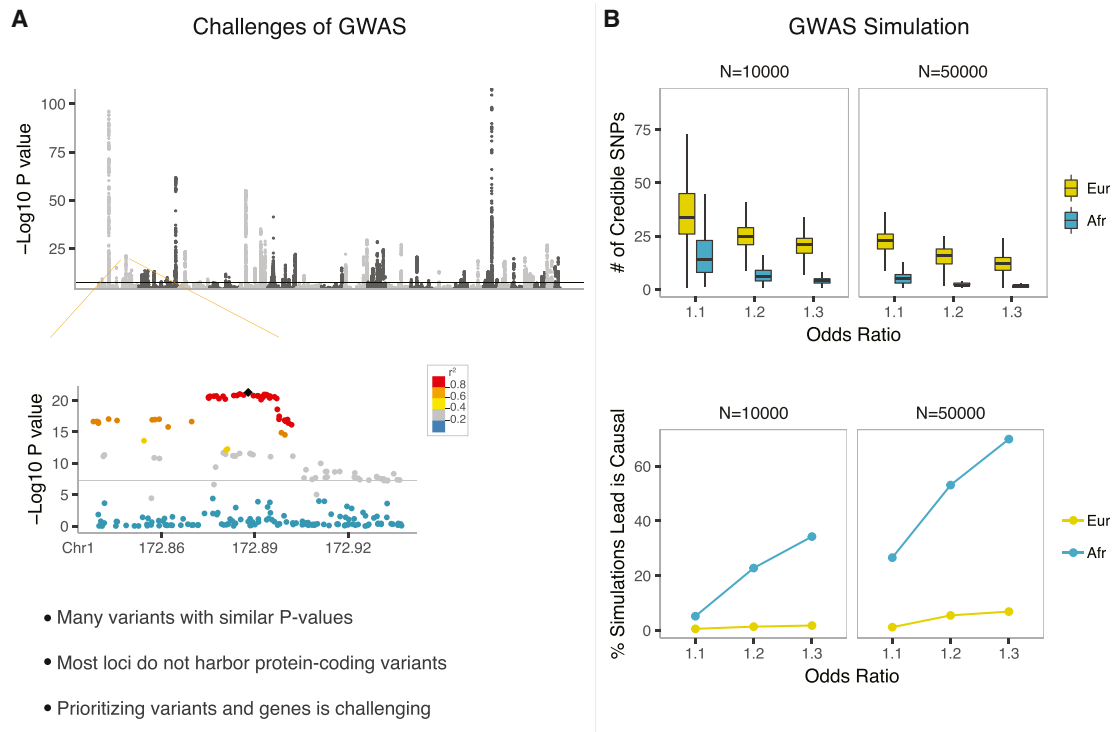**Figure 1. Challenges of GWASs**

(A) An example Manhattan plot displaying Crohn disease GWAS p values.[3] The zoomed-in plot of the chr 1q24.3 locus[3] illustrates the correlation between variants (r^2) and association p values.

(B) The upper figure shows the number of statistically prioritized (95% credible SNP set) variants based on 1,000 simulations under different odds-ratio, sample-size, and ethnicity scenarios. The lower figure shows the percentage of simulations that identified the assigned causal variant as the lead variant. Haplotypes were constructed from 1000 Genomes European (CEU, TSI, FIN, GBR, and IBS) or African (YRI, LWK, GWD, MSL, ESN, ASW, and ACB) subjects. Common genetic variants (MAF > 0.01) across 100 kb region in the chr 1q24.3 locus shown in Figure 1A were included in haplotype construction. Among observed credible SNP set variants, a randomly chosen SNP was assigned as the causal variant for simulation purposes. GWAS p values were simulated with the simGWAS package[4] under different odds-ratio, sample-size, and ethnicity scenarios. 95% credible SNP sets were calculated as implemented in the coloc package.[5] Numbers of credible SNP set variants in each simulation and the percentage of simulations that identified the assigned causal SNP as the lead SNP were plotted.

(SS; MIM: 181750). Dozens to hundreds of GWAS loci are associated with each of these autoimmune diseases, yet there has not been a comprehensive study summarizing results from GWAS fine-mapping efforts in autoimmunity. Here, we used a systematic approach and formed a catalog of published studies with evidence of fine mapping for at least one autoimmune GWAS locus. We summarized the lessons learned about the genetic architecture of complex autoimmune diseases, highlighted some of the inconsistencies that are yet to be resolved in the literature, and used a novel approach to prioritize candidate trait-relevant genes in GWAS loci on the basis of both genome-wide and locus-specific fine-mapping evidence.

A catalog of autoimmune GWAS fine-mapping efforts

We put together a list of 22 manuscripts that are commonly cited as examples of "GWAS fine-mapping studies" (Table S1). Using these 22 manuscripts, we defined

a set of informative terms that often co-occur in the abstracts of the manuscripts on GWAS fine-mapping efforts. We found that in 18 of these 22 abstracts either the "GWAS" or "genome-wide association" term co-occurs with one of the following six terms: "functional," "causal," "mechanisms," "causative," "mechanism," or "mechanistic." Accordingly, we downloaded PubMed abstracts from the National Library of Medicine and searched all abstracts published between January 2005 and January 2020 for the co-occurrence of "GWAS"/"genome-wide association" with any of the aforementioned six terms (Figure 2A). We further filtered the abstracts to retrieve those that contained one or more autoimmune disease ID (diseases under "experimental factor ontology" (EFO) autoimmune disease category: 0005140) shown in Table S2. After performing these two searches sequentially, we identified a total of 837 manuscripts that most likely describe autoimmune GWAS fine-mapping studies. We manually reviewed these 837 manuscripts and assessed whether they include results based on at least one of the variant- or gene-prioritization approaches shown in
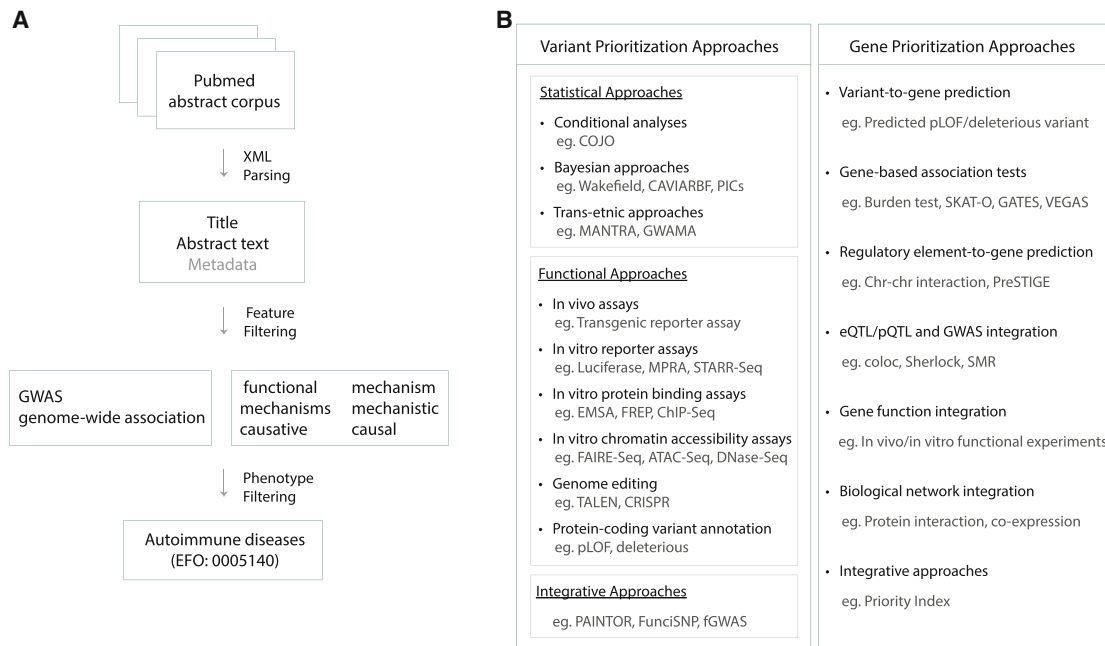
Figure 2. A catalog of autoimmune GWAS fine-mapping efforts
(A) Approach used to identify manuscripts that are most likely on the autoimmune GWAS fine-mapping topic. PubMed abstracts published between January 2005 and January 2020 were retrieved from the National Library of Medicine. There were 9,744 manuscripts following feature (GWAS and fine-mapping terms) filtering and 837 manuscripts following autoimmune disease filtering.
(B) The list of causal-variant and gene-prioritization approaches that were assessed during review and identification of autoimmune GWAS fine-mapping studies.

Figure 2B. For statistical variant-prioritization approaches, we additionally required that the manuscripts fulfill at least one of the following two criteria: (1) deep resequencing followed by genotyping or imputation of discovered variants in the same ancestry population and (2) dense genotyping in a population that matches the ancestry of the population for which the genotyping array was optimized. The list of fine-mapping approaches we used (Figure 2B) have been previously described and discussed in detail by a number of review papers on GWAS fine-mapping approaches.[6–9] In total, we cataloged 85 manuscripts with evidence of fine mapping for at least one autoimmune GWAS locus (Table S3). There were seven (CD, UC, IBD, RA, PSO, T1D, and MS), 11 (CD, UC, IBD, RA, JIA, PSO, T1D, VIT, MS, SLE, and SS), and 15 (CD, UC, IBD, RA, JIA, PSO, T1D, MS, SLE, SS, VIT, Behcet disease [BD; MIM: 109650], psoriatic arthritis [PSA; MIM: 607507], Graves disease [GD; MIM: 275000], and ankylosing spondylitis [AS; MIM: 106300]) autoimmune diseases with allelic heterogeneity, variant prioritization, and gene-prioritization estimates, respectively (Tables S5, S6, and S8).

22% of complex autoimmune-disease-associated loci display allelic heterogeneity

A critical step in fine mapping is to address allelic heterogeneity, i.e., the presence of multiple independent variants that affect the function of the same gene and lead to the same phenotype. Allelic heterogeneity is common for Mendelian diseases, which are caused by rare (risk-allele frequency < 1% in the general population), highly penetrant variants. For example, the ClinVar database[16] includes 34 pathogenic *AIRE* (MIM: 607358) variants that cause polyglandular autoimmune syndrome and 16 pathogenic *FAS* (MIM: 134637) variants that cause autoimmune lymphoproliferative syndrome (Table S4). In fact, allelic heterogeneity was reported for 74% of the Mendelian-disease-causing genes in the ClinVar database (Table S4).

Linkage disequilibrium complicates efforts to assess allelic heterogeneity for common variants associated with complex diseases. Researchers have developed three categories of methods to address this challenge: conditional analyses, haplotype-association tests, and Bayesian approaches. In conditional analyses, independence of a variant is tested after it is conditioned on the lead variant in the locus; if a nearby variant remains significant, it is considered as having a distinct additional signal. Haplotype-based association tests directly model combinations of alleles at multiple variants. In so doing, they can help researchers to understand the effect sizes and directions of the variants that are not fully independent from each other. Bayesian approaches use model selection strategies to identify independent signals of association within the same locus, leveraging the distribution of effect sizes and local patterns of linkage disequilibrium from reference panels. Here, through our systematic review we compiled allelic heterogeneity estimates for 129 loci associated

with autoimmune disease.[17–25] Across the 129 loci examined, 28 (22%) were reported to have more than one signal based on at least one study (Table S5).

Expectedly, the HLA locus (chr 6p21.32), known for its extreme polymorphism, extensive LD, and high gene content, was reported to display the largest number of independent signals associated with complex autoimmune diseases (Figure 3A). This locus was exceptional in that independent signals identified were common, large-effect-size protein-coding variants that act through multiple different genes in the locus. Although evidence for multiple independent autoimmune-disease signals in the HLA locus is mounting, results from this locus are particularly difficult to summarize because there is no consensus on how genetic variations in this locus should be captured and reported. The four-digit "classical HLA alleles," for instance, consider each unique protein sequence as a different "allele" and provide haplotype-level associations. Amino-acid-based models test each variable amino acid individually and provide a finer-scale resolution of the association signals. However, the complex nature of the HLA locus challenges the interpretation of the amino-acid-based models as well because a large fraction of amino acid positions in this locus are multi-allelic, and association p values are influenced by both the degree of phenotypic association and the level of complexity at each position.[21] For these reasons, despite the importance of the HLA locus in autoimmune diseases, throughout the rest of the manuscript we will focus primarily on results from the non-HLA loci. Among non-HLA loci with allelic heterogeneity, the NOD2 locus (chr 16q12.1) contained the largest number of independent signals associated with a complex autoimmune disease (Figures 3A and 3B). This locus as well had unique properties in that it also contained low-frequency, large-effect-size protein-coding association signals that had high penetrance.[26] In addition to the protein-coding variants with predicted deleterious effects (rs2066845, rs2066844, rs104895444, and rs104895467), the NOD2 locus also includes independent common non-coding variants (rs2357623 and rs72796367) and other rare variants with predicted benign (rs5743271) or unknown (rs184788345) effects that are associated with Crohn disease. In fact in most loci with allelic heterogeneity, there was at least one signal that was driven by a non-coding variant (Table S5), necessitating further gene-prioritization approaches to dissect whether all independent signals impact a single gene or whether multiple variants that act through different genes in the same locus contribute to the same disease. Given that the immune genes tend to cluster together in the human genome,[27] the latter possibility should also be considered while following up on these signals experimentally. Mendelian randomization approaches or eQTL-GWAS colocalization analyses that allow more than one causal variant in any given locus could also help with addressing this question.

We also noted that additional independent signals identified in conditional analyses were enriched to represent rare- or low-frequency variants with larger effect sizes than those of the lead variants (mean MAF = 0.25, mean risk-allele odds ratio = 1.26). the Welch's t test p value was 3.8×10^{-6} for MAF comparison and 0.018 for the odds-ratio comparison (Figure 3C and Table S5). Although in fewer loci, common variants have also been identified as additional signals in autoimmune-disease loci (Figures 3B–3D). Allelic heterogeneity is expected to hamper complex-disease-association signals as multiple trait-relevant variants with distinct phenotypic effects may cancel out each other's signals when population-based association studies are performed.[28,29] In essence, for loci that could be identified through GWAS, one population genetics theory predicts that there is likely negligible allelic heterogeneity or that the risk allele frequency/effect size of the predominant signal is significantly different than those of additional signals in the locus.[28,29] Alternatively, the depletion of allelic heterogeneity driven by multiple independent common variants could be due to the presence of several causal variants that are in high LD. It is important to note that when several causal variants are correlated, statistical approaches fail to differentiate between their effects, but we cannot rule out the existence of allelic heterogeneity either.

Causal variants and the extent of genetic overlap between different autoimmune diseases

Even when all variants in a locus are included in an association study, there is a reasonable chance that, as a result of statistical fluctuations, the lead variant representing each independent signal is not the true causal variant. For example, we simulated GWAS p values in a randomly chosen Crohn disease GWAS locus under different odds-ratio, sample-size, and ethnicity scenarios. Even in the "best-case" scenario where the association signal was driven by a relatively large-effect variant (OR = 1.3) in a well-powered study (n = 50,000, 1:1 case-control ratio) from an African American population sample (with expected small LD windows), there was only a 70% chance that the lead variant identified was the true causal variant (Figure 1B).

Given the observed pattern of association in a locus, Bayesian approaches can help to infer the probability that a variant is causal and can output a list of variants that include the causal variant with a certain probability threshold. Additionally, functional experiments including *in vivo* transgenic reporter assays; *in vitro* reporter, protein binding, and chromatin-accessibility assays; genome editing tools; and algorithms that integrate both functional annotations and statistical approaches could help prioritize the causal variants in loci associated with complex diseases. There is no gold standard in variant fine-mapping strategies, but consistent results across different studies and approaches can improve confidence in the fine-mapping findings.
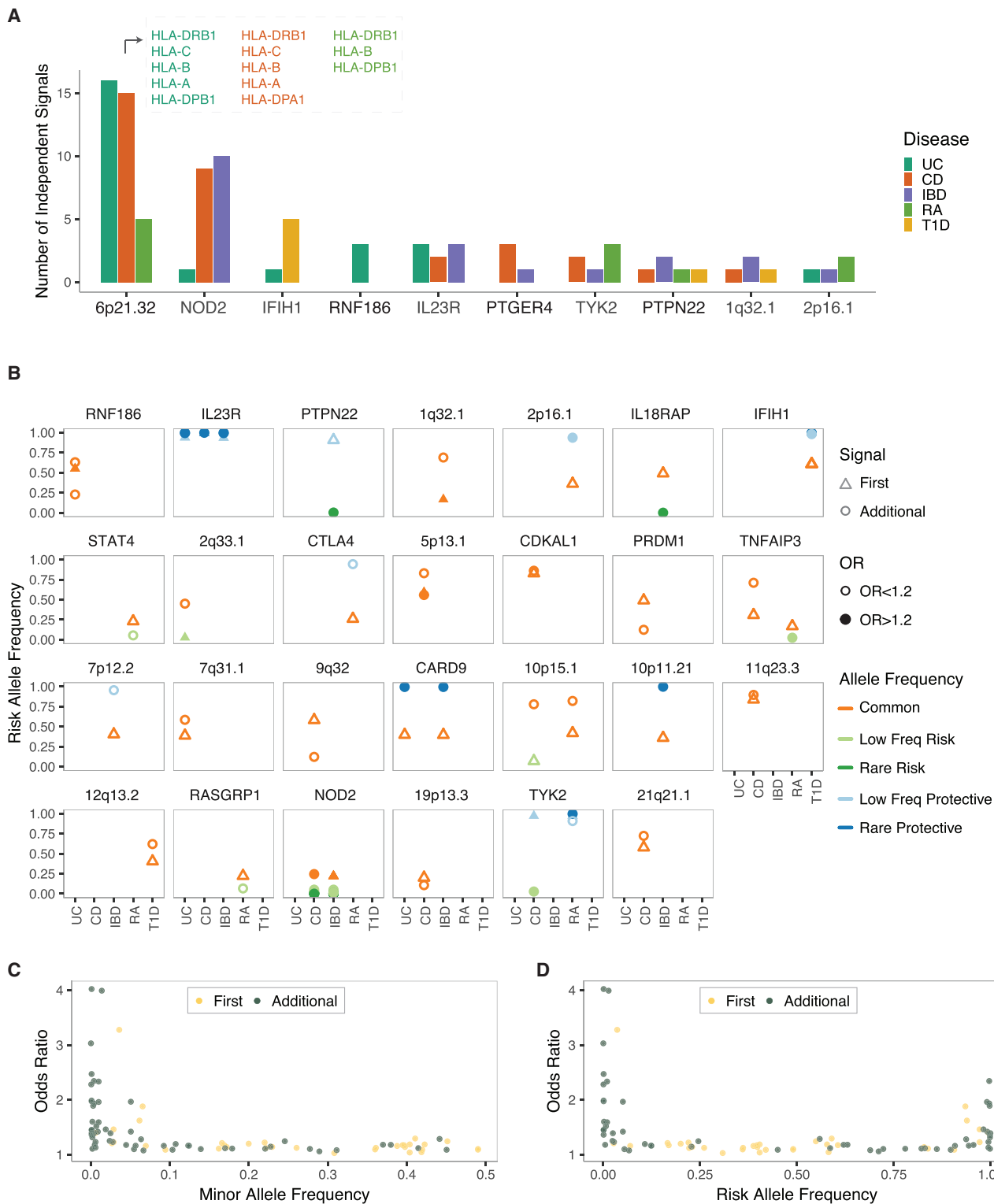
Figure 3. The extent of allelic heterogeneity in complex autoimmune diseases

(A) Example GWAS loci where allelic heterogeneity was reported for at least one autoimmune disease.

(B) Risk-allele frequency and effect size of the first and additional signals in non-HLA loci with allelic heterogeneity. Gene names are included only when a single gene was predominantly reported as the candidate trait-relevant gene in the literature. Note that the HLA locus is not included because of the complexity of the locus and that there are inconsistencies in how genetic associations in this locus were reported (i.e, reporting was at the haplotype level, amino acid level, and nucleotide level). Common-allele frequency (AF), AF > 0.1; low-allele frequency, 0.01 < AF < 0.1; and rare-allele frequency, AF < 0.01. Abbreviations are as follows: UC, ulcerative colitis; CD, Crohn disease; IBD, inflammatory bowel disease; RA, rheumatoid arthritis; T1D, and type 1 diabetes.

(C) The relationship between odds ratios and minor-allele frequencies.

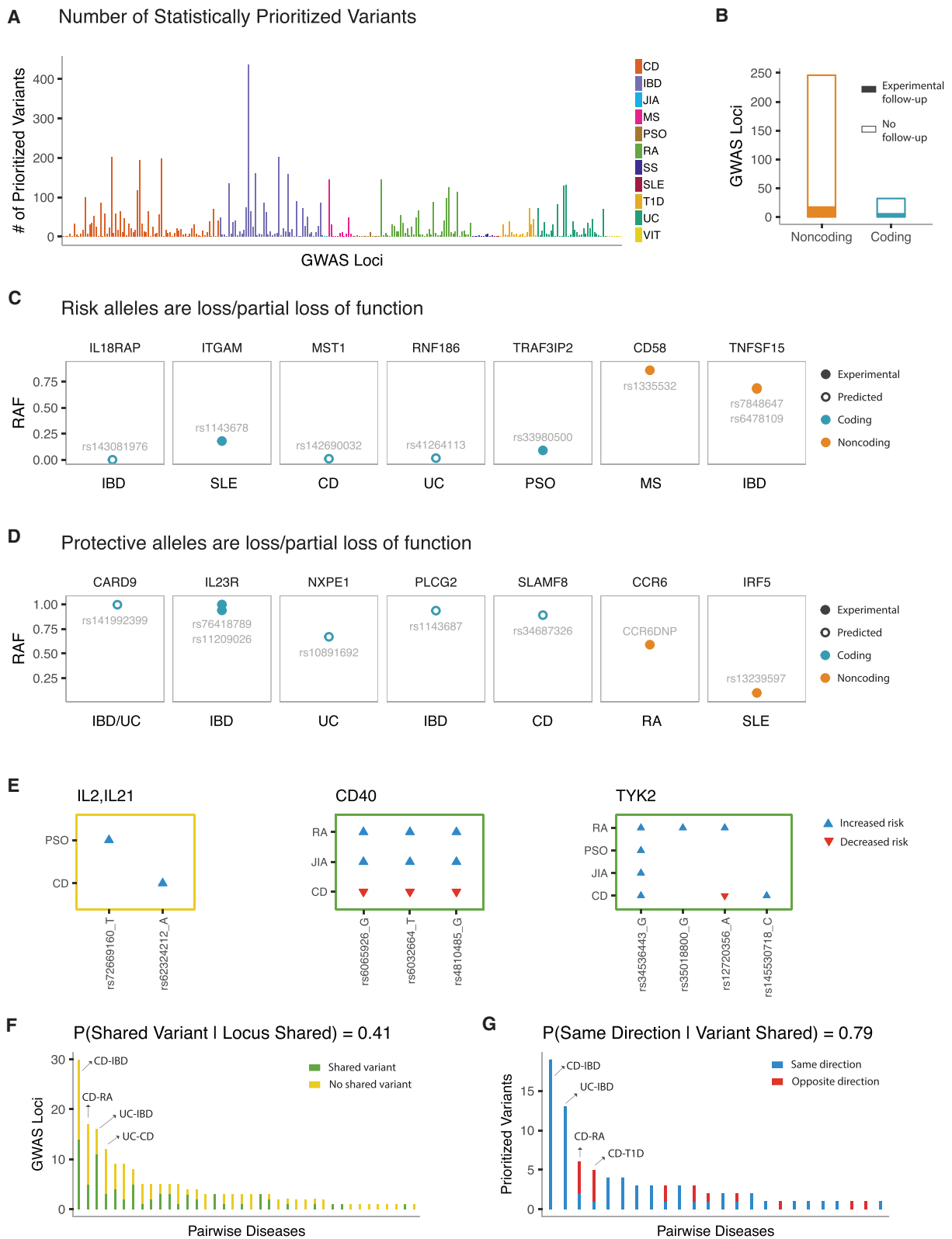(D) The relationship between odds ratios and risk-allele frequencies.

Figure 4. An overview of the candidate causal variants in autoimmune GWAS loci

(A) Number of prioritized variants based on statistical estimates across all locus-by-disease combinations.

(B) Locus-by-disease combinations stratified by whether a predicted damaging protein-coding variant was prioritized (n = 32) or not (n = 246). 22 signals (7 protein-coding, 15 non-coding) in total were experimentally followed up on.

(C) Example gene-disease pairs where loss- or partial-loss-of-function alleles increase risk of autoimmune disease.

(D) Example gene-disease pairs where loss- or partial-loss-of-function alleles protect against autoimmune disease.

(legend continued on next page)

Here we compiled 164 autoimmune-disease loci (Table S6) for which the causal variants were prioritized via at least one of the fine-mapping approaches shown in Figure 2. These 164 loci included 278 combinations of locus-by-disease association signals with 11 autoimmune diseases. The vast majority of the association signals (n = 257, 92.4%) were fine-mapped via statistical approaches.[3,2,17–20,22–25,30,31] A median of eight variants were prioritized on the basis of statistical estimates across all locus-by-disease combinations (Figure 4A). Only 22 (7.9%) signals were fine-mapped via functional experiments,[32–48] and 17 (6.1%) were fine-mapped via integrative approaches.[49,50]

At any disease-associated variant, the disease risk allele might increase or decrease the function of the target gene or might lead to acquisition of a new function. Elucidating the mechanism of action of the risk and protective alleles could help researchers to understand how target genes are affected and how they should be modulated in patient cohorts to reduce or eliminate the disease symptoms. Among fine-mapped autoimmune-disease signals, 32 (11.5%) included at least one predicted pLOF (based on the Variant Effect Predictor) or missense variant of predicted damaging effect (based on SIFT/PolyPhen). Five of these 32 signals were further validated with functional experiments (Tables S6 and S7). Although the vast majority of signals (246 out of 278; 88.5%) were non-coding, only 17 such signals were followed up on experimentally (Figure 4B, Table S6).

In silico classification of protein-coding variants as either deleterious or benign is imperfect,[51] and functional experiments at a given locus might not have assayed all statistically prioritized variants. Nonetheless, these approaches have the potential to improve understanding of the biological mechanisms and enable us to strategize interventions needed to mimic the effects of the protective alleles. For example, Sivanesan et al.[34] experimentally validated that protective alleles of the IL23R protein display loss of function due to impairment of receptor maturation and reduced protein stability, suggesting antagonists of IL23R as a relevant therapeutic strategy for treating IBD. Therapeutic inhibition of IL23 is an approved treatment for psoriasis. Here, in agreement with the causal role of IL23 signaling in autoimmune diseases, GWAS and fine-mapping experiments suggest that strategies that directly inhibit the IL23 receptor (IL23R) could be of even greater potential therapeutic value.[34] In another autoimmune GWAS locus, Richard et al.[35] used allele-specific expression measurements to demonstrate that IBD-protective alleles in the chromosome 9q32 locus enhance expression of *TNFSF15* (MIM: 604052),

suggesting agonists of TNFSF15 as potential therapies for treating autoimmune diseases. Similar examples of genes with fine-mapped loss- or partial-loss-of-function autoimmune-disease risk alleles included *IL18RAP* (MIM: 604509), *ITGAM* (MIM: 120980), *MST1* (MIM: 142408), *RNF186* (MIM: 617163), *TRAF3IP2* (MIM: 607043), and *CD58* (MIM: 153420),[32,35–37,41,45,46,48] implying agonists of the protein products of these genes as potential therapeutic interventions (Figure 4C and Tables S6 and S7). *CARD9* (MIM: 607212), *NXPE1*, *PLCG2* (MIM: 600220), *SLAMF8* (MIM: 606620), *CCR6* (MIM: 601835), and *IRF5* (MIM: 607218) were among the genes that had loss- or partial-loss-of-function-protective alleles,[33,34,39,40] predicting antagonists of the protein products of these genes as potential therapies (Figure 4D and Tables S6 and S7).

Genetic correlations across subsets of autoimmune diseases have been estimated from genetic-risk scores[52] and LD score regression.[53] Here, we asked a related question to estimate the extent of overlap in fine-mapping results across diseases at shared GWAS loci. We found that many loci implicated in more than one autoimmune disease displayed differences in the sets of prioritized variant(s) and/or direction of genotype effect on disease risk (e.g., Figure 4E, Table S6). Across loci with fine-mapping estimates for two or more autoimmune diseases, we found at least partial overlap among the prioritized variants in 41% of the pairwise comparisons (Figure 4F and Table S6). When the prioritized variant or variants were shared, we chose one variant to represent each independent signal and showed that the direction of effect was identical for 79.5% of the cases (Figure 4G and Tables S6 and S11). Although limited by the compiled fine-mapping manuscripts and not based on robust statistical signals such as evidence of a shared causal variant from pairwise colocalization analyses, these observations suggest that when the same locus is implicated in two or more autoimmune diseases, differences in the causal variants underlying each phenotype rather than differences in the direction of effects is the primary source of genetic discordance between different autoimmune phenotypes.

Assigning confidence scores to estimates of candidate genes to identify most consistently dissected GWAS loci

GWAS have clearly shown that most complex-disease associations are driven by non-coding variants (Figure 4B). Although protein-coding variants can be readily linked to

(E) Example loci with discordance in the set of prioritized variant(s) or direction of genotype effect on the risk of different autoimmune diseases.
(F) The x axis shows the 38 pairwise autoimmune disease combinations for which at least one GWAS locus is shared and fine-mapped for both diseases. The y axis shows the number of loci that harbor at least one shared prioritized variant (green) versus the number of loci that do not harbor any shared prioritized variant (yellow).
(G) The x axis shows the 24 pairwise autoimmune disease combinations that share at least one prioritized variant in at least one locus. The y axis shows the number of shared variant(s) that have the same direction of effect on both diseases (blue) versus the number of variants with opposite direction of effect on each disease (red). Note that when multiple prioritized variants overlapped for two diseases, only one variant per each independent association signal was included during assessment of the direction of effect.
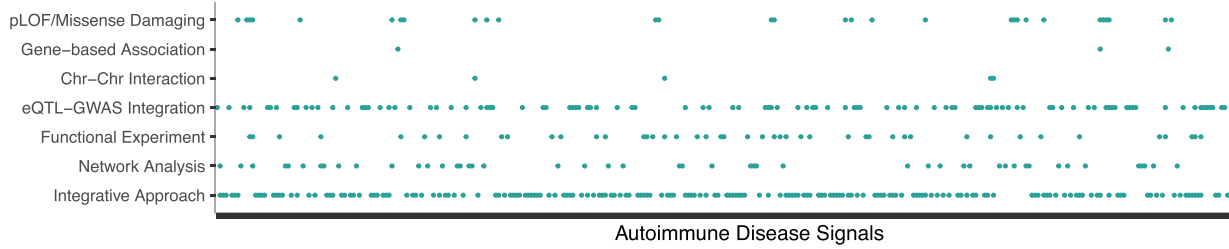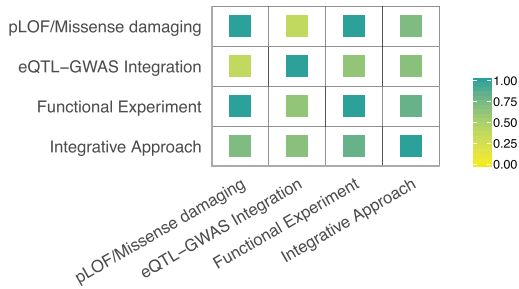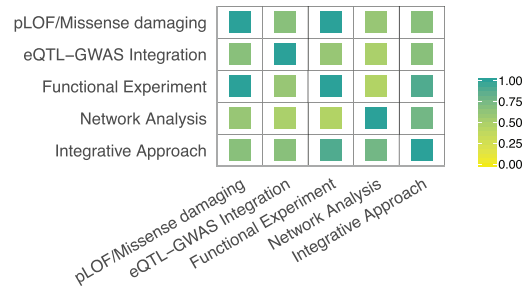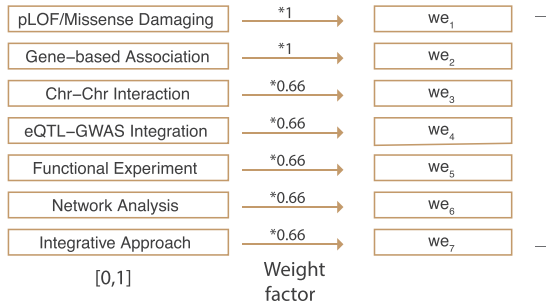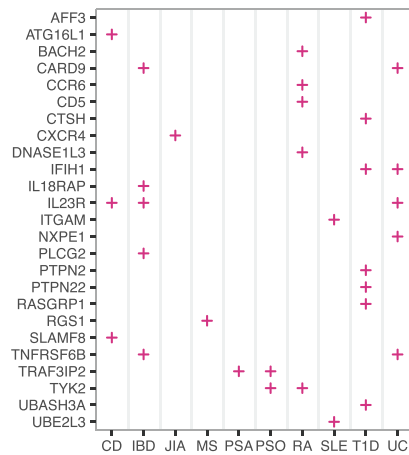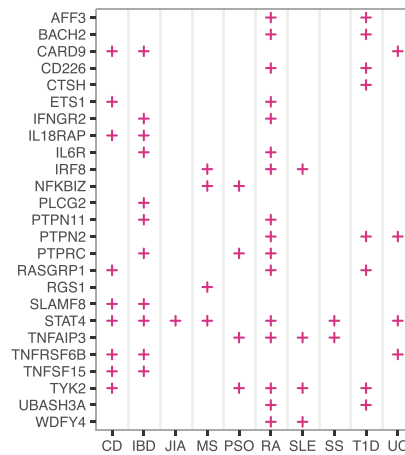
Figure 5. An integrated approach to assigning confidence scores to candidate-gene estimates

(A) Gene-prioritization evidence compiled for 342 combinations of locus-by-disease association signals across 191 unique loci and 15 autoimmune diseases. Detailed information about each locus and fine-mapping evidence are included in Table S8.

(B) This heatmap shows the frequency of at least one overlapping gene when the same locus-by-disease combination was dissected via multiple strategies. Results were plotted for pairwise strategies that co-dissected at least five locus-by-disease combinations.

the genes that they affect, identifying the gene targets of non-coding variants has not been straightforward. Each GWAS locus often spans multiple genes, and non-coding variants do not necessarily regulate the closest gene. Additionally, gene regulation can be context and tissue specific, which adds an extra layer of complexity to identifying the gene involved and predicting the direction of gene regulatory effect on disease risk. After a GWAS, many genome-wide studies have assayed intermediate molecular traits including epigenetic marks, gene expression levels, and chromatin interaction datasets.[3,30,54–66] An array of computational approaches have been developed to test the impact of genetic variation on these intermediate traits and to help researchers infer causal relationships between the molecular phenotypes and the downstream disease risk. Gene-based association tests of rare variants have been used to increase the statistical power and to enable direct identification of trait-relevant genes.[45,67,68] Functions of the compelling candidate genes in GWAS loci have been tested with *in vitro* experiments and model organisms.[32,34,45,69–91] Additionally, protein-protein interaction and co-expression networks have been integrated with GWAS results to both identify trait-relevant genes and pinpoint key network modules important in disease pathogenesis. In this study, we compiled candidate gene estimates from 342 combinations of locus-by-disease association signals across 191 unique loci and 15 autoimmune diseases (Table S8, Figure 5A). We had seven broad categories of gene-prioritization approaches (Figures 2 and 5A). We note that we only considered the results of studies that focused on identifying *cis* gene targets of GWAS variants. Approaches that focus on the *trans* gene targets or interacting pairs of the GWAS *cis* gene targets were beyond the scope of our study.

When the same association signal was dissected by different approaches, we found that there was often at least one overlap among the prioritized gene(s) (Figures 5B and 5C). Several integrative algorithms have been developed to rank genes in GWAS loci by incorporating results from different gene-prioritization approaches.[92,93] These algorithms mainly rely on integrating results of genome-wide functional datasets such as genome-wide QTLs and chromatin interactions. However, they do not capture locus-specific evidence that often comes from GWAS follow-up studies that experimentally characterize one or a few GWAS loci.[32–35,37,38,40,42,44,45,47,48,69–73,75–88,90,91] For example, while available integrative approaches would capture the colocalization signal between *BACH2* (MIM:

605394) eQTLs and type 1 diabetes, they would not integrate the results of a GWAS follow-up study that functionally linked the BACH2 protein to type 1 diabetes.[80] Furthermore, a gene such as *CCR6* (MIM: 601835) could have been missed if locus-specific TALEN gene editing evidence in this locus had not been considered, as there is no reported statistical colocalization between the *CCR6* eQTLs and autoimmune-disease GWASs.[40] Additionally, previously published integrative approaches[92,93] allow prioritization of multiple genes in the same disease locus. Although it is biologically possible that multiple trait-relevant genes exist in the same locus, identifying loci where a single gene was consistently implicated by different approaches is crucial for drug-discovery scientists to prioritize among dozens or hundreds of loci with gene candidates. To address these key gaps for complex autoimmune diseases, we integrated the candidate gene predictions we compiled in a single framework. For each prioritized gene *g*, we calculated a weighted total evidence score (E_g) as follows:

$$E_g = \sum_{i=1}^{7} we_i,$$

where we_i corresponds to the weighted evidence that led to the prioritization of gene *g* on the basis of each gene-prioritization category *i*. We chose the weight factors of each gene-prioritization category *i* to reflect the consensus of confidence in evidence. The presence of predicted loss-of-function (pLOF) or missense damaging variant(s) among statistically prioritized variants in a GWAS locus is considered as the most direct evidence implying the trait-relevant genes. As such, we down-weighted all prioritization categories relative to "pLOF/missense damaging" and "gene-based association" categories (Figure 5D). For each prioritized gene *g* in each locus *l*, we then computed a confidence score (CS_{gl}), which adjusts the total evidence score (E_g) of each gene by the presence or absence of evidence supporting other genes (S) in the same locus:

$$CS_{gl} = E_g - \left(\max(E_g | g \in s) \right).$$

We applied this approach to each autoimmune disease separately (i.e., by calculating Eg scores based on evidence from individual diseases) (Table S9 and Figure 5E) and also performed a disease agnostic analysis where we considered gene prioritization evidence across all autoimmune diseases that had a GWAS signal in the same locus jointly (Table S10 and Figure 5F). Locus-based disease agnostic

(C) This heatmap shows the frequency of at least one overlapping gene when the same autoimmune disease locus was dissected via multiple strategies. Results were plotted for pairwise strategies that co-dissected at least five autoimmune-disease loci.
(D) Integration of gene-prioritization approaches to calculating confidence scores.
(E) Top 25 prioritized genes when each autoimmune disease was considered individually (i.e., gene prioritization evidence of each disease was integrated separately even when multiple autoimmune diseases had GWAS signals and genes prioritized in the same locus).
(F) Top 25 prioritized genes from disease agnostic analysis where gene-prioritization evidence across all autoimmune diseases was considered jointly in shared loci. Gene and disease names are sorted by alphabetical order. Abbreviations are as follows: CD, Crohn disease; IBD, inflammatory bowel disease; JIA, juvenile idiopathic arthritis; MS, multiple sclerosis; PSO, psoriasis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SS, systemic sclerosis; T1D, type 1 diabetes; and UC, ulcerative colitis.

analysis can be advantageous in loci where a single gene underlies multiple autoimmune-disease associations. In contrast, this same approach can introduce noise to confidence scores when different genes in the same GWAS locus underlie different autoimmune diseases. Colocalization analyses[5] across diseases could be used to guide use of disease agnostic approach. However, it is worth noting that the absence of a shared causal variant (either as a result of true absence or because of a lack of statistical power) does not rule out the presence of a shared trait-relevant gene underlying different autoimmune diseases. For example, different causal variants in distinct genomic elements might control regulation of the same gene to influence risk of multiple autoimmune diseases. As such, when a disease agnostic approach is used, a further comparison of the prioritized genes for each individual disease might be necessary to assess whether combining functional evidence could be justified or not. A comparison of the results of these two analyses (Tables S9 and S10) suggests that in most loci there is one candidate gene underlying different autoimmune diseases, and combining information across GWAS follow-up studies of different autoimmune phenotypes might help with the gene prioritization efforts. However, there are loci where different genes have been prioritized for different autoimmune diseases. For example, in the chromosome 6q27 locus, CCR6 (MIM: 601835) was prioritized for rheumatoid arthritis,[40,55] and RNASET2 (MIM: 612944) was prioritized for Grave disease, vitiligo, and Crohn disease.[59,64,72] Similarly, in the chromosome 11q12.2 locus, although CD6 (MIM: 186720) was prioritized for multiple sclerosis and Crohn disease,[22,62,85] CD5 (MIM: 153340) was the gene prioritized for rheumatoid arthritis via multiple approaches.[20,55]

By applying both these strategies to the published GWAS follow-up evidence for 15 autoimmune diseases, we prioritized 189 candidate trait-relevant genes (Tables S9 and S10). Among the genes with top confidence scores were known drug target genes (e.g., CD5 [MIM: 153340], CXCR4 [MIM: 162643], IL6R [MIM: 147880], and TYK2 [MIM: 176941]), genes with known immune-related functions (e.g., IFIH1 [MIM: 606951], IL18RAP [MIM: 604509], STAT4 [MIM: 600558], and TNFAIP3 [MIM: 191163]), and genes that are lesser characterized in the context of autoimmune diseases (e.g., AFF3 [MIM: 601464], NXPE1, and UBE2L3 [MIM: 603721]). Although not definitive, the approach we used provided a straightforward means of integrating multiple sources of gene-prioritization evidence and allowed us to distinguish consistently dissected GWAS loci from those with conflicting results and/or multiple genes with similar degree of functional support.

Discussion

Genetic variation underlies differences in disease pathogenesis. After the discovery of DNA, human geneticists largely focused on understanding how rare, large-effect size mutations can cause rare diseases that run in families with predictable inheritance patterns. Only with the development of DNA microarray and high-throughput sequencing technologies in the 2000s did it become possible to reveal the genetic architecture of common, complex diseases. Remarkable advances have been achieved in complex-trait genetics in the past 15 years. The current focus of many researchers in the field is to translate this growing genetic knowledge into therapeutics with the ultimate goal of reducing the burden of these diseases worldwide.

Human genetics has the potential to transform healthcare, and it is imperative that this transformation impact all affected human lives equally. Yet, there is an unfortunate lack of diversity in the subjects included in GWASs,[94] most of which are based on populations of European ancestry. There are many reports indicating that lack of ethnic diversity in human genetic studies might lead to inaccurate predictions of disease risk and lack of interventions in under-studied populations.[94] Moreover, trans-ethnic analyses of GWAS results across diverse populations can take advantage of population-specific LD patterns and narrow down disease-causing variants that are shared between populations.[7,8] Despite studies reporting that associations of SNPs with autoimmune diseases are largely consistent across populations,[95,96] we were unable to find autoimmune trans-ethnic studies that fully and equally captured human genetic variation across different populations. Specifically, in most trans-ethnic studies, genotyping arrays used in non-European populations were designed for capturing genetic variation in Europeans. Hence, these studies were reporting a rather incomplete and potentially biased picture. As research moves forward, we see great need and promise in carefully designed GWAS and fine-mapping studies in diverse human populations. Such studies not only will uncover the full genetic architecture underlying human disease but also have the potential to improve the risk prediction and precision in medical care.

Another major obstacle that we encountered while reviewing GWAS fine-mapping studies was the difficulty of summarizing the results from the HLA locus. Genetic variation in the HLA locus represents the strongest predisposing signal for almost all autoimmune diseases[97] and yet it is practically impossible to review the fine-mapping evidence in this locus in a systematic way. Among the characteristic features of the HLA locus are the high degree of polymorphism and strong LD. As such, fine mapping in this locus is by default challenging. An additional layer of complexity is added when fine-mapping studies in this locus report their association results at different molecular resolutions (i.e., haplotype level, amino acid level, and nucleotide level). Direct conversions between different molecular resolutions are not always possible. There are also many multi-allelic variants with different numbers of distinct alleles in the HLA locus. A commonly ignored potential issue in HLA fine-mapping studies is that

association p values could be inflated by the degree of complexity of the variants.[21] Therefore, p values attributed to different models with variable variant complexities could be misleading during prioritization of causal variants. Clinical laboratories have a relatively long history of using classical HLA alleles at different resolutions depending on the patient needs. As the field moves forward, we anticipate that at least within the GWAS fine-mapping community, there will be better-standardized ways to capture, report, and analyze the variation in this complex and important locus.

Matching therapeutic modality to genetic mechanism of action has the potential to facilitate the translational research in a way that had not been possible before GWASs.[98] For protein-coding variants with predicted damaging effects, it is relatively straightforward to determine whether the risk or the protective allele has the deleterious effect on the protein function. The genetic mechanism of action for non-coding variants is often predicted with gene expression levels as an intermediate phenotype between genotype and the disease outcome. Although such approaches prioritize trait-relevant genes more accurately than baselines, interpretation of these analyses could be convoluted.[99] First, true or artifactual co-regulation might lead to multiple gene prioritizations within the same locus. Second, even when a single gene is prioritized in a locus, the allelic direction of effect on gene expression might be tissue specific. Hence, it might not be possible to infer a therapeutic strategy if the mechanistically disease-related tissue is unknown or when genotype effect on gene expression is opposite in two potentially disease-relevant tissues. We expect that overcoming these challenges will involve expansion of current gene-expression panels to cover additional cell and stimulation types as well as methodologies that focus on identifying the relevant cell type for each disease or each locus-by-disease combination. Additionally, revealing causal intermediate phenotypes (eg., blood cell indices) of complex diseases could improve the inference on causal variants and genes in GWAS loci, in particular when fine-mapping signals of disease and their underlying intermediate phenotypes are considered jointly. Until all loci are fine-mapped and genetic-based therapeutic modalities are reliably predicted, we see great benefit in allocating resources to the most consistently dissected loci where a single gene is clearly and consistently prioritized on the basis of multiple sources of functional evidence.

In the very early stages of GWASs, Hindorff et al. realized the importance of forming a catalog of GWAS results, manually curated the published SNP-trait associations, and built the commonly used NHGRI GWAS Catalog.[100] As a collective effort and largely in a project-centric model, researchers then started sharing full summary statistics to follow up on and fine-map primary GWAS findings. Ongoing efforts include standardizing heterogeneous formats and harmonizing content of GWAS summary statistics.[1] The availability of a harmonized, downloadable catalog of GWAS summary statistics has facilitated many aspects of GWAS fine-mapping studies. Here, in parallel to these efforts, we present a systematic approach cataloging and summarizing published GWAS fine-mapping evidence for autoimmune diseases. We believe the work presented here could expedite the process of translating GWAS findings into therapeutics and be expanded into collaborative efforts to catalog and evaluate published fine-mapping evidence for all complex human diseases.

Data and code availability

The code generated during this study is available at https://github.com/minalc44/autoimmune/. Supplemental information includes all datasets analyzed during this study.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.03.009.

Declaration of interests

M.C. is an employee of Bristol-Myers Squibb. J.C.M. is an employee and a shareholder of Bristol-Myers Squibb. C.D.B. declares no competing interests.

References

1. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (D1), D1005–D1012.

2. Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

3. de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat. Genet. *49*, 256–261.

4. Fortune, M.D., and Wallace, C. (2019). simGWAS: a fast method for simulation of large scale case-control GWAS summary statistics. Bioinformatics *35*, 1901–1906.

5. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

6. Cannon, M.E., and Mohlke, K.L. (2018). Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. Am. J. Hum. Genet. *103*, 637–653.

7. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.

8. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet. *102*, 717–730.

9. Price, A.L., Spencer, C.C., and Donnelly, P. (2015). Progress and promise in understanding the genetic basis of common diseases. Proc. Biol. Sci. *282*, 20151684.

10. King, E.A., Davis, J.W., and Degner, J.F. (2019). Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet. *15*, e1008489.

11. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. Nat. Genet. *47*, 856–860.

12. Zhernakova, A., van Diemen, C.C., and Wijmenga, C. (2009). Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat. Rev. Genet. *10*, 43–55.

13. Richard-Miceli, C., and Criswell, L.A. (2012). Emerging patterns of genetic overlap across autoimmune disorders. Genome Med. *4*, 6.

14. González-Serna, D., Ochoa, E., López-Isac, E., Julià, A., Degenhardt, F., Ortego-Centeno, N., Radstake, T.R.D.J., Franke, A., Marsal, S., Mayes, M.D., et al.; Scleroderma Genetic Consortium (2020). A cross-disease meta-GWAS identifies four new susceptibility loci shared between systemic sclerosis and Crohn's disease. Sci. Rep. *10*, 1862.

15. Wang, K., Baldassano, R., Zhang, H., Qu, H.Q., Imielinski, M., Kugathasan, S., Annese, V., Dubinsky, M., Rotter, J.I., Russell, R.K., et al. (2010). Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum. Mol. Genet. *19*, 2059–2067.

16. Landrum, M.J., and Kattman, B.L. (2018). ClinVar at five years: Delivering on the promise. Hum. Mutat. *39*, 1623–1630.

17. Beaudoin, M., Goyette, P., Boucher, G., Lo, K.S., Rivas, M.A., Stevens, C., Alikashani, A., Ladouceur, M., Ellinghaus, D., Törkvist, L., et al.; Quebec IBD Genetics Consortium; NIDDK IBD Genetics Consortium; and International IBD Genetics Consortium (2013). Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet. *9*, e1003723.

18. Cortes, A., Field, J., Glazov, E.A., Hadler, J., Stankovich, J., Brown, M.A., Brown, M.A.; and ANZgene Consortium (2013). Resequencing and fine-mapping of the chromosome 12q13-14 locus associated with multiple sclerosis refines the number of implicated genes. Hum. Mol. Genet. *22*, 2283–2292.

19. Diogo, D., Bastarache, L., Liao, K.P., Graham, R.R., Fulton, R.S., Greenberg, J.D., Eyre, S., Bowes, J., Cui, J., Lee, A., et al. (2015). TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. PLoS ONE *10*, e0122271.

20. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K., et al.; Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; and Wellcome Trust Case Control Consortium (2012). High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. *44*, 1336–1340.

21. Goyette, P., Boucher, G., Mallon, D., Ellinghaus, E., Jostins, L., Huang, H., Ripke, S., Gusareva, E.S., Annese, V., Hauser, S.L., et al.; International Inflammatory Bowel Disease Genetics Consortium; Australia and New Zealand IBDGC; Belgium IBD Genetics Consortium; Italian Group for IBD Genetic Consortium; NIDDK Inflammatory Bowel Disease Genetics Consortium; United Kingdom IBDGC; Wellcome Trust Case Control Consortium; and Quebec IBD Genetics Consortium (2015). High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat. Genet. *47*, 172–179.

22. Huang, H., Fang, M., Jostins, L., Umićević Mirkov, M., Boucher, G., Anderson, C.A., Andersen, V., Cleynen, I., Cortes, A., Crins, F., et al.; International Inflammatory Bowel Disease Genetics Consortium (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature *547*, 173–178.

23. Hunt, K.A., Mistry, V., Bockett, N.A., Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O., et al. (2013). Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature *498*, 232–235.

24. Nejentsev, S., Walker, N., Riches, D., Egholm, M., and Todd, J.A. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science *324*, 387–389.

25. Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; and International Inflammatory Bowel Disease Genetics Consortium (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. *43*, 1066–1073.

26. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M., et al. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature *411*, 599–603.

27. Trowsdale, J. (2002). The gentle art of gene arrangement: the meaning of gene clusters. Genome Biol. *3*, COMMENT 2002.

28. Reich, D.E., and Lander, E.S. (2001). On the allelic spectrum of human disease. Trends Genet. *17*, 502–510.

29. Pritchard, J.K., and Cox, N.J. (2002). The allelic architecture of human disease genes: common disease-common variant...or not? Hum. Mol. Genet. *11*, 2417–2423.

30. López-Isac, E., Acosta-Herrera, M., Kerick, M., Assassi, S., Satpathy, A.T., Granja, J., Mumbach, M.R., Beretta, L., Simeón, C.P., Carreira, P., et al.; European Scleroderma Group†; and Australian Scleroderma Interest Group (ASIG) (2019). GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nat. Commun. *10*, 4955.

31. Keene, K.L., Quinlan, A.R., Hou, X., Hall, I.M., Mychaleckyj, J.C., Onengut-Gumuscu, S., and Concannon, P. (2012). Evidence for two independent associations with type 1 diabetes at the 12q13 locus. Genes Immun. *13*, 66–70.

32. Wang, S., Wen, F., Wiley, G.B., Kinter, M.T., and Gaffney, P.M. (2013). An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression. PLoS Genet. *9*, e1003750.

33. Thynn, H.N., Chen, X.F., Hu, W.X., Duan, Y.Y., Zhu, D.L., Chen, H., Wang, N.N., Chen, H.H., Rong, Y., Lu, B.J., et al. (2020). An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. J. Invest. Dermatol. *140*, 348–360.e311.

34. Sivanesan, D., Beauchamp, C., Quinou, C., Lee, J., Lesage, S., Chemtob, S., Rioux, J.D., and Michnick, S.W. (2016). IL23R (Interleukin 23 Receptor) Variants Protective against Inflammatory Bowel Diseases (IBD) Display Loss of Function due to Impaired Protein Stability and Intracellular Trafficking. J. Biol. Chem. *291*, 8673–8685.

35. Richard, A.C., Peters, J.E., Savinykh, N., Lee, J.C., Hawley, E.T., Meylan, F., Siegel, R.M., Lyons, P.A., and Smith, K.G.C. (2018). Reduced monocyte and macrophage TNFSF15/TL1A expression is associated with susceptibility to inflammatory bowel disease. PLoS Genet. *14*, e1007458.

36. Ong, L.T., Tan, H.F., Feng, C., Qu, J., Loh, S.C., Bhattacharyya, S., and Tan, S.M. (2017). The Systemic Lupus Erythematosus-Associated Single Nucleotide Polymorphism rs1143678 in Integrin αM Cytoplasmic Tail Generates a 14-3-3ζ Binding Site That Is Proinflammatory. J. Immunol. *198*, 883–894.

37. Mitkin, N.A., Muratova, A.M., Korneev, K.V., Pavshintsev, V.V., Rumyantsev, K.A., Vagida, M.S., Uvarova, A.N., Afanasyeva, M.A., Schwartz, A.M., and Kuprash, D.V. (2018). Protective C allele of the single-nucleotide polymorphism rs1335532 is associated with strong binding of Ascl2 transcription factor and elevated CD58 expression in B-cells. Biochimica et biophysica acta Molecular basis of disease. Biochim. Biophys. Acta Mol. Basis Dis. *1864*, 3211–3220.

38. Matthews, S.M., Eshelman, M.A., Berg, A.S., Koltun, W.A., and Yochum, G.S. (2019). The Crohn's disease associated SNP rs6651252 impacts MYC gene expression in human colonic epithelial cells. PLoS ONE *14*, e0212850.

39. Li, G., Martínez-Bonet, M., Wu, D., Yang, Y., Cui, J., Nguyen, H.N., Cunin, P., Levescot, A., Bai, M., Westra, H.J., et al. (2018). High-throughput identification of noncoding functional SNPs via type IIS enzyme restriction. Nat. Genet. *50*, 1180–1188.

40. Li, G., Cunin, P., Wu, D., Diogo, D., Yang, Y., Okada, Y., Plenge, R.M., and Nigrovic, P.A. (2016). The Rheumatoid Arthritis Risk Variant CCR6DNP Regulates CCR6 via PARP-1. PLoS Genet. *12*, e1006292.

41. Hüffmeier, U., Uebe, S., Ekici, A.B., Bowes, J., Giardina, E., Korendowych, E., Juneblad, K., Apel, M., McManus, R., Ho, P., et al. (2010). Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat. Genet. *42*, 996–999.

42. Gregory, A.P., Dendrou, C.A., Attfield, K.E., Haghikia, A., Xifara, D.K., Butter, F., Poschmann, G., Kaur, G., Lambert, L., Leach, O.A., et al. (2012). TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. Nature *488*, 508–511.

43. Gao, P., Uzun, Y., He, B., Salamati, S.E., Coffey, J.K.M., Tsalikian, E., and Tan, K. (2019). Risk variants disrupting enhancers of T_H1 and T_{REG} cells in type 1 diabetes. Proc. Natl. Acad. Sci. USA *116*, 7581–7590.

44. Ellinghaus, D., Zhang, H., Zeissig, S., Lipinski, S., Till, A., Jiang, T., Stade, B., Bromberg, Y., Ellinghaus, E., Keller, A., et al. (2013). Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. Gastroenterology *145*, 339–347.

45. Cardinale, C.J., Wei, Z., Panossian, S., Wang, F., Kim, C.E., Mentch, F.D., Chiavacci, R.M., Kachelries, K.E., Pandey, R., Grant, S.F., et al. (2013). Targeted resequencing identifies defective variants of decoy receptor 3 in pediatric-onset inflammatory bowel disease. Genes Immun. *14*, 447–452.

46. Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.A., and Mann, M. (2012). Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet. *8*, e1002982.

47. Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., Mari, B., Barbry, P., Mosnier, J.F., Hébuterne, X., et al. (2011). A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat. Genet. *43*, 242–245.

48. Adrianto, I., Wen, F., Templeton, A., Wiley, G., King, J.B., Lessard, C.J., Bates, J.S., Hu, Y., Kelly, J.A., Kaufman, K.M., et al.; BIOLUPUS and GENLES Networks (2011). Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat. Genet. *43*, 253–258.

49. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. *47*, 955–961.

50. Kichaev, G., and Pasaniuc, B. (2015). Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. Am. J. Hum. Genet. *97*, 260–271.

51. Ghosh, R., Oak, N., and Plon, S.E. (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. *18*, 225.

52. Grassmann, F., Kiel, C., Zimmermann, M.E., Gorski, M., Grassmann, V., Stark, K., Heid, I.M., Weber, B.H.; and International AMD Genomics Consortium (IAMDGC) (2017). Genetic pleiotropy between age-related macular degeneration and 16 complex diseases and traits. Genome Med. *9*, 29.

53. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al.; ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

54. Wallace, C., Rotival, M., Cooper, J.D., Rice, C.M., Yang, J.H., McNeill, M., Smyth, D.J., Niblett, D., Cambien, F., Tiret, L., et al.; Cardiogenics Consortium (2012). Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. Hum. Mol. Genet. *21*, 2815–2824.

55. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; and GARNET consortium (2014). Genetics of

rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

56. Odhams, C.A., Cunninghame Graham, D.S., and Vyse, T.J. (2017). Profiling RNA-Seq at multiple resolutions markedly increases the number of causal eQTLs in autoimmune disease. PLoS Genet. *13*, e1007071.

57. Mo, X., Guo, Y., Qian, Q., Fu, M., Lei, S., Zhang, Y., and Zhang, H. (2020). Mendelian randomization analysis revealed potential causal factors for systemic lupus erythematosus. Immunology *159*, 279–288.

58. McGovern, A., Schoenfelder, S., Martin, P., Massey, J., Duffus, K., Plant, D., Yarwood, A., Pratt, A.G., Anderson, A.E., Isaacs, J.D., et al. (2016). Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. Genome Biol. *17*, 212.

59. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

60. Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., et al. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat. Commun. *6*, 10069.

61. Márquez, A., Kerick, M., Zhernakova, A., Gutierrez-Achury, J., Chen, W.M., Onengut-Gumuscu, S., González-Álvaro, I., Rodriguez-Rodriguez, L., Rios-Fernández, R., González-Gay, M.A., et al.; Coeliac Disease Immunochip Consortium; Rheumatoid Arthritis Consortium International for Immunochip (RACI); International Scleroderma Group; and Type 1 Diabetes Genetics Consortium (2018). Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. Genome Med. *10*, 97.

62. Hussman, J.P., Beecham, A.H., Schmidt, M., Martin, E.R., McCauley, J.L., Vance, J.M., Haines, J.L., and Pericak-Vance, M.A. (2016). GWAS analysis implicates NF-κB-mediated induction of inflammatory T cells in multiple sclerosis. Genes Immun. *17*, 305–312.

63. Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A., and Wallace, C. (2015). 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.

64. Fryett, J.J., Inshaw, J., Morris, A.P., and Cordell, H.J. (2018). Comparison of methods for transcriptome imputation through application to two common complex diseases. Eur. J. Hum. Genet. *26*, 1658–1667.

65. Fotsing, S.F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., and Gymrek, M. (2019). The impact of short tandem repeat variation on gene expression. Nat. Genet. *51*, 1652–1659.

66. Davison, L.J., Wallace, C., Cooper, J.D., Cope, N.F., Wilson, N.K., Smyth, D.J., Howson, J.M., Saleh, N., Al-Jeffery, A., Angus, K.L., et al.; Cardiogenics Consortium (2012). Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Hum. Mol. Genet. *21*, 322–333.

67. Motegi, T., Kochi, Y., Matsuda, K., Kubo, M., Yamamoto, K., and Momozawa, Y. (2019). Identification of rare coding variants in *TYK2* protective for rheumatoid arthritis in the Japanese population and their effects on cytokine signalling. Ann. Rheum. Dis. *78*, 1062–1069.

68. Finkel, T.H., Li, J., Wei, Z., Wang, W., Zhang, H., Behrens, E.M., Reuschel, E.L., Limou, S., Wise, C., Punaro, M., et al. (2016). Variants in CXCR4 associate with juvenile idiopathic arthritis susceptibility. BMC Med. Genet. *17*, 24.

69. Zhou, J., Wu, R., High, A.A., Slaughter, C.A., Finkelstein, D., Rehg, J.E., Redecke, V., and Häcker, H. (2011). A20-binding inhibitor of NF-κB (ABIN1) controls Toll-like receptor-mediated CCAAT/enhancer-binding protein β activation and protects from inflammatory disease. Proc. Natl. Acad. Sci. USA *108*, E998–E1006.

70. Yuan, Q., Li, Y., Li, J., Bian, X., Long, F., Duan, R., Ma, X., Gao, F., Gao, S., Wei, S., et al. (2018). WDFY4 Is Involved in Symptoms of Systemic Lupus Erythematosus by Modulating B Cell Fate via Noncanonical Autophagy. J. Immunol. *201*, 2570–2578.

71. Yang, L., Feng, D., Bi, X., Stone, R.C., and Barnes, B.J. (2012). Monocytes from Irf5-/- mice have an intrinsic defect in their response to pristane-induced lupus. J. Immunol. *189*, 3741–3750.

72. Wang, Q., Jiang, M., Wu, J., Ma, Y., Li, T., Chen, Q., Zhang, X., and Xiang, L. (2014). Stress-induced RNASET2 overexpression mediates melanocyte apoptosis via the TRAF2 pathway in vitro. Cell Death Dis. *5*, e1022.

73. Waddell, A., Ahrens, R., Steinbrecher, K., Donovan, B., Rothenberg, M.E., Munitz, A., and Hogan, S.P. (2011). Colonic eosinophilic inflammation in experimental colitis is mediated by Ly6C(high) CCR2(+) inflammatory monocyte/macrophage-derived CCL11. J. Immunol. *186*, 5993–6003.

74. Tsoi, L.C., Spain, S.L., Ellinghaus, E., Stuart, P.E., Capon, F., Knight, J., Tejasvi, T., Kang, H.M., Allen, M.H., Lambert, S., et al. (2015). Enhanced meta-analysis and replication studies identify five new psoriasis susceptibility loci. Nat. Commun. *6*, 7001.

75. Scott-Hewitt, N.J., Folts, C.J., Hogestyn, J.M., Piester, G., Mayer-Pröschel, M., and Noble, M.D. (2017). Heterozygote galactocerebrosidase (GALC) mutants have reduced remyelination and impaired myelin debris clearance following demyelinating injury. Hum. Mol. Genet. *26*, 2825–2837.

76. Santin, I., Moore, F., Colli, M.L., Gurzov, E.N., Marselli, L., Marchetti, P., and Eizirik, D.L. (2011). PTPN2, a candidate gene for type 1 diabetes, modulates pancreatic β-cell apoptosis via regulation of the BH3-only protein Bim. Diabetes *60*, 3279–3288.

77. Peer, S., Cappellano, G., Hermann-Kleiter, N., Albrecht-Schgoer, K., Hinterleitner, R., Baier, G., and Gruber, T. (2018). Regulation of Lymphatic GM-CSF Expression by the E3 Ubiquitin Ligase Cbl-b. Front. Immunol. *9*, 2311.

78. Moore, F., Colli, M.L., Cnop, M., Esteve, M.I., Cardozo, A.K., Cunha, D.A., Bugliani, M., Marchetti, P., and Eizirik, D.L. (2009). PTPN2, a candidate gene for type 1 diabetes, modulates interferon-gamma-induced pancreatic beta-cell apoptosis. Diabetes *58*, 1283–1291.

79. Mehto, S., Jena, K.K., Nath, P., Chauhan, S., Kolapalli, S.P., Das, S.K., Sahoo, P.K., Jain, A., Taylor, G.A., and Chauhan, S. (2019). The Crohn's Disease Risk Factor IRGM Limits NLRP3 Inflammasome Activation by Impeding Its Assembly and by Mediating Its Selective Autophagy. Mol. Cell *73*, 429–445.e427.

80. Marroquí, L., Santin, I., Dos Santos, R.S., Marselli, L., Marchetti, P., and Eizirik, D.L. (2014). BACH2, a candidate risk gene

for type 1 diabetes, regulates apoptosis in pancreatic β-cells via JNK1 modulation and crosstalk with the candidate gene PTPN2. Diabetes 63, 2516–2527.

81. Liu, Z., Lee, J., Krummey, S., Lu, W., Cai, H., and Lenardo, M.J. (2011). The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. Nat. Immunol. 12, 1063–1070.

82. Lindahl, H., Guerreiro-Cacais, A.O., Bedri, S.K., Linnerbauer, M., Lindén, M., Abdelmagid, N., Tandre, K., Hollins, C., Irving, L., Glover, C., et al. (2019). IL-22 Binding Protein Promotes the Disease Process in Multiple Sclerosis. J. Immunol. 203, 888–898.

83. Leshchiner, E.S., Rush, J.S., Durney, M.A., Cao, Z., Dančík, V., Chittick, B., Wu, H., Petrone, A., Bittker, J.A., Phillips, A., et al. (2017). Small-molecule inhibitors directly target CARD9 and mimic its protective variant in inflammatory bowel disease. Proc. Natl. Acad. Sci. USA 114, 11392–11397.

84. Lapaquette, P., Glasser, A.L., Huett, A., Xavier, R.J., and Darfeuille-Michaud, A. (2010). Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly. Cell. Microbiol. 12, 99–113.

85. Kofler, D.M., Severson, C.A., Mousissian, N., De Jager, P.L., and Hafler, D.A. (2011). The CD6 multiple sclerosis susceptibility allele is associated with alterations in CD4+ T cell proliferation. J. Immunol. 187, 3286–3291.

86. Khor, B., Conway, K.L., Omar, A.S., Biton, M., Haber, A.L., Rogel, N., Baxt, L.A., Begun, J., Kuballa, P., Gagnon, J.D., et al. (2019). Distinct Tissue-Specific Roles for the Disease-Associated Autophagy Genes ATG16L2 and ATG16L1. J. Immunol. 203, 1820–1829.

87. Hartmann, F.J., Khademi, M., Aram, J., Ammann, S., Kockum, I., Constantinescu, C., Gran, B., Piehl, F., Olsson, T., Codarri, L., and Becher, B. (2014). Multiple sclerosis-associated IL2RA polymorphism controls GM-CSF production in human TH cells. Nat. Commun. 5, 5056.

88. Dominguez, S., Montgomery, A.B., Haines, G.K., 3rd, Bloomfield, C.L., and Cuda, C.M. (2017). The caspase-8/RIPK3 signaling axis in antigen presenting cells controls the inflammatory arthritic response. Arthritis Res. Ther. 19, 224.

89. Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., and Scacheri, P.C. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 24, 1–13.

90. Castro-Dopico, T., Dennison, T.W., Ferdinand, J.R., Mathews, R.J., Fleming, A., Clift, D., Stewart, B.J., Jing, C., Strongili, K., Labzin, L.I., et al. (2019). Anti-commensal IgG Drives Intestinal Inflammation and Type 17 Immunity in Ulcerative Colitis. Immunity 50, 1099–1114.e10.

91. Anderson, A.C., Sullivan, J.M., Tan, D.J., Lee, D.H., and Kuchroo, V.K. (2015). A T cell extrinsic mechanism by which IL-2 dampens Th17 differentiation. J. Autoimmun. 59, 38–42.

92. Fang, H., De Wolf, H., Knezevic, B., Burnham, K.L., Osgood, J., Sanniti, A., Lledó Lara, A., Kasela, S., De Cesco, S., Wegner, J.K., et al.; ULTRA-DD Consortium (2019). A genetics-led approach defines the drug target landscape of 30 immune-related traits. Nat. Genet. 51, 1082–1091.

93. Koscielny, G., An, P., Carvalho-Silva, D., Cham, J.A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., et al. (2017). Open Targets: a platform for therapeutic target identification and validation. Nucleic Acids Res. 45 (D1), D985–D994.

94. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. Cell 177, 26–31.

95. Yin, X., Low, H.Q., Wang, L., Li, Y., Ellinghaus, E., Han, J., Estivill, X., Sun, L., Zuo, X., Shen, C., et al. (2015). Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. Nat. Commun. 6, 6916.

96. Okada, Y., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Kawaguchi, T., Stahl, E.A., Kurreeman, F.A., Nishida, N., et al. (2012). Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat. Genet. 44, 511–516.

97. Matzaraki, V., Kumar, V., Wijmenga, C., and Zhernakova, A. (2017). The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. 18, 76.

98. Marshall, E. (2013). Biomedicine. NIH seeks better database for genetic diagnosis. Science 342, 27.

99. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. 51, 592–599.

100. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 106, 9362–9367.