Journal Pre-proof

Triangulating molecular evidence to prioritize candidate causal genes at established atopic dermatitis loci

Maria K. Sobczyk, Tom G. Richardson, Verena Zuber, Josine L. Min, Tom R. Gaunt, Lavinia Paternoster, eQTLGen Consortium, BIOS Consortium, GoDMC

PII: S0022-202X(21)01160-X

DOI: https://doi.org/10.1016/j.jid.2021.03.027

Reference: JID 2912

To appear in: *The Journal of Investigative Dermatology*

Received Date: 8 January 2021

Revised Date: 12 March 2021

Accepted Date: 25 March 2021

Please cite this article as: Sobczyk MK, Richardson TG, Zuber V, Min JL, Gaunt TR, Paternoster L, eQTLGen Consortium, BIOS Consortium, GoDMC, Triangulating molecular evidence to prioritize candidate causal genes at established atopic dermatitis loci, *The Journal of Investigative Dermatology* (2021), doi: https://doi.org/10.1016/j.jid.2021.03.027.

Triangulating molecular evidence to prioritize candidate causal genes at established atopic dermatitis loci

Maria K Sobczyk (ORCID: 0000-0003-0000-4100)[1], Tom G Richardson (ORCID: 0000-0002-7918-2040)[1], Verena Zuber (ORCID: 0000-0001-9827-1877) [2,3], Josine L Min (ORCID: 0000-0003-4456-9824)[1], eQTLGen Consortium, BIOS Consortium, GoDMC, Tom R Gaunt (ORCID: 0000-0003-0924-3247)[1], Lavinia Paternoster (ORCID: 0000-0003-2514-0889)[1]

1) MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK

2) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

3) MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

Work was done in Bristol, United Kingdom.

Short title: Prioritizing genes at atopic dermatitis loci

Corresponding author: Dr Lavinia Paternoster, MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Bristol BS8 2BN, UK

E-mail: l.paternoster@bristol.ac.uk Telephone: +44 117 3310 135 Twitter: @drpaternoster

Abbreviations:

AD: atopic dermatitis; caQTL - chromatin accessibility quantitative trait locus; AS: ankylosing spondylitis; CD: Crohn's disease; DGE: differential gene expression; eQTL: expression quantitative trait locus; GTEx: Genotype-Tissue Expression project; GWAS:

genome-wide association study; hQTL: histone quantitative trait locus; LCL – lymphoblastoid cell line; LD: linkage disequilibrium; MHC: major histocompatibility complex; mQTL: DNA methylation quantitative trait locus; MS: multiple sclerosis; PP: posterior probability; pQTL: protein quantitative trait locus; RA: rheumatoid arthritis; SLE: Systemic lupus erythematosus; SNP: single nucleotide polymorphisms; sQTL: splicing quantitative trait locus; T1D: type I diabetes; UC: ulcerative colitis

ABSTRACT

Genome-wide association studies for atopic dermatitis (AD) have identified 25 reproducible loci. We attempt to prioritize candidate causal genes at these loci using extensive molecular resources compiled into a bioinformatics pipeline.

We identified a list of 103 molecular resources for AD aetiology, including expression, protein and DNA methylation QTL datasets in skin or immune-relevant tissues which were tested for overlap with GWAS signals. This was combined with functional annotation using regulatory variant prediction, and features such as promoter-enhancer interactions, expression studies and variant fine-mapping. For each gene at each locus, we condensed the evidence into a prioritization score.

Across the investigated loci, we detected significant enrichment of genes with adaptive immune regulatory function and epidermal barrier formation among the top prioritized genes. At 8 loci, we were able to prioritize a single candidate gene (*IL6R*, *ADO*, *PRR5L*, *IL7R*, *ETS1*, *INPP5D*, *MDM1*, *TRAF3*). In addition, at 6 of the 25 loci, our analysis prioritizes less familiar candidates (*SLC22A5*, *IL2RA*, *MDM1*, *DEXI*, *ADO*, *STMN3*).

Our analysis provides support for previously implicated genes at several AD GWAS loci, as well as evidence for plausible additional candidates at others, which may represent potential targets for drug discovery.

INTRODUCTION

Defined by inflamed dry, hyperplastic eczematous skin and pruritus, atopic dermatitis (AD) is among the world's top 50 common diseases, with prevalence in 2010 estimated at close to 230 million cases and increasing (Hay et al. 2014). AD is highly heritable - with estimates of up to 75% in twin studies (Elmose and Thomsen 2015). The largest and most recent genome-wide association study (GWAS) of AD, undertaken by the EAGLE consortium in 2015 identified 25 loci associated with AD in individuals of European descent (Paternoster et al. 2015). The majority of disease-associated variants are located in non-coding regions, implying that they have a regulatory role rather than affecting protein function. Thus, integrating various biological data resources can provide complementary evidence about GWAS causal genes (Hormozdiari et al. 2018).

Since publication of the EAGLE GWAS, there has been an explosion of new datasets from many cell types and new methods that offer an opportunity to refine prioritization of genes at the GWAS loci. In this paper, we aim to comprehensively dissect AD GWAS loci by prioritizing candidate causal genes and illuminating biological mechanisms through which candidate genes can impact AD risk. We integrate several established fine-mapping and gene prioritization methods in a unique AD-focused gene prioritization pipeline to comprehensively evaluate the causal genetic evidence at each locus and utilize an exhaustive set of 103 molecular datasets in AD-relevant tissues to best support these methods. We explicitly model our assumptions about the importance of different types of evidence as well as strength of the associations relating the features to genes and variants. Our pipeline in combining these methods generates a score for each gene, used to assess the magnitude of evidence of each tested gene at a locus of being causal. Such a score can serve as a metric

which allows rapid gene prioritization by molecular biologists and other interested parties, such as pharmaceutical companies.

RESULTS

Identification of key tissues and cell types in AD GWAS loci

In order to determine which tissues and cell types should be part of the pipeline, we tested for enrichment of expression at our GWAS loci across a wide range of tissues and cell types (53 tissues from GTEx ver.7 and 79,249,533 cell types from the Gene Atlas, Immunological Genomics and FANTOM CAGE, respectively) and determined that all immune cell, skin (including fibroblast), spleen and whole blood datasets should be included (see Supplementary Results). We reviewed the literature to identify 103 separate datasets from these tissue-types with relevant data (Figure S1, S2).

Prioritization of candidate genes

Gene prioritization scores ranged from 0 to 1405 (SNP scores ranged from 0.5 to 968) (Dataset S1). For 8 loci the top prioritized SNP was not the index SNP, and for 10 loci the closest gene did not score best (Table 1). In detailing the results, we focus on genes ranked in the top 3 and SNPs ranked in top 10 at each locus as this limit agrees with the sharp score decay observed in scores (Figure S3 & S4, Dataset S2).

Excluding the complex MHC locus, the highest gene scores were seen for genes at 5 loci: *IL18R1* (score=1384) and *IL18RAP* (score=1341) at the 2q12.1 locus, *PPP2R3C* (score=996) at the 14q13.2 locus, *IL7R* (score=965) at the 5p13.2 locus, *TRAF3* (score=848) at 14q32.32 locus and *IL6R* (score=743) at 1q21.3 locus (Table 1, Figure 1; Dataset S3 for all loci). Assuming that the true model is one of a single causal gene at each locus, prioritization can also be evaluated by comparing the score of the top prioritized gene at a locus with all other

genes at that locus. Eight loci (1q21.3-*IL6R*, 10q21.2-*ADO*, 11p13-*PRR5L*, 5p13.2-*IL7R*, 11q24.3-*ETS1*, 2q37.1-*INPP5D*, 12q15-*MDM1*, 14q32.32-*TRAF3*; Table 1) have a single stand-out candidate causal gene – with the top gene contributing more than 50% of the total score of top 10-ranked genes. The top candidate by that metric is *PRR5L* (79% of top 10 genes at 11p13 locus), with a score of 598 compared to 65 for the second-ranked gene at this locus. Most top-prioritized genes by the total score are also prioritized by this metric. Two further loci show good evidence (>75% cumulative score) shared across two candidate genes (*IL18R1* and *IL18RAP at 2q12.1 and EMSY* and *LRRC32* at 11q13.5, which share 77% and 84% of the cumulative score respectively). At 2q12.1 (where *IL18R1* and *IL18RAP* reside) there is evidence for two independent genetic signals, and these may affect each of the prioritized genes.

For five loci, the pipeline prioritizes genes in top position (and with a score >300) that were not considered in the original GWAS annotation (Paternoster et al. 2015); *MDM1* at 12q15 (score=728), *ADO* at 10q21.2 (score=615), *STMN3* at 20q13.33 (score=608), *SLC22A5* at 5q31.1 (score=461) and *DEXI* at 16p13.13 (score=376). Some in this list (such as *SLC22A5*) represent promising candidates.

For each locus, as well as evaluating the overall prioritization scores of each gene, we present a summary figure that shows how different evidence sources have contributed to the overall score (Figure S5) - the loci with the most compelling evidence are displayed in Figure 2. In addition, the individual results from each source are also available for deeper evaluation (Dataset S4). Full discussion of each locus in Table 1 integrating evidence from the pipeline with knowledge from literature is available in Supplementary Results.

Validation of gene prioritization

In the absence of "gold standard" true positive genes to which we could compare our prioritization of candidate genes at GWAS loci, we evaluated our results in two indirect ways. Firstly, we tested if our top 3 prioritized genes across all loci are enriched in any gene sets, using enrichr (Kuleshov et al. 2016) and compared those with categories enriched among previously implicated AD genes (Table S1). We found that both lists are significantly enriched for immune system-related genes (Figure 3), but often with stronger evidence in our prioritized gene sets. In particular, cytokine categories were overrepresented, e.g. GO cytokine-mediated signalling pathway (adjusted *p*-value for our prioritized genes = $1x10^{-9}$ versus 0.004 for other previously implicated AD genes). The genes in the cytokine pathways identified by the pipeline include *IL6R*, *IL22, INPP5D, IL2RA*, *IFNG*, *IL18R1*, *IL18RAP*, *IL1RL1* and *IL7R*. Signalling involved in regulation of response to interferon γ (GO, p=0.039 ver. 0.043), JAK1-/JAK2-STAT3-interacting genes and JAK-STAT signalling pathway in general (KEGG, $p=4x10^{-5}$ ver. $2x10^{-4}$), also overlapped between the two gene sets, as terms relating to T cell differentiation. We did not find enrichment of genes in any specific type of immunity – with all of Th1, Th2, Th17, Th22 represented and previously shown to play a role in certain subsets of AD patients, despite overall particular importance of Th2 and Th22 (Esaki et al. 2016; Leung and Guttman-Yassky 2014; Suárez-Fariñas et al. 2013). Genes concerned with establishment of the skin barrier were marginally enriched for in the pipeline (due to the prioritization of cornified envelope genes, *HRNR* and *RPTN*), but less than the previously reported AD genes (GO, $p=0.045$ ver. $8x10^{-8}$, Table S2).

The second way we validated our results was to test if our candidates interacted with each other and with the genes with established roles in AD pathogenesis using STRING (Szklarczyk et al. 2019) to visualize the highest-confidence interactions. The analysis revealed an extensive network that included 25 prioritized genes, centred on key immune regulators (see Supplementary Results and Figure S6).

DISCUSSION

Previous annotations of AD GWAS loci have been limited in their ability to identify likely causal genes (Paternoster et al. 2015). Here we provide a thorough investigation of the 25 European AD loci, by integrating all relevant available data that can be used to provide evidence for hypothesizing causal genes, and combine this data in such a way as to produce a ranking for every gene at each locus.

As there are a vast number of methods that can be employed to attempt to establish the causal genes for GWAS signals, we integrate several of these, which represent the most useful and robust approaches that span experimentally generated functional annotations, predictions for regulatory impact generated by machine learning models, as well as linking back to AD physiology through evaluation of differential gene expression and DNA methylation studies and proteome comparisons involving eczema patients.

We employed the most robust methods where possible, for example statistical methods (coloc, TWAS) were used to formally compare the association patterns in QTL studies and GWAS when full summary statistics were available, as ~50% of common variants are associated with one eQTL or more across 53 tissues in GTEx (Liu et al. 2019) so simple lookups for variant overlap alone will result in many false positives. Where full summary statistics were not available, we still included such look-ups, but gave such evidence much lower weight in the overall score (weight adjustment of 2 compared to 20 for colocalization).

For 10 loci the top ranked gene is not the gene closest to the index GWAS SNP. Eight loci have a single stand-out candidate causal gene (score >50% of the top 10 gene cumulative score) and 7 genes score particularly high (>700) and/or have a particular stand-out score (>75%). Whilst in many cases our analysis strengthens the evidence for existing candidate

causal genes at these loci, at 6 loci our score ranks alternative candidates as the most likely causal gene.

One of these 6 can be considered an interesting validation of our approach. *IL15RA* was previously considered the most plausible candidate gene at the 10p15.1 locus due to the limited eQTL evidence that was available at the time. Our approach however prioritized *IL2RA* over *IL15RA*. Since the publication of the GWAS in 2015 this locus has been followed up with CRISPR experiments, which reported that the T-allele at rs61839660 down-regulates IL2RA expression (Simeonov et al. 2017), suggesting that our prioritization at this locus is correct.

At another locus - 11q13.5, experimental evidence has emerged supporting candidate role of the top two prioritised genes - LRRC32 (encoding the GARP receptor) and EMSY. Rare missense mutations found in LRRC32 in eczema patients decrease GARP expression on the activated T regulatory cell surface and reduce conversion of naïve T cells into T regulatory cells (Manz et al. 2016). On the other hand, EMSY has been characterized as a potent regulator of skin barrier formation (Elias et al. 2019). Another top-prioritised gene with recent evidence for role in skin barrier formation is KIF3A (locus 5q31.1b, Stevens et al. 2020), further details in supplementary results.

Other validations of our approach are provided by tests of enrichment of ontology terms and evidence of protein-protein interactions amongst the top ranked genes across all loci. Enrichment was found for ontology terms associated with: skin barrier integrity, T helper cell polarization, cytokine signalling and JAK-STAT signalling. The importance of JAK-STAT signalling has recently been highlighted by its enrichment among genes prioritized for inflammatory skin diseases (including AD) with HiChIP-derived T cell enhancer connectome (Jeng et al. 2019) and over-representation of rare coding variants in JAK1/2 in a new AD

study (Mucha et al. 2019). In investigating protein-protein interactions with the STRING database among our prioritized candidate genes and other established candidates, interactions between genes with immune regulation (but not skin barrier) functions were found amongst the established AD players: *TSLP* and its receptor, *TLR2*, *STAT6*, *IL4* and interferon γ receptor. STRING data is not entirely comprehensive and omits other functional relationships between prioritized genes, described in Supplementary Results.

In general, the results of our GWAS prioritization analysis remind us that interpretation of a GWAS locus is complicated due to varying regulation between cell types and widespread co-regulation that makes identification of the true causal gene difficult. Indeed, recent GWAS research reveals that on top of each locus being able to contain multiple signals (Mahajan et al. 2018), each signal can influence multiple co-regulated genes (Cannon and Mohlke 2018). Associations with molecular phenotypes follow the same pattern, with at least 9% of human eQTLs quantified to contain secondary signals (Wood et al. 2011) and multiple genes implicated for 50% of human eQTLs (Gamazon et al. 2018). According to the multiple enhancer variant hypothesis, several variants in LD can influence multiple enhancers and cooperatively affect expression of target gene(s). Corradin *et al.* (2014) provide evidence for it in 6 autoimmune diseases, including RA, Crohn's disease and SLE (Corradin et al. 2014). Therefore, it is not surprising that many of our loci showed multiple colocalizations for different genes and tissues, especially in gene-dense regions, with the caveat that not all may be causal. A recent analysis of the TWAS colocalization method claims that around 75% of hits will be non-causal in the instance of correlated gene expression at the locus (Wainberg et al. 2019), and we hypothesize that may be the case at loci 11q13.1, 14q13.2, and 20q13.33, where expression of as many as 4-6 genes colocalizes with AD GWAS signal in the TWAS results. Still, due to a distinct possibility of detection of multiple target genes and variants at a locus, we do not focus only on top-rated hits in our gene and variant ranking. AD GWAS loci

which we believe should be further experimentally investigated in that regard, include: 2q12.1 (*IL18R1*, *IL18RAP*, *IL1R1*), 5q31.1 (*KIF3A*, *PDLIM4*, *SLC22A4*, *IRF1*) and 20q13.33 (*STMN3*, *LIME1*, *ARFRP1*) – the first two especially due to containing at least two independent signals in the GWAS analysis.

Most of the genes with eQTL colocalization across tissues exhibit the same direction of effect, e.g. *PRR5L* (at 11p13), where the protective allele is associated with increased expression in the skin, whole blood and immune cell subsets. However, at three loci (2q12.1, 14q13.2, and 20q13.33) there may be tissue-dependent effects on expression, with opposite directions of effect on *STMN3, LIME1, APFRP1, IL18RAP* and *PP2R3C*. This indicates that causal variants potentially reside in tissue type-specific regulatory regions and context-dependent effect of these genes could impact atopic dermatitis phenotype.

Our pipeline for follow-up of GWAS signals, whilst focused on the integration of AD-relevant resources in this use case, can be adapted for other diseases or traits, following identification of the most relevant molecular datasets. The best evidence would come from consistent and clear prioritisation of a single gene from multiple sources (e.g. variants of interest at a locus showing physical interaction with enhancers and promoters of the same genes implicated by eQTL and pQTL data and validation of such genes in differential expression analyses, all in consistent cell/tissue types). However, for several reasons this situation is uncommon. Available datasets include evidence from limited tissues and cell states - reflecting transcriptional dynamics which are often transient, and low basepair resolution offered by high-throughput Hi-C which results in large, non-specific overlap regions (Mora et al. 2016). Ideally, data on specific blood and skin cell types would be available rather than bulk tissue, which will average out any cell-specific signals (Cano-Gamez and Trynka 2020). Furthermore, available sources do not cover the full spectrum of

variants or genes/proteins and so absence of evidence cannot be equated to evidence of absence. Predictions will improve as evidence from across more tissue types, especially at a single-cell resolution become available. Such rich datasets are already being generated for related disorders, such as asthma (Vieira Braga et al. 2019); considering trans- and isoform-level mechanism of action, and explicitly modelling network connectivity via protein-protein interactions and co-expression. It is also important to note that all the methods described in the pipeline are purely correlational, and so will require experimental manipulation for establishing causality of target genes, via e.g. CRISPR screening.

Our gene prioritisation score method assigns weight to different evidence sources, effectively up-weighting evidence with expected lower false discovery rate (such as TWAS and coloc) – which are also rarer and down-weighting weaker evidence (such as single eQTL look-ups) – that have been shown to often be purely coincidental and are numerous, so could easily overwhelm the overall score. There is currently no consensus on the best way to quantitatively integrate such evidence. Prior efforts for single trait GWAS annotation have taken other approaches: assigning equal weights (Schlosser et al. 2020) - which has obvious down-sides or attempting automatic weight assignment (Schwartzentruber et al. 2021) - which essentially optimised for closest genes. A promising approach uses 'gold standard' gene assignments at select GWAS loci for training (Ghoussaini et al. 2021), however, this type of method requires a number of GWAS as input with evidence sources limited to those relevant to all traits and selection bias inherent to the choice of "gold standard genes" used for training. It is of note that many different approaches all up-weight colocalization evidence, in agreement with our pipeline. Whilst there is some arbitrariness in our weighting assumptions, we believe our score calculation procedure has clear assumptions and justifiably balances some of the trade-offs.

Whilst there are limitations in our approach, as outlined in the sections above, we find it useful as an approach to easily flag the genes where we find most evidence which can then be carefully evaluated and potentially characterized as future drug targets. Loci where we are more confident in prioritization of single genes, especially lend themselves to direct experimental investigation, such as *TRAF3* at the 14q32.32 locus and *PRR5L* at the 11p13 locus. Additionally, investigating loci with clear candidate genes and association with multiple inflammatory diseases showing consistent direction of effect, such as 11p13 (*PRR5L* – MS, asthma), 11q24 (*ETS1* - psoriasis, celiac disease) and 16p13.13 (*DEXI* and *CLEC16A* - T1D, MS, alopecia areata, SLE, asthma) may reveal promising targets with potential drug repurposing future. Others with opposing direction of effect may reveal potential adverse side effects for consideration in therapeutic development (e.g. with anti-IL6 biologics for RA).

MATERIALS AND METHODS

This is an abbreviated version. For additional technical details, see Supplementary Materials and Methods.

Source GWAS

We investigate 25 loci, which show either genome-wide significance and for novel loci are replicated in independent European ancestry sample (21 loci), or are significant loci prioritized by the gene set enrichment analysis presented in the original paper (Paternoster et al. 2015).

Bayesian fine-mapping

To identify likely causal genetic variants in the regions harbouring AD GWAS signals, we used three different Bayesian fine-mapping methods: Finemap (Benner et al. 2016), fastPaintor (Kichaev et al. 2017) and JAM (Newcombe et al. 2016). Each method relies on

different prior assumptions and model formulation leading to divergent results (Cannon et al. 2017). The aim of our finemapping was not necessarily to identify the causal variants per se, but to prioritize SNPs which in turn provide evidence for what genes in the region are likely to be causal (further details in Supplementary M&M).

Variant filtering

In subsequent gene analyses, described below, we limited ourselves to SNPs within the region in significant LD with the index SNP in 1000 Genomes EUR population, henceforth referred to as the GWAS locus interval. The region in each case was defined by the positions of the furthest away 5' and 3' SNP with r2>=0.2 relative to the index SNP (limited to a maximum of 500kb in either direction). All the SNPs within that boundary were then considered (further details in Supplementary M&M).

Identification of key tissues and cell types

In order to focus on key tissues/cell types associated with eczema variants, we used gene set enrichment in SNPSea (Slowikowski et al. 2014) with the supplied gene expression datasets: Gene Atlas Affymetrix expression in 79 human tissues (Su et al. 2004), Immunological Genome Project (Heng et al. 2008) Affymetrix expression in 249 murine blood cell types and FANTOM CAGE (Kawaji et al. 2014) in 533 human cell types.

Secondly, we used MAGMA (de Leeuw et al. 2015) gene enrichment analysis on GTEx 7.0 (GTEx Consortium 2017) data as carried out by FUMA (Watanabe et al. 2017) (further details in Supplementary M&M).

eQTL identification

We used genotype array data and RPKM (reads per kilobase of transcript, per million mapped reads)-normalized expression in lymphoblastoid cell line (LCL) and skin tissue from

the TwinsUK cohort (Buil et al. 2015). *cis*-eQTLs 1.5Mbp upstream and downstream of TSS were identified using linear mixed model implemented in GEMMA (Zhou and Stephens 2012). eQTL associations were identified using the Wald test.

In the analysis involving the CEDAR cohort (Momozawa et al. 2018), we used the publicly available data: imputed genotypes and normalized gene expression values from blood and intestinal cell types ($CD4^+$ T lymphocytes, $CD8^+$ T lymphocytes, $CD19^+$ B lymphocytes, $CD14^+$ monocytes, $CD15^+$ granulocytes, platelets, ileum, colon, rectum). We used GEMMA's linear mixed model and Wald test to re-identify *cis*-eQTLs within 1.5Mbp upstream and downstream of TSS (further details in Supplementary M&M).

Colocalization with coloc and TWAS

We obtained full summary statistic results for cis-eQTLs detected in whole blood in the eQTLGen dataset (Võsa et al. 2018) – accessed on 08/08/2018, eQTLs from GTEx ver.7 dataset identified in the following tissues: whole blood, spleen, sun-exposed and unexposed skin, transformed fibroblasts and EBV-transformed lymphocytes, eQTLs published from the study investigating monocyte response to microbe-associated molecular patterns (Kim-Hellmuth et al. 2017), eQTLs in the monocytes, neutrophils and CD4+ T cells from the BLUEPRINT project (Chen et al. 2016), and pQTLs from whole blood in the Sun et al. dataset (Sun et al. 2018) as well as TwinsUK and CEDAR eQTLs identified above (Dataset S5). Subsequently, colocalization signal between betas from GWAS and eQTLs/pQTLs for genes within 1.5Mbp upstream and downstream of index SNP was evaluated with the coloc (Giambartolomei et al. 2014) R package. In coloc analysis, we considered loci with posterior probability of H4 (PPH4) >0.5 as informative enough to be included (Table S3, as done previously (Kim-Hellmuth et al. 2019); with H4 stating the hypothesis of both traits being associated and sharing a single causal variant.

We also carried out a TWAS (Gusev et al. 2016) analysis, where reference datasets with gene expression and genotype data (GTEx ver.7.0, CEDAR and TwinsUK) were used to predict gene expression in our target GWAS. The analysis pipeline for the Summary-based Mendelian Randomization analysis has been described previously (Richardson et al. 2020) (further details in Supplementary M&M).

Complementary gene prioritization methods

To further prioritize GWAS gene targets, we used two gene prioritization methods: regfm (Shooshtari et al. 2017) and PrixFixe (Taşan et al. 2015). PrixFixe strategy relies on prioritization of groups of candidate genes from multiple GWAS loci based on 'cofunction' networks (CFNs). Regfm's workflow involves intersection of fine-mapped credible interval SNPs with consensus DHS sites and genes whose expression they control predicted based on ROADMAP (Roadmap Epigenomics Consortium et al. 2015) chromatin accessibility and gene expression data to prioritize target genes.

Variant functional prediction

KGGSeq (Li et al. 2016) was used to measure non-coding variant regulatory potential and coding variant deleteriousness using functional scores derived by combining scores from 7 algorithms. fathmm-XF (Rogers et al. 2018), GWAS4D (Huang et al. 2018) and fitCons (Gulko et al. 2015) were also used independently. Overlap with ChIP-Seq defined binding sites of transcriptional regulators was cross-referenced in the ReMap2018 database (Chèneby et al. 2018). Splicing regulatory potential of variants was evaluated with SPIDEX (Xiong et al. 2014).

We also looked at variant overlap within different regulatory regions: insulator (Wang et al. 2015), promoter-enhancer interactions (9 studies), regulatory non-coding RNAs (5 studies), topologically associating domains (TADs, 6 studies), and CTCF binding sites (Ziebarth et al.

2012) using giggle (Layer et al. 2018) search engine (further details in Supplementary M&M).

Independent lookups

We have also performed gene and variant lookups among published significant results (see Dataset S5 for references) from 29 eQTL studies, 3 mQTL (including GoDMC results, Min et al. 2020), 2 pQTL studies, 2 hQTL studies, and a caQTL study where full GWAS results were not available, as well as differential expression (5 studies), DNA methylation (2 studies) and 2 proteome comparisons between skin in AD patients and healthy controls. We also interrogated the GWAS Catalog (Milano et al. 2016) (accessed on 11/01/2019) for any variants that have been identified as genome-wide significant in previous GWAS studies on related inflammatory conditions (further details in Supplementary M&M).

Generation of candidate gene and SNP rankings

The results of analyses and lookups listed above were then integrated to provide two rankings of: 1) all the SNPs within each GWAS locus interval and 2) all the genes within 3Mbp window centred around index SNP. This was achieved by assigning a score to each piece of evidence and summing across these sources to generate a causal prioritization score for every SNP and every gene tested. These scores represent the strength of evidence for a causal role of the SNP or gene in AD. Detailed method of calculation of basic score per gene or variant in a given experiment/analysis is presented in Supplementary Materials & Methods and visualized in Figure S1. Briefly, each source of evidence was assigned a weight based on subjective strength of evidence: highest (20) for results from statistical tests using full set of summary statistics, such as molecular QTL colocalization methods; lowest (1) for prediction results from machine learning models such as variant functional prediction software and intermediate (2) for positional overlap with significant experimental results, such as identified

promoter-enhancer loops. In calculating the final score, we also considered the magnitude of result significance or effect, specificity (overall number of SNPs/genes significant in a given experiment), independence of evidence (number of experiments conducted in the same study, such as measuring both expression and DNA methylation levels). The final score was adjusted by heterogeneity of evidence (i.e. genes or variants consistently supported by a range of evidence sources - alternative functional assays and statistical methods – were upweighted in proportion to the square root of mean number of unique study types and unique study IDs), as well as absolute number of studies providing supportive evidence.

DATA AVAILABILITY STATEMENT

All the code written to carry out the analysis is archived under:

https://doi.org/10.5281/zenodo.3775865

Datasets related to this article can be found in the following Figshare repositories:

Dataset S1. Full gene and variant loci rankings:

Gene: https://doi.org/10.6084/m9.figshare.12130824

Variant: https://doi.org/10.6084/m9.figshare.12130857

Dataset S2. Data evidence for top 10-ranked gene and variants at each locus:

Gene: https://doi.org/10.6084/m9.figshare.12130863

Variant: https://doi.org/10.6084/m9.figshare.12130878

Dataset S3. Locus Zoom-style gene and variant score plots for each locus:

Gene: https://doi.org/10.6084/m9.figshare.12221006

Variant: https://doi.org/10.6084/m9.figshare.12221033

Dataset S4. The final evidence dataset used in ranking genes and variants:

https://doi.org/10.6084/m9.figshare.12130701

Dataset S5. Reference table detailing individual datasets used.

https://doi.org/10.6084/m9.figshare.12222731

CONFLICT OF INTEREST STATEMENT

TRG receives funding from GlaxoSmithKline and Biogen for unrelated research.

ACKNOWLEDGEMENTS

This work was supported by a Springboard award (SBF003\1094), awarded to LP, supported by the Academy of Sciences (AMS), the Wellcome Trust, the Government Department of Business, Energy and Industrial Strategy (BEIS) and the British Heart Foundation (BHF). TRG acknowledges support from the UK Medical Research Council (MC_UU_00011/4). TGR is a UKRI Innovation research fellow (MR/S003886/1). TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. Fine-mapping analysis used the UK Biobank genetic data resource as an LD reference [Application #10074].

AUTHOR CONTRIBUTIONS

Conceptualization: MKS, TRG, LP; Data Curation: MKS; Funding Acquisition: LP; Investigation: MKS; Methodology: MKS,TRG, LP; Resources: VZ, JLM, LP; Supervision: TRG, LP; Validation: TGR; Writing - Original Draft Preparation: MKS, TRG, LP; Writing - Review and Editing: MKS, TGR, VZ, JLM, TRG, LP

REFERENCES

Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: Efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493–501

Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng H-F, et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. 2015;47(1):88–91

Cannon ME, Duan Q, Wu Y, Zeynalzadeh M, Xu Z, Kangas AJ, et al. Trans -ancestry Fine Mapping and Molecular Assays Identify Regulatory Variants at the ANGPTL8 HDL-C GWAS Locus. G3 genes - genomes - Genet. 2017;7(9):3217–27

Cannon ME, Mohlke KL. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. Am. J. Hum. Genet. ElsevierCompany.; 2018;103(5):637–53 Available from: https://doi.org/10.1016/j.ajhg.2018.10.001

Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front. Genet. 2020;11(May):1–21

Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell. 2016;167(5):1398-1414.e24

Chèneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: An updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res. 2018;46(D1):D267–75

Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sallari R, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of

gene expression to confer susceptibility to common traits. Genome Res. 2014;24(1):1–13

Elias MS, Wright SC, Remenyi J, Abbott JC, Bray SE, Cole C, et al. EMSY expression affects multiple components of the skin barrier with relevance to atopic dermatitis. J. Allergy Clin. Immunol. Elsevier Inc.; 2019;144(2):470–81 Available from: https://doi.org/10.1016/j.jaci.2019.05.024

Elmose C, Thomsen SF. Twin Studies of Atopic Dermatitis: Interpretations and Applications in the Filaggrin Era. J. Allergy. Hindawi; 2015;902359:1–7

Esaki H, Brunner PM, Renert-Yuval Y, Czarnowicki T, Huynh T, Tran G, et al. Early-onset pediatric atopic dermatitis is TH2 but also TH17 polarized in skin. J. Allergy Clin. Immunol. 2016;138(6):1639–51 Available from: http://www.sciencedirect.com/science/article/pii/S0091674916307953

Gamazon ER, Segrè A V, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. 2018;50(7):956–67 Available from: https://doi.org/10.1038/s41588-018-0154-4

Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 2021;49(D1):D1311–20 Available from: https://doi.org/10.1093/nar/gkaa840

Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014;10(5)

GTEx Consortium. Genetic effects on gene expression across human tissues. Nature.

2017;550(7675):204–13

Gulko B, Hubisz MJ, Gronau I, Siepel A. A method for calculating probabilities of fitness consequences for point mutations across the human genome. Nat. Genet. Nature Publishing Group; 2015;47(3):276–83 Available from: http://dx.doi.org/10.1038/ng.3196

Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. Nature Publishing Group; 2016;48(3):245–52 Available from: http://dx.doi.org/10.1038/ng.3506

Hay RJ, Johns NE, Williams HC, Bolliger IW, Dellavalle RP, Margolis DJ, et al. The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions. J. Invest. Dermatol. Elsevier Masson SAS; 2014;134(6):1527–34 Available from: http://dx.doi.org/10.1038/jid.2013.446

Heng TSP, Painter MW, Elpek K, Lukacs-Kornek V, Mauermann N, Turley SJ, et al. The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 2008;9(10):1091–4 Available from: http://www.nature.com/doifinder/10.1038/ni1008-1091

Hormozdiari F, Gazal S, Van De Geijn B, Finucane HK, Ju CJT, Loh PR, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet. 2018;50(7):1041–7

Huang D, Yi X, Zhang S, Zheng Z, Wang P, Xuan C, et al. GWAS4D: multidimensional analysis of context-specific regulatory variant for human complex diseases and traits. Nucleic Acids Res. Oxford University Press; 2018;46(W1):W114–20 Available from: https://academic.oup.com/nar/article/46/W1/W114/4996581

Jeng MY, Mumbach MR, Granja JM, Satpathy AT, Chang HY, Chang ALS. Enhancer

Connectome Nominates Target Genes of Inherited Risk Variants from Inflammatory Skin Disorders. J. Invest. Dermatol. The Authors; 2019;139(3):605–14 Available from: https://doi.org/10.1016/j.jid.2018.09.011

Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, et al. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Res. 2014;24(4):708–17

Kichaev G, Roytman M, Johnson R, Eskin E, Lindström S, Kraft P, et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. Bioinformatics. 2017;33(2):248–55

Kim-Hellmuth S, Aguet F, Oliva M, Muñoz-Aguirre M, Wucher V, Kasela S, et al. Cell type specific genetic regulation of gene expression across human tissues. bioRxiv. Cold Spring Harbor Laboratory; 2019; Available from: https://www.biorxiv.org/content/early/2019/10/17/806117

Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat. Commun. Springer US; 2017;8(1) Available from: http://dx.doi.org/10.1038/s41467-017-00366-1

Kuleshov M V, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7

Layer RM, Pedersen BS, Disera T, Marth GT, Gertz J, Quinlan AR. GIGGLE: a search engine for large-scale integrated genome analysis. Nat. Methods. 2018;15(2):123–6

de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput. Biol. 2015;11(4):1–19

Leung DYM, Guttman-Yassky E. Deciphering the complexities of atopic dermatitis: Shifting paradigms in treatment approaches. J. Allergy Clin. Immunol. 2014;134(4):769–79 Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84908095090&doi=10.1016%2Fj.jaci.2014.08.008&partnerID=40&md5=08507f4bf59402dce42ec8d2da04537c

Li MJ, Pan Z, Liu Z, Wu J, Wang P, Zhu Y, et al. Predicting regulatory variants with composite statistic. Bioinformatics. 2016;32(18):2729–36

Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. Nat. Genet. Springer US; 2019;50 Available from: http://www.nature.com/articles/s41588-019-0404-0

Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. 2018;50(11):1505–13 Available from: https://doi.org/10.1038/s41588-018-0241-6

Manz J, Rodrı E, Elsharawy A, Oesau E, Petersen B, Mayr G, et al. Targeted Resequencing and Functional Testing Identifies Low-Frequency Missense Variants in the Gene Encoding GARP as Significant Contributors to Atopic Dermatitis Risk. 2016;136

Milano A, McMahon A, Welter D, Bowler E, Hastings E, Cunningham F, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2016;45(D1):D896–901 Available from: https://dx.doi.org/10.1093/nar/gkw1133

Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. medRxiv. Cold Spring Harbor Laboratory Press; 2020; Available from:

https://www.medrxiv.org/content/early/2020/09/03/2020.09.01.20180406

Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat. Commun. 2018;9(1):1–18

Mora A, Sandve GK, Gabrielsen OS. In the loop : promoter – enhancer interactions and bioinformatics. 2016;17(July 2015):980–95

Mucha S, Baurecht H, Novak N, Rodríguez E, Bej S, Mayr G, et al. Protein-coding variants contribute to the risk of atopic dermatitis and skin-specific gene expression. J. Allergy Clin. Immunol. American Academy of Allergy, Asthma & Immunology; 2019; Available from: http://www.ncbi.nlm.nih.gov/pubmed/31707051

Newcombe PJ, Conti D V., Richardson S. JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. Genet. Epidemiol. 2016;40(3):188–201

Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat. Genet. 2015;47(12):1449–56

Richardson TG, Hemani G, Gaunt TR, Relton CL, Davey Smith G. A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. Nat. Commun. 2020;11(1):185

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–29

Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. Bioinformatics. 2018;34(3):511–3

Schlosser P, Li Y, Sekula P, Raffler J, Grundner-Culemann F, Pietzner M, et al. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. Nat. Genet. 2020;52:167–176 Available from: https://doi.org/10.1038/s41588-019-0567-8

Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. Nat. Genet. 2021; Available from: https://doi.org/10.1038/s41588-020-00776-w

Shooshtari P, Huang H, Cotsapas C. Integrative Genetic and Epigenetic Analysis Uncovers Regulatory Mechanisms of Autoimmune Disease. Am. J. Hum. Genet. ElsevierCompany.; 2017;101(1):75–86 Available from: http://dx.doi.org/10.1016/j.ajhg.2017.06.001

Simeonov DR, Gowen BG, Boontanrart M, Roth TL, Gagnon JD, Mumbach MR, et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature. Nature Publishing Group; 2017;549(7670):111–5 Available from: http://dx.doi.org/10.1038/nature23875

Slowikowski K, Hu X, Raychaudhuri S. SNPsea: An algorithm to identify cell types, tissues and pathways affected by risk loci. Bioinformatics. 2014;30(17):2496–7

Stevens ML, Zhang Z, Johansson E, Ray S, Jagpal A, Ruff BP, et al. Disease-associated KIF3A variants alter gene methylation and expression impacting skin barrier and atopic dermatitis risk. Nat. Commun. 2020;11(1):4092 Available from: https://doi.org/10.1038/s41467-020-17895-x

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. 2004;101(16):6062–7

Suárez-Fariñas M, Dhingra N, Gittler J, Shemer A, Cardinale I, de Guzman Strong C, et al. Intrinsic atopic dermatitis shows similar TH2 and higher TH17 immune activation compared with extrinsic atopic dermatitis. J. Allergy Clin. Immunol. 2013;132(2):361–70 Available from: http://www.sciencedirect.com/science/article/pii/S0091674913006933

Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. Springer US; 2018;558(7708):73–9 Available from: http://dx.doi.org/10.1038/s41586-018-0175-2

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. Oxford University Press; 2019;47(D1):607–13

Taşan M, Musso G, Hao T, Vidal M, Macrae CA, Roth FP. Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. Nat. Methods. 2015;12(2):154–9

Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, Simon LM, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. 2019;25(7):1153–63 Available from: https://doi.org/10.1038/s41591-019-0468-5

Võsa U, Claringbould A, Westra H, Bonder MJ, Zeng B, Kirsten H, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta- analysis. bioRxiv. 2018;1–57 Available from: https://www.biorxiv.org/content/10.1101/447367v1

Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. Springer US; 2019;51(4):592–9 Available from: http://www.nature.com/articles/s41588-019-0385-z

Wang J, Vicente-García C, Seruggia D, Moltó E, Fernandez-Miñán A, Neto A, et al. MIR retrotransposon sequences provide insulators to the human genome. Proc. Natl. Acad. Sci. 2015;112(32):E4428–37 Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1507253112

Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. Springer US; 2017;8(1):1–10 Available from: http://dx.doi.org/10.1038/s41467-017-01261-5

Wood AR, Hernandez DG, Nalls MA, Yaghootkar H, Gibbs JR, Harries LW, et al. Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. Hum. Mol. Genet. 2011;20(20):4082–92 Available from: https://doi.org/10.1093/hmg/ddr328

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science (80-.). American Association for the Advancement of Science; 2014;18(December) Available from: http://science.sciencemag.org/content/early/2014/12/17/science.1254806

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. Nature Publishing Group; 2012;44(7):821–4 Available from: http://dx.doi.org/10.1038/ng.2310

Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Res. 2012;41(Database issue):D188-94 Available from: http://nar.oxfordjournals.org/content/early/2012/11/26/nar.gks1165.abstract

Table 1 Genes prioritized at atopic dermatitis GWAS loci

The closest genes to the index variant (in either direction) are marked in **bold**. The two values given in parentheses in top 3 ranked gene columns correspond to the gene prioritization score and percentage of the total score for locus top 10 genes, respectively.

Locus	GWAS Index variant	Nearest genes	Top ranked gene	2nd ranked gene	3rd ranked gene
1q21.3 - a	rs61813875	CRCT1/LCE3E	HRNR (464, 28%)	RPTN (285, 17%)	CRNN (249, 15%)
1q21.3 - b	rs12730935	IL6R	**IL6R (743, 62%)**	UBE2Q1 (93, 8%)	ADAR (61, 5%)
2p13.3	rs112111458	CD207/VAX2	**CD207 (272, 45%)**	CLEC4F (62, 10%)	VAX2 (56, 9%)
2q12.1	rs6419573/rs3917265*	IL18R1/IL18RAP	**IL18R1 (1384, 39%)**	**IL18RAP (1341, 38%)**	IL1RL1 (224, 6%)
2q37.1	rs1057258	INPP5D	**INPP5D (296, 57%)**	ATG16L1 (106, 20%)	RN7SL32P (29, 6%)
4q27	rs6827756/rs13152362*	KIAA1109	**KIAA1109 (220, 35%)**	BBS12 (112, 18%)	TRPC3 (100, 16%)
5p13.2	rs10214237	IL7R/CAPSL	**IL7R (965, 65%)**	SPEF2 (203, 14%)	UGT3A2 (89, 6%)
5q31.1 - a	rs12188917	TH2LCRR	SLC22A5 (461, 35%)	IRF1 (303, 23%)	RAD50 (122, 9%)
5q31.1 - b	rs4705962*	KIF3A	**KIF3A (249, 23%)**	SLC22A5 (247, 23%)	PDLIM4 (142, 13%)
6p21.32	rs4713555	STAT3	HLA-DRA (1405, 30%)	HLA-DQB1 (689, 15%)	**HLA-DRB1 (566, 12%)**
6p21.33	rs41293864	MICB	HSPA1B (173, 15%)	HCG27 (165, 14%)	CSNK2B (152, 13%)
8q21.13	rs6473227	MIR5708/ZBTB10	**ZBTB10 (192, 41%)**	TPD52 (70, 15%)	PAG1 (69, 15%)
10p15.1	rs6602364	IL2RA/IL15RA	**IL2RA (333, 45%)**	RBM17 (111, 15%)	PFKFB3 (51, 7%)
10q21.2	rs2944542	ZNF365	ADO (615, 61%)	**ZNF365 (101, 10%)**	EGR2 (90, 9%)
11p13	rs2592555/rs12295535*	PRR5L	**PRR5L (598, 79%)**	TRAF6 (65, 9%)	COMMD9 (34, 5%)
11q13.1	rs10791824	OVOL1	CTSW (336, 23%)	**OVOL1 (236, 16%)**	EFEMP2 (168, 11%)
11q13.5	rs2212434	C11orf30/LRRC32	**LRRC32 (545, 43%)**	**EMSY (521, 41%)**	THAP12 (47, 4%)
11q24.3	rs7127307	–/ETS1	**ETS1 (298, 75%)**	FLII (35, 9%)	APLP2 (18, 5%)
12q15	rs2227483	IL22	MDM1 (728, 70%)	**IL22 (99, 10%)**	IFNG (57, 5%)
14q13.2	rs2038255	PPP2R3C	**PPP2R3C (996, 31%)**	KIAA0391 (814, 25%)	SRP54 (433, 13%)
14q32.32	rs7146581	TRAF3	**TRAF3 (848, 55%)**	AMN (281, 18%)	CDC42BPB (186, 12%)
16p13.13	rs2041733	CLEC16A	DEXI (376, 34%)	**CLEC16A (364, 33%)**	RMI2 (108, 10%)
17q21.2	rs12951971	STAT3	DHX58 (254, 32%)	**STAT3 (101, 13%)**	RAB5C (100, 13%)
17q25.3	rs11657987	PGS1	**PGS1 (205, 46%)**	DNAH17 (73, 16%)	SOCS3 (52, 12%)
19p13.2	rs2918307	ADAMTS10/ACTL9	**ACTL9 (115, 41%)**	**ADAMTS10 (57, 20%)**	MAP2K7 (34, 12%)
20q13.33	rs4809219	RTEL1/TNFRSF6B	STMN3 (608, 27%)	LIME1 (473, 21%)	ARFRP1 (257, 12%)

*index SNP for secondary signal, where the pipeline did not give different gene prioritizations for the two signals, these are presented on one row.

FIGURE LEGENDS

Figure 1. Gene scores within the 3Mbp interval of lead SNP in the 6 highest-scoring loci. Top prioritised gene marked with a black square. a) locus 1q21.3 – b; b) locus 2q12.1; c) locus 5p13.2; d) locus 11p13; e) locus 14q13.2; f) locus 14q32.32.

Figure 2. Score by type of evidence for top 3 ranked genes in the 6 highest-scoring loci. Scores for top 3 ranked genes at each locus are shown partitioned by category of evidence – here including the top 10 categories contributing the highest proportion of total score at the top 10 ranked genes for all loci. Order of loci corresponds to the order in Table 1.

Figure 3. Network visualization of the functional terms enriched among locus top 3 prioritized genes. The ontology categories are depicted as blue hexagons, with their size linearly proportional to $-\log_{10}$ of adjusted enrichment *p*-value. AD genes are depicted as pink rectangles, with the intensity of the colour fill proportional to gene score and thickness of the green border marking the gene rank at the locus, with rank 1 the thickest.