Integrative multiomics analysis highlights immune-cell regulatory mechanisms and shared genetic architecture for 14 immune-associated diseases and cancer outcomes

Claire Prince,[1] Ruth E Mitchell,[1] Tom G. Richardson*[1,2]

[1] Medical Research Council Integrative Epidemiology Unit (MRC IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom
Claire.prince@bristol.ac.uk; R.Mitchell@bristol.ac.uk; Tom.G.Richardson@bristol.ac.uk.
[2] Novo Nordisk Research Centre, Headington, Oxford, OX3 7FZ, United Kingdom

*Corresponding author: Tom G. Richardson, MRC Integrative Epidemiology Unit, Bristol Medical School (Population Health Sciences), University of Bristol, Tom.G.Richardson@bristol.ac.uk

Abstract

Developing functional insight into the causal molecular drivers of immunological disease is a critical challenge in genomic medicine. Here we systematically apply Mendelian randomization (MR), genetic colocalization, immune cell-type enrichment and phenome-wide association methods to investigate the effects of genetically predicted gene expression on 10 immune-associated diseases and 4 cancer outcomes. Using whole blood derived estimates for regulatory variants from the eQTLGen consortium (n=31,684) we constructed genetic risk scores for 10,104 genes. Applying the inverse-variance weighted MR method transcriptome-wide whilst accounting for linkage disequilibrium structure identified 664 unique genes with evidence of a genetically predicted effect on at least one disease outcome ($P<4.81 \times 10^{-5}$). We next undertook genetic colocalization to investigate cell-type specific effects at these loci using gene expression data derived from 18 types of immune cells. This highlighted many cell-type dependent effects, such as *PRKCQ* expression and asthma risk (posterior probability=0.998), which was T-cell specific. Phenome-wide analyses on 311 complex traits and endpoints allowed us to explore shared genetic architecture and prioritize key drivers of disease risk, such as *CASP10* which provided evidence of an effect on 7 cancer-related outcomes. Our atlas of results can be used to characterize known and novel loci in immune-associated disease and cancer susceptibility, both in terms elucidating cell-type dependent effects as well as dissecting shared disease pathways and pervasive pleiotropy. As an exemplar, we have highlighted several key findings in this study, although similar evaluations can be conducted using our interactive web platform.

Introduction

The widespread application of genome-wide association studies (GWAS) has had profound success in detecting robust associations between genetic variants and complex disease outcomes. This includes diseases with a large immunological basis, such as rheumatoid arthritis [MIM 180300], inflammatory bowel disease [MIM 266600] and asthma [MIM 600807]. [1] The immune system also plays a crucial role in the pathogenesis of other types of disease, such as cancer outcomes. [2] There is now extensive interest in the field of genetic epidemiology in integrating findings from GWAS with regulatory molecular datasets. [3-5] In doing so, studies aim to bring to light the underlying functional and biological mechanisms responsible for GWAS signals and translate findings for disease prevention purposes.

A challenge encountered by these endeavours is obtaining molecular trait datasets derived from tissues and cell-types relevant to the disease being studied in sufficient samples. A recent review highlights this by comparing the differences between affected and unaffected tissues for heritable traits and diseases, [6] demonstrating that molecular traits such as gene expression can have elevated or even exclusive expression in disease-relevant tissue types. For the majority of disease outcomes, this diminishes the utility of whole blood-derived datasets, which to date typically have by far the largest sample sizes on molecular traits due to their non-invasive accessibility. Notable exceptions to this are diseases with a large immune basis, given that whole blood is responsible for carrying innate and adaptive immune cells through the body from the lymphatic system to the site of injury or infection. [7] As such, initial analyses of transcriptomic datasets derived from whole blood provides optimal statistical power to detect association signals for immune-associated diseases [8; 9], which can then be dissected and characterized in detail using cell-type specific data. Doing so can help develop mechanistic insight into the cell-types which play a role on the causal pathway from genetic variants to immune-associated diseases and cancer outcomes. [10]

Furthermore, diseases with an immune component and types of cancer are known to have shared genetic architecture. [11] For example, previous work in this area has identified evidence

of shared genetic architecture between Crohn's disease [MIM 266600] and multiple sclerosis [MIM 126200], [12] across several paediatric autoimmune diseases, [13] and also amongst various cancer endpoints due to immune-associated mechanisms. [14] Moreover, genetic evidence of horizontal pleiotropy at loci which encode a therapeutic target may be informative in terms of flagging potential adverse effects unrelated to immune-associated and cancer outcomes, which is particularly attractive given the increasing interest in using human genetics to help validate drug targets. [15] Taken together, these findings highlight the importance of conducting evaluations of pervasive pleiotropy at immune disease and cancer susceptibility loci, to prioritise candidate genes where in-depth follow up analyses would be worthwhile.

In this study, we constructed an analytical pipeline to systematically triangulate evidence from Mendelian randomization and genetic colocalization methods to evaluate the effects of genetically predicted gene expression for 10,104 genes on 14 immune-associated disease and cancer outcomes. MR is a form instrumental variable analysis which uses genetic variants to infer causal relationships between exposures and disease outcomes, which are more robust to confounding and reverse causation given that they are inherited at birth. [16; 17] Although MR studies of molecular traits have typically been limited to single instrument analyses in the past, regulatory variants derived from whole blood by the eQTLGen consortium (n=31,684) provide an opportunity to harness multiple variants in an MR framework whilst accounting for their linkage disequilibrium (LD) structure. [18] We then explored whether putative genetic effects may be cell-type specific using expression data from 18 different immune-cell regulatory datasets from the BLUEPRINT consortium and DICE database using genetic colocalization. [19; 20] Finally, we undertook a phenome-wide association study (PheWAS) of genes highlighted by these analyses to assess their shared architecture and pathways using data on a total of 311 curated complex traits and outcomes. A schematic diagram of this analysis pipeline can be found in **Figure 1**.

Material and Methods

The eQTLGen consortium

We obtained expression quantitative trait loci (eQTL) data derived from 31,684 blood and peripheral blood mononuclear cell samples using data from 37 studies as part of the eQTLGen consortium. Detailed methods have been described previously [21]. In this study, we only selected eQTL for analyses that were located within 100 kilobases (kbs) either side of a protein-coding gene region whose expression they were robustly associated with (based on $P<5x10^{-8}$). This was to mitigate the likelihood that instruments for our MR analyses were influencing disease outcomes via alternate biological pathways due to the co-expression of neighbouring genes.

We constructed genetic scores using weakly independent cis-regulatory variants based on an $r^2<0.1$. LD calculations were based on a reference panel of 10,000 unrelated UK Biobank participants of European descent. [22-24] Genetic scores were only analysed where we had at least 2 weakly independent eQTL to reduce the likelihood of false positive effects due to only being able to instrument genes using a single variant.

Immune-cell type specific datasets

We obtained immune-cell specific eQTL data from the BLUEPRINT and DICE projects, [19; 20] eQTL from 3 cell types were available from BLUEPRINT (monocytes, neutrophils, and t-cells) and a further 15 from DICE (a full list of these cell types can be found in **Table S1**). The sample sizes for these immune-cell type specific datasets ranged from n=89 to n=194, meaning that using these data for MR would typically be restricted to single SNP analyses. As such, we used these datasets specifically for follow up analyses of genetically predicted effects identified in whole blood to evaluate cell-type specificity.

Complex trait and disease outcome datasets

Our primary analyses were based on 14 disease outcomes consisting of 10 immune-associated diseases and 4 types of cancer. These were asthma, breast cancer [MIM 114480], Crohn's

disease, eczema [MIM 603165], hypothyroidism [608175], inflammatory bowel disease, multiple sclerosis, ovarian cancer [MIM 167000], prostate cancer [MIM 176807], rheumatoid arthritis, systemic lupus erythematosus [MIM 152700], total number of cancers, type 1 diabetes [MIM 222100], and ulcerative colitis [MIM 266600]. [25-33] Full details on the GWAS for each of these outcomes can be found in **Tables S2 & S3.** These disease outcomes were selected due to their studies being undertaken on large sample sizes (i.e., n > 10,000) and also those providing access to the full summary statistics which were necessary to conduct MR. Although this meant that we were unable to analyse other endpoints where our systematic pipeline would be of value (e.g., leukaemia), large-scale GWAS data of these outcomes may become accessible for future work.

All other GWAS summary statistics analysed in our PheWAS were accessed through the IEU Open GWAS project. [34] Estimates were extracted using the *'TwoSampleMR'* R package. [35] In total, 297 traits and outcomes were included which related to broad range of outcomes from across the complex disease spectrum (**Table S4**). These endpoints were based on a curated list of outcomes analysed previously using an MR framework. [5] The 14 endpoints from our primary analysis were also included in these analyses for comparative purposes meaning overall 311 outcomes were analysed.

Statistical analysis

We used the *'MendelianRandomization'* R package to undertake two-sample MR analyses using the inverse-variance weighted (IVW) method accounting for correlation structure between instruments. [23; 36] We firstly applied this approach transcriptome-wide on each of the 14 immune and cancer-related outcomes in turn to highlight genes whose expression provided strong evidence of a genetically predicted effect. As a heuristic, we used a Bonferroni corrected threshold based on the number of genes analysed across the transcriptome for each outcome (i.e., P<0.05/number of genes analysed). For results that survived Bonferroni corrections, we performed a leave-one-out MR analysis. This involved repeating analyses after removing each SNP in turn with replacement and allowed us to determine whether any individual SNPs were

driving genetically predicted effects. Results for genes where the largest leave-one-out p value still survived the heuristic Bonferroni corrected threshold used in the initial analysis were considered to be the most robust to individual SNPs driving genetically predicted effects.

Next, we conducted genetic colocalization analyses using the *'coloc'* R package with default parameters. This allowed us to investigate the cell-type specificity of putative effects at each locus robust to Bonferroni corrections in the previous analysis. We also only carried forward genes located outside the human leukocyte antigen (HLA) region of the genome due to the extensive LD structure at HLA which can results in false positive findings when using techniques such as genetic colocalization. In our colocalization analysis, we evaluated whether there was a causal variant at each locus responsible for conferring risk of disease that was also driving variation in gene expression derived from each of the 18 cell-type specific datasets in turn. A posterior probability (PPA) threshold ≥ 0.80 was used to indicate evidence of a shared a common causal variant between disease outcome and cell-type specific gene expression.

We applied MR only to genes with at least 2 instruments to mitigate false positive results which single SNP MR of molecular traits may be particularly prone to due to co-expression between neighbouring genes. Furthermore, this allowed us to harness the power of the large-scale whole blood dataset from eQTLGen (n=31,684) which provided an unprecedented number of eQTL for MR analyses. Genetic colocalization was then selected as a method to investigate the cell-type specificity of findings given that the immune cell datasets analysed were derived based on comparatively modest sample sizes (n=89 to 194). As such, only a very small proportion of genes would have had at least 2 instruments for MR analyses based on our criteria of $P<5x10^{-8}$ and $r^2<0.1$ (mean=~141 genes across the 18 immune-cell types, **Table S5**). Conversely, a much larger proportion of genes are likely to satisfy the single variant assumption of the coloc method using these datasets, particularly based on default Bayesian prior distributions. Moreover, this overall approach allowed us to adopt the principles of triangulation, whereby separate approaches (with different underlying assumptions) are applied to different datasets to investigate multiple lines of evidence. [37] We were therefore

able to corroborate findings from MR and colocalization in this study, which have different strengths and weaknesses, in order to provide evidence implicating a gene's role in disease risk.

Using the results from our genetic colocalization analysis, we applied a hypergeometric test to assess evidence of enrichment for cell-type specific effects across the genome for each disease outcome. The *'phyper'* R package was used to perform enrichment analysis. [38] Background comparisons were based on the other loci identified by our MR analyses which did not provide evidence of colocalization with gene expression from the same cell-type. For immune-associated disease outcomes we used other immune-associated disease loci for comparisons, and likewise used cancer loci for cancer outcome enrichment evaluations. [39] We considered a Bonferroni corrected threshold as robust evidence of enrichment in these analyses based on the 18 immune-cell types assessed for each outcome (i.e. $P<0.05/18 = 2.78x10^{-3}$).

Lastly, we performed a PheWAS using the IVW MR method accounting for local LD structure as before using eQTLGen data. However, this analysis was restricted to genes that survived Bonferroni corrections in the initial MR analysis based on the 14 immune- and cancer-related outcomes. We used a Bonferroni-corrected threshold based on the number of outcomes analysed in the PheWAS (i.e., $P<0.05/311 = 1.61x10^{-4}$) to identify evidence of genetically predicted effects on an outcome. However, as with the other thresholds applied in this study, this cut-off was used as a heuristic to highlight noteworthy findings and users of our web atlas may wish to apply more stringent or lenient thresholds as they see fit.

Comparison with previous findings from transcriptome-wide association studies

We assessed the reliability of our initial MR analysis by comparing results reported by transcriptome-wide association studies (TWAS) of whole blood eQTL datasets from TWAS-Hub . [40] These were based on the Netherlands Twins Register (NTR) (n=1247) and, [41] GTEx whole blood (n=338) datasets. [42] Only 10 of the 14 primary outcomes were compared in this analysis as inflammatory bowel disease, multiple sclerosis, number of self-reported cancers and type 1 diabetes were not available in TWAS-Hub. We only compared genes which were robust to

Bonferroni corrections (i.e., P<0.05/number of genes analysed) from both our analyses and those from TWAS-Hub.

All analyses were undertaken using R version 3.6.1 and 3.6.2. Manhattan and PheWAS plots were generated using the R package *'ggplot2'*, [43] locus zoom plots using the code adapted from the *'gassocplot'* package and enrichment plots using the *'pheatmap'* package. We also developed a web application to disseminate findings for all results generated using our analytical pipeline using the *'shiny'* R package.

Results

An integrative Mendelian randomization analysis of 14 immune-associated diseases and cancer outcomes using multiple cis-regulatory instruments

Constructed genetic scores using weakly independent cis-regulatory variants identified 10,104 genes that were eligible for analysis using the IVW MR method. (**Table S6**) In total, 734 genes provided evidence of a genetically predicted effect on at least 1 of the 14 immune- and cancer-related outcomes after accounting for multiple testing using the Bonferroni corrected threshold for each outcome separately (ranging from $P=4.09x10^{-6}$ to $4.81x10^{-5}$) **(Tables S7 & S8**). Of these, 664 were located outside the HLA region of the genome and were carried forward for subsequent analyses. **Figure 2** illustrates various exemplar signals identified for 4 outcomes: asthma, hypothyroidism, breast cancer and inflammatory bowel disease. Full results for all 14 outcomes can be investigated using the interactive web browser (see "Web Resources" for URL).

Amongst these results were genetically predicted effects at various well-established loci known to confer risk of autoimmune disease, including *CARD9* [MIM 607212] and *STAT3* [MIM 102582] ($P=1.03x10^{-15}$ and $P=1.77x10^{-8}$ respectively with inflammatory bowel disease), [44] *ORMDL3* [MIM 610075] associated with asthma ($P=4.82x10^{-10}$), [45] and cytokines such as interleukin-24 (*IL24* [MIM 604136]*)* and interleukin-2 receptor alpha chain *(IL2RA* [MIM 147730]) which were associated with systemic lupus erythematosus and asthma respectively ($P=1.28x10^{-6}$ and $P=1.34x10^{-6}$). [46; 47] A number of novel or emerging loci were also identified for autoimmune disease outcomes, such as *RORC* [MIM 602943], a transcription factor predominantly expressed in T helper 17 cells [48] which was most strongly associated with asthma risk ($P=4.13x10^{-27}$), as well as *CCDC88B* [MIM 611205] ($P=1.07x10^{-6}$ and $P=1.15x10^{-5}$ with hypothyroidism/myxoedema and multiple sclerosis respectively). These findings were additionally supported by evidence from our leave-one-out sensitivity analysis to highlight signals which were not dependent on single cis-instruments (**Table S9**).

Similarly, there were various findings highlighted by this approach at known cancer loci, such as *CASP10* [MIM 601762] ($P=1.82x10^{-17}$ for prostate cancer)*,* [49] and *FAM175A* [MIM 611143] ($P=1.56x10^{-16}$ for breast cancer)*,* [50] as well as genes that have been identified in relation to a number of cancers including *CDKN2A* [MIM 600160] ($P= 5.08 x10^{-14}$ with breast cancer)*, IRF1* [MIM 147575] ($P=7.40x10^{-7}$ with breast cancer) and *IGF2* [MIM 147470] ($P=1.67x10^{-6}$ with prostate cancer). [51] There were also loci highlighted by our analyses on cancer outcomes with limited previous evidence of an association with cancer outcomes based on the current literature and may therefore be worthwhile prioritising for further evaluation, such as *PSMD8* [MIM 617844] ($P=5.60x10^{-9}$ with prostate cancer) and *TTC16* ($P=1.10x10^{-9}$ with prostate cancer).

Comparing our results with those reported by TWAS-Hub found that a large proportion of genes robust to Bonferroni correction using whole blood eQTL datasets from this resource were replicated by our analyses (**Table S10)** (NTR: 55 out of 157, GTEx: 42 out of 111)**.** However, our analyses also highlighted a large number of genes highlighted robust to multiple testing corrections in comparison to TWAS-Hub, which is likely predominantly due to using an eQTL dataset derived from a substantially larger sample. For instance, for Crohn's disease, while we replicated 22 genes identified in TWAS-Hub, there were 40 genes with evidence of a genetically predicted effect in our analysis which did not meet Bonferroni corrections (NTR: $P< 2.07x10^{-5}$, GTEx: $P<2.49x10^{-5}$) in TWAS-Hub.

Identifying immune-cell specific effects at immune-associated and cancer associated loci

We performed genetic colocalization at each of the 664 non-HLA loci identified in the previous analysis using 14 immune-cell datasets from the DICE database and 3 immune cell-type datasets from the BLUEPRINT consortium. In total, 531 genetic effects colocalised across the immune-associated disease and cancer outcomes with immune-cell type expression (based on PPA≥0.80), which may provide mechanistic insight into the disease pathogenesis at these loci **Table S11**.

For example, we identified strong evidence of colocalization between *PRKCQ* [MIM 600448] expression in T cells and asthma risk (PPA=0.998), whereas there was very weak evidence of

colocalization when analysing any of the other immune cell-types (**Figure 3a, Table S12**). *PRKCQ* has been previously implicated in allergic disease risk and is involved in T cell activation. [52] There was also evidence of colocalization between *KSR1* [MIM 601132] expression and Crohn's disease in classical monocytes (PPA=0.998) (**Figure 3b, Table S12**), which is known to be an important cell-type in relation to Crohn's disease. [53] Amongst cancer loci, there was evidence for colocalization between prostate cancer and *C2orf43* [MIM 613570] in non-classic monocytes (PPA=0.887)) (**Figure 3c, Table S12**). *C2orf43* has been found to be expressed in monocytes and the loss of this gene has previously been associated with risk of prostate cancer. [54; 55] All other effects with evidence of genetic colocalization are shown in **Table S11** as well as on our web browser where effects across all cell-type can be compared visually.

Enrichment of immune-cell types amongst disease-associated loci

We performed enrichment analyses using results from the colocalization analyses to investigate whether effects in certain immune-cell types were overrepresented amongst each outcome. We did not include the number of reported cancers results in the background set as there was no strong evidence of colocalization identified in the previous analysis for any gene using this outcome. This may reflect that associated loci are more likely to be involved in risk factors for cancer rather than being directly involved in cancer pathogenesis themselves.

As illustrated in **Figure 4**, we identified evidence of enrichment for various cell-types amongst rheumatoid arthritis loci and in particular for activated naïve CD8 T cells ($P= 1.79 \times 10^{-4}$). Increased levels of these cells have been previously observed in the peripheral blood of individuals with rheumatoid arthritis. [56] Monocytes were enriched amongst multiple sclerosis loci ($P= 6.34 \times 10^{-4}$) which have previously been implicated in the pathology of this disease. [57] The strongest evidence of enrichment for breast cancer loci was for regulatory memory T cells (P= 0.018); which have previously been reported to restrict anti-tumour immune mechanisms, although this finding was not robust to multiple testing corrections. [58] (**Tables S13 & S14**)

Conducting phenome-wide association studies to explore shared genetic architecture and elucidate pleiotropic loci

For the 664 non-HLA genes identified in our primary analysis using whole blood, we repeated analyses using the IVW MR analysis accounting for LD structure but on a set of 311 curated traits and outcomes. This phenome-wide analysis allowed us to highlight loci where there is evidence of shared genetic architecture amongst various immune-associated and cancer outcomes. For instance, *IL24*, which encodes an interleukin cytokine involved in promoting the development and differentiation of T, B, and hematopoietic cells, and plays an essential role in both innate and adaptive immunity, [59] provided evidence of an effect on multiple autoimmune outcomes (**Figure 5a**). Similarly, analyses of *CASP10*, which encodes caspase 10 and is a known cancer susceptibility locus, identified genetically predicted effects on 7 different cancer disease outcomes. [49] (**Figure 5b**) There was also evidence of shared architecture at emerging immune disease loci, such as *CCDC88B*, which has recently been implicated in the pathogenesis of inflammatory bowel disease. [60]

Along with evaluations of loci with shared architecture for immune-associated and cancer outcomes, our atlas of phenome-wide results may be valuable in highlighting genes with more specific effects on disease outcomes. For example, *PRKCQ* was highlighted by our cell-type analysis as having a T-cell specific mediated effect on asthma risk (PPA=0.998), and only provided robust evidence of an effect on asthma as discovered in our initial analysis based on the number of tests undertaken ($P<1.61 \times 10^{-4}$=0.05/311 tests) (**Figure 5c**). Similar evaluations of pleiotropy may have translatable benefit for drug target prioritization efforts. For instance, *TPM3* [MIM 191030] has recently been postulated as a potential therapeutic target for cancer therapy. [61] Although our cell-type analysis detected evidence of a monocyte-specific role of *TPM3* in prostate cancer risk (PPA=0.821), phenome-wide results indicated that it may influence risk of outcomes such as hypertension [MIM 145500] ($P=1.49 \times 10^{-7}$) and angina ($P=2.81 \times 10^{-9}$) with the opposite direction of effect. These results therefore suggest that loci which exhibit horizontal pleiotropic effects such as *TPM3* should be deprioritised as therapeutic

targets due to putative adverse effects. Results depicted in **Figure 5** can also be found in **Tables S15-S18**.

Discussion

In this study, we have performed a transcriptome-wide Mendelian randomization study to investigate the genetically predicted effects of gene expression on risk of 10 immune-associated diseases and 4 cancer outcomes. The results of this investigation provide a comprehensive atlas of genetic effects which highlight both known and novel susceptibility loci for these outcomes. We conducted in-depth analyses of these loci using genetic colocalization and phenome-wide MR to further characterize their role in disease, both in terms of developing mechanistic insight into cell-type dependent effects as well as elucidating shared biological pathways. As exemplar, we have highlighted several key findings in this manuscript, however all our results can be investigated interactively (see "Web Resources" for URL). We envisage this atlas of results will benefit future research endeavours interested in dissecting the molecular drivers of immune-associated disease and cancer outcomes, as well as help guide functional studies to validate and strengthen evidence for loci highlighted in our study.

Integrating molecular regulatory signatures derived from whole blood with findings from GWAS has been considered a limitation for the majority of complex disease outcomes studied to date. [62] However, it presents a viable strategy for immune system-related diseases given that whole blood is responsible for carrying innate and adaptive immune cells through the body from the lymphatic system to the site of injury or infection [7]. This has allowed us to harness the unparalleled sample size of transcriptome-wide data made available by the eQTLGen consortium. As a consequence, we were able to instrument genes using multiple regulatory variants and address another conventional limitation of previous studies in the paradigm which have typically been confined to single-SNP MR analyses. [5]

Amongst these findings are many previously reported autoimmune disease and cancer genes. For example, *CARD9,* identified as having a genetically predicted effect on Crohn's disease and

ulcerative colitis, has previously been reported to confer risk of both these forms of inflammatory bowel disease. [63] Additionally, it is known to be involved in innate immunity and inflammation, as well as being specifically expressed in myeloid cells. [64] Likewise, *STAT3* has been identified in relation to inflammatory bowel diseases and type 1 diabetes and is thought to be involved in autoimmunity both due to its role as a mediator on the IL-6 signalling pathway and as a transcription factor in the differentiation of Th1 cells. [65; 66] Amongst established cancer loci was *CTBP1* [MIM 602618] which we identified evidence as having a genetically predicted effect on breast and prostate cancer risk. *CTBP1* is an oncogenic transcriptional co-regulator which has been shown to be overexpressed in a number of cancers. It functions by regulating the expression of tumour suppressers and oncogenic factors, which has led to its identification as a potential therapeutic target. [67]

There were also less well-established loci identified in the MR analysis. *RORC* was highlighted in relation to inflammatory bowel diseases, eczema, and asthma. It is a transcription factor of *IL-17* [MIM 603149] expression and Th17 cells, which are key in the immune system and has been suggested as a potential target for autoimmune diseases. [68] *IL-17* is a pro inflammatory cytokine which recruits immune cells to the site of inflammation and its overproduction has been reported to lead to inflammation and autoimmune conditions. [69; 70] *CCDC88B* provided evidence of a genetically predicted effect on hypothyroidism/myxoedema and multiple sclerosis, which is a gene previously shown to be highly expression in immune cells. Furthermore, it has been identified as an important regulator of T cell function and previously implicated to play a role in inflammation pathways. [71]

Applying genetic colocalization revealed many cell-type dependent effects in our study, which may help develop understanding into the pathways and mechanisms behind these disease outcomes. [72] For example, there was strong evidence of colocalization between *CARD9* expression and inflammatory bowel diseases in monocytes (PPA=0.986) and neutrophils (PPA=0.986). The effect of *KSR1* expression on risk of Crohn's disease colocalized with data from monocytes (PPA=0.998). *KSR1,* which encodes a kinase suppressor of Ras 1, has previously

been identified in relation to Crohn's disease, [44] and blood monocytes have been previously reported to be elevated in individuals with this disease. [53]

PRKCQ, which provided evidence of a genetically predicted effect on asthma risk in our MR analysis, is a member for the protein kinase C family and encodes the enzyme protein kinase C theta which has an important role in the regulation of signalling pathways and the activation of T cells. Moreover, *PRKCQ* has been identified as having a crucial role in autoimmunity through T cell activation. [73; 74] Findings in this study provided evidence of genetic colocalization for this gene with asthma in T cells but none of the other immune cell types assessed, suggesting that *PRKCQ*'s role in conferring autoimmune disease risk may be confined to T cells. [52] *C2orf43* was identified in relation to prostate cancer and provided evidence of colocalization for prostate cancer in monocytes and T cells. It has previously been shown that loss of *C2orf43* may be associated with risk of prostate cancer, and is a gene found to be expressed in lysates of human monocytes and monocyte-derived macrophages. [55]

We found that the 531 effects which provided evidence of genetic colocalization using cell-type specific gene expression were enriched for certain disease outcomes. For instance, rheumatoid arthritis loci were enriched for evidence of genetic colocalization with gene expression derived from activated naïve CD8 T cells. Increased levels of these cells have been previously observed in the peripheral blood of individuals with rheumatoid arthritis. [56] It has also been suggested that $CD8^+$ T cells have a role in the initiation and maintenance of rheumatoid arthritis. [75] Colocalization evidence with gene expression from monocytes was enriched amongst multiple sclerosis loci, which have previously been implicated in the pathology of this disease by increasing levels of cytokines leading to increased cellular activation and proliferation, tissue damage and altered blood brain barrier. [57]

Our PheWAS analysis highlighted genes which are involved in conferring risk of multiple immune-associated disease outcomes, such as well-established autoimmune locus *IL24*. This gene is in the interleukin family of cytokines which are involved in signalling in the immune

system and regulating immune cells,[46] *IL24* has been identified as a key mediator for both pro-inflammatory diseases and allergic disorders. [76] Similarly, *CASP10* was identified in relation to various cancer outcomes in this analysis supporting previous findings from the literature. [49] *CASP10* encodes the enzyme caspase-10 which is a member of the caspase family which have a role in cell apoptosis. [77] This family of protease enzymes have been previously considered as potential therapeutic targets for cancer. [78]

There were also genes that provided evidence of genetically predicted effects on very few outcomes across the disease spectrum. For example, *PRKCQ* provided robust evidence of an effect on asthma, but no other outcomes assessed based on multiple testing corrections, and as previously mentioned this gene has been shown as important in asthma pathology. We also note that our PheWAS results may help elucidate pleiotropic loci which should be valuable for therapeutic validation endeavours. As an example, we demonstrate that previously postulated target *TPM3* for cancer therapy had genetically predicted effects on various disease endpoints, some of which had the opposite direction of effect to lower cancer risk. Evidence of horizontal pleiotropy may be useful in terms of deprioritising drug targets, whereas those which appear to be more specifically associated with disease, such as *PRKCQ*, may be worthwhile prioritising and pursuing further. However, results from our genetic analysis are but one line of evidence to be used in conjunction with findings from other studies such as functional wet lab work.

Although there various strengths to our study there are also limitations. Firstly, whilst the use of immune cell datasets and genetic colocalization allowed us to identify cell-type dependent effects in this study, the sample sizes used to derive these eQTL are modest in scale compared to GWAS and thus may not explain a large proportion of heritability. [79] Therefore, genetic effects which are not supported with colocalization evidence could be explained by low statistical power. Future datasets generated at scale once technologies become more feasible should facilitate more comprehensive evaluations of cell-type specific regulatory mechanisms. Moreover, larger sample sizes for immune cell specific datasets would allow a more robust investigation into whether results identified using whole blood are subject to molecular

pleiotropy (i.e. co-expression amongst neighbouring genes which can make pinpointing the causal gene at a locus challenging). Another limitation is horizontal pleiotropy, which may play a role in these results despite the support of colocalization and can be defined here as a causal variant influencing immune-cell expression and disease risk via two separate biological pathways. Finally, gene expression was not derived from disease related datasets and were mostly from "healthy" individuals. As such future work using genetic effects on gene expression derived from individuals diagnosed with immune-associated disease or cancer may potentially capture signatures not detected by our analyses. Lastly, our study was focused on cell-type dependent effects using immune-cell datasets, although investigating our results in conjunction with those from investigations into tissue specificity may facilitate further mechanistic insight. [5; 72]

The results of this study provide a map of genetically predicted regulatory mechanisms that may influence disease outcomes with an immune basis. These findings should prove valuable for future studies to further characterize susceptibility loci and translate genetic evidence for disease prevention and treatment purposes.

Declaration of Interests

TGR has become a part-time employee of Novo Nordisk whilst this manuscript was under review. All other authors declare no conflicts of interest.

Acknowledgements

We are extremely grateful to the eQTLGen, BLUEPRINT, DICE and GWAS consortia (Bentham et al. 2015, de Lange et al. 2017, Beecham et al. 2013, Phelan et al. 2017, Michailidou et al. 2017, Okada et al. 2014, Schumacher et al. 2018, Censin et al. 2017 and the MRC IEU) who kindly made their data available for the benefit on this study.

CP is supported by a Wellcome Trust PhD studentship in Molecular, Genetic and Lifecourse Epidemiology (108902/B/15/Z). TGR was a UKRI Innovation Research Fellow whilst supervising this project (MR/S003886/1). This work was supported by the MRC Integrative Epidemiology Unit which receives funding from the UK Medical Research Council and the University of Bristol (MC_UU_00011/1). The reference panel used in this study was created using data from the UK Biobank (app 15825).

Web Resources

All results generated in this study can be interactively queried and downloaded at: http://mrcieu.mrsoftware.org/immuno_MR/.

Data and Code Availability

eQTL data from the eQTLGen consortium was obtained from http://www.eqtlgen.org - accessed on 18/10/2018. TWAS of whole blood eQTL datasets were obtained from TWAS Hub: http://twas-hub.org/

Code relating to R packages used in this study were obtained from the following: coloc: https://github.com/chr1swallace/coloc; phyper: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Hypergeometric.html; gassocplot: https://github.com/jrs95/gassocplot; pheatmap: https://cran.r-

project.org/web/packages/pheatmap/index.html; shiny: (https://cran.r-project.org/web/packages/shiny/index.html

All further code and data used in this study has been described in the material and methods section.

References

1. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101, 5-22.
2. Ogino, S., Galon, J., Fuchs, C.S., and Dranoff, G. (2011). Cancer immunology--analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol 8, 711-719.
3. Hormozdiari, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet 99, 1245-1260.
4. Neumeyer, S., Hemani, G., and Zeggini, E. (2020). Strengthening Causal Inference for Complex Disease Using Molecular Quantitative Trait Loci. Trends Mol Med 26, 232-241.
5. Richardson, T.G., Hemani, G., Gaunt, T.R., Relton, C.L., and Davey Smith, G. (2020). A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. Nat Commun 11, 185.
6. Hekselman, I., and Yeger-Lotem, E. (2020). Mechanisms of tissue and cell-type specificity in heritable traits and diseases. Nat Rev Genet 21, 137-150.
7. Chaussabel, D., Pascual, V., and Banchereau, J. (2010). Assessing the human immune system through blood transcriptomics. BMC Biol 8, 84.
8. Chaussabel, D. (2015). Assessment of immune status using blood transcriptomics and potential implications for global health. Semin Immunol 27, 58-66.
9. Pascual, V., Chaussabel, D., and Banchereau, J. (2010). A genomic approach to human autoimmune diseases. Annu Rev Immunol 28, 535-571.
10. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat Genet 49, 600-605.
11. Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., and Richards, J.B. (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet 19, 110-124.
12. Restrepo, N.A., Butkiewicz, M., McGrath, J.A., and Crawford, D.C. (2016). Shared Genetic Etiology of Autoimmune Diseases in Patients from a Biorepository Linked to De-identified Electronic Health Records. Front Genet 7, 185.
13. Li, Y.R., Li, J., Zhao, S.D., Bradfield, J.P., Mentch, F.D., Maggadottir, S.M., Hou, C., Abrams, D.J., Chang, D., Gao, F., et al. (2015). Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. Nat Med 21, 1018-1027.
14. Bien, S.A., and Peters, U. (2019). Moving from one to many: insights from the growing list of pleiotropic cancer risk genes. Br J Cancer 120, 1087-1089.
15. 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.

16. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1-22.
17. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 23, R89-98.
18. Porcu, E., Rueger, S., Lepik, K., e, Q.C., Consortium, B., Santoni, F.A., Reymond, A., and Kutalik, Z. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun 10, 3300.
19. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martin, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell 167, 1398-1414 e1324.
20. Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell 175, 1701-1715 e1716.
21. Võsa, U., Claringbould, A., Westra, H., Bonder, M.J., Deelen, P., and Zeng, B. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv.
22. Kibinge, N.K., Relton, C.L., Gaunt, T.R., and Richardson, T.G. (2020). Characterizing the Causal Pathway for Genetic Variants Associated with Neurological Phenotypes Using Human Brain-Derived Proteome Data. Am J Hum Genet.
23. Burgess, S., Dudbridge, F., and Thompson, S.G. (2016). Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 35, 1880-1906.
24. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12, e1001779.
25. Mitchell, R. (Creator), Elsworth, B. (Creator), Raistrick, C. (Creator), Paternoster (Cox), L. (Creator), Hemani, G. (Creator), Gaunt, T. (Creator) (14 Dec 2017). MRC IEU UK Biobank GWAS pipeline version 1. University of Bristol. 10.5523/bris.2fahpksont1zi26xosyamqo8rr
26. Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92-94.
27. 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.
28. International Multiple Sclerosis Genetics, C., Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kemppinen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 45, 1353-1360.

29. Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., Lawrenson, K., Winham, S.J., Dennis, J., Pirie, A., Riggan, M.J., Chornokur, G., et al. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet 49, 680-691.
30. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 50, 928-936.
31. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376-381.
32. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombleson, P., Behrens, T.W., Martin, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 47, 1457-1464.
33. Censin, J.C., Nowak, C., Cooper, N., Bergsten, P., Todd, J.A., and Fall, T. (2017). Childhood adiposity and risk of type 1 diabetes: A Mendelian randomization study. PLoS Med 14, e1002362.
34. Lyon, M.S., Andrews, S.J., Elsworth, B., Gaunt, T.R., Hemani, G., and Marcora, E. (2021). The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biol 22, 32.
35. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife 7.
36. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 46, 1734-1739.
37. Munafo, M.R., and Davey Smith, G. (2018). Robust research needs many lines of evidence. Nature 553, 399-401.
38. Johnson, N.L., Kotz, S., and Kemp, A.W. (1992). Univariate Discrete Distributions, Second Edition.(New York: Wiley.).
39. Falcon, S., and Gentleman, R. (2008). Hypergeometric Testing Used for Gene Set Enrichment Analysis. In Bioconductor Case Studies. (Springer, New York, NY), pp 207-220.
40. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48, 245-252.
41. Jansen, R., Hottenga, J.J., Nivard, M.G., Abdellaoui, A., Laport, B., de Geus, E.J., Wright, F.A., Penninx, B., and Boomsma, D.I. (2017). Conditional eQTL analysis reveals allelic heterogeneity of gene expression. Hum Mol Genet 26, 1444-1451.
42. Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, et al. (2017). Genetic effects on gene expression across human tissues. Nature 550, 204-213.

43. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., and Wilke, C. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
44. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119-124.
45. James, B., Milstien, S., and Spiegel, S. (2019). ORMDL3 and allergic asthma: From physiology to pathology. J Allergy Clin Immunol 144, 634-640.
46. Chen, J., Caspi, R.R., and Chong, W.P. (2018). IL-20 receptor cytokines in autoimmune diseases. J Leukoc Biol 104, 953-959.
47. Martin, J.E., Carmona, F.D., Broen, J.C., Simeon, C.P., Vonk, M.C., Carreira, P., Rios-Fernandez, R., Espinosa, G., Vicente-Rabaneda, E., Tolosa, C., et al. (2012). The autoimmune disease-associated IL2RA locus is involved in the clinical manifestations of systemic sclerosis. Genes Immun 13, 191-196.
48. Ivanov, II, McKenzie, B.S., Zhou, L., Tadokoro, C.E., Lepelley, A., Lafaille, J.J., Cua, D.J., and Littman, D.R. (2006). The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. Cell 126, 1121-1133.
49. Yan, S., Li, Y.Z., Zhu, J.W., Liu, C.L., Wang, P., and Liu, Y.L. (2012). Role of CASP-10 gene polymorphisms in cancer susceptibility: a HuGE review and meta-analysis. Genet Mol Res 11, 3998-4007.
50. Bose, M., Sachsenweger, J., Laurila, N., Parplys, A.C., Willmann, J., Jungwirth, J., Groth, M., Rapakko, K., Nieminen, P., Friedl, T.W.P., et al. (2019). BRCA1 mislocalization leads to aberrant DNA damage response in heterozygous ABRAXAS1 mutation carrier cells. Hum Mol Genet 28, 4148-4160.
51. Alsamman, K., and El-Masry, O.S. (2018). Interferon regulatory factor 1 inactivation in human cancer. Biosci Rep 38.
52. Brezar, V., Tu, W.J., and Seddiki, N. (2015). PKC-Theta in Regulatory and Effector T-cell Functions. Front Immunol 6, 530.
53. Gren, S.T., and Grip, O. (2016). Role of Monocytes and Intestinal Macrophages in Crohn's Disease and Ulcerative Colitis. Inflamm Bowel Dis 22, 1992-1998.
54. Long, Q.Z., Du, Y.F., Ding, X.Y., Li, X., Song, W.B., Yang, Y., Zhang, P., Zhou, J.P., and Liu, X.G. (2012). Replication and fine mapping for association of the C2orf43, FOXP4, GPRC6A and RFX6 genes with prostate cancer in the Chinese population. PLoS One 7, e37866.
55. Goo, Y.H., Son, S.H., Kreienberg, P.B., and Paul, A. (2014). Novel lipid droplet-associated serine hydrolase regulates macrophage cholesterol mobilization. Arterioscler Thromb Vasc Biol 34, 386-396.
56. Wang, L.Q., Wang, Y., Wang, R., Zhao, R.X., Meng, Y., Ren, L.L., Cao, D., Sun, X.L., and Zhang, X.W. (2020). [Changes and clinical significance of peripheral blood CD8(+)CD25(+)T cells in rheumatoid arthritis patients]. Zhonghua Yi Xue Za Zhi 100, 1557-1561.
57. Filion, L.G., Graziani-Bowering, G., Matusevicius, D., and Freedman, M.S. (2003). Monocyte-derived cytokines in multiple sclerosis. Clin Exp Immunol 131, 324-334.
58. Hashemi, V., Maleki, L.A., Esmaily, M., Masjedi, A., Ghalamfarsa, G., Namdar, A., Yousefi, M., Yousefi, B., and Jadidi-Niaragh, F. (2020). Regulatory T cells in breast cancer as a potent anti-cancer therapeutic target. Int Immunopharmacol 78, 106087.

59. Bryant, C., and Fitzgerald, K.A. (2009). Molecular mechanisms involved in inflammasome activation. Trends Cell Biol 19, 455-464.
60. Fodil, N., Moradin, N., Leung, V., Olivier, J.F., Radovanovic, I., Jeyakumar, T., Flores Molina, M., McFarquhar, A., Cayrol, R., Bozec, D., et al. (2017). CCDC88B is required for pathogenesis of inflammatory bowel disease. Nat Commun 8, 932.
61. Wang, Y., Stear, J.H., Swain, A., Xu, X., Bryce, N.S., Carnell, M., Alieva, I.B., Dugina, V.B., Cripe, T.P., Stehn, J., et al. (2020). Drug Targeting the Actin Cytoskeleton Potentiates the Cytotoxicity of Low Dose Vincristine by Abrogating Actin-Mediated Repair of Spindle Defects. Mol Cancer Res 18, 1074-1087.
62. Taylor, K., Davey Smith, G., Relton, C.L., Gaunt, T.R., and Richardson, T.G. (2019). Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization. Genome Med 11, 6.
63. Zhernakova, A., Festen, E.M., Franke, L., Trynka, G., van Diemen, C.C., Monsuur, A.J., Bevova, M., Nijmeijer, R.M., van 't Slot, R., Heijmans, R., et al. (2008). Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. Am J Hum Genet 82, 1202-1210.
64. Vornholz, L., and Ruland, J. (2020). Physiological and Pathological Functions of CARD9 Signaling in the Innate Immune System. Curr Top Microbiol Immunol.
65. Hillmer, E.J., Zhang, H., Li, H.S., and Watowich, S.S. (2016). STAT3 signaling in immunity. Cytokine Growth Factor Rev 31, 1-15.
66. Gharibi, T., Babaloo, Z., Hosseini, A., Abdollahpour-Alitappeh, M., Hashemi, V., Marofi, F., Nejati, K., and Baradaran, B. (2020). Targeting STAT3 in cancer and autoimmune diseases. Eur J Pharmacol 878, 173107.
67. Dcona, M.M., Morris, B.L., Ellis, K.C., and Grossman, S.R. (2017). CtBP- an emerging oncogene and novel small molecule drug target: Advances in the understanding of its oncogenic action and identification of therapeutic inhibitors. Cancer Biol Ther 18, 379-391.
68. Rutz, S., Eidenschenk, C., Kiefer, J.R., and Ouyang, W. (2016). Post-translational regulation of RORgammat-A therapeutic target for the modulation of interleukin-17-mediated responses in autoimmune diseases. Cytokine Growth Factor Rev 30, 1-17.
69. Khan, D., and Ansar Ahmed, S. (2015). Regulation of IL-17 in autoimmune diseases by transcriptional factors and microRNAs. Front Genet 6, 236.
70. Tabarkiewicz, J., Pogoda, K., Karczmarczyk, A., Pozarowski, P., and Giannopoulos, K. (2015). The Role of IL-17 and Th17 Lymphocytes in Autoimmune Diseases. Arch Immunol Ther Exp (Warsz) 63, 435-449.
71. Kennedy, J.M., Fodil, N., Torre, S., Bongfen, S.E., Olivier, J.F., Leung, V., Langlais, D., Meunier, C., Berghout, J., Langat, P., et al. (2014). CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. J Exp Med 211, 2519-2535.
72. McGowan, L.M., Davey Smith, G., Gaunt, T.R., and Richardson, T.G. (2019). Integrating Mendelian randomization and multiple-trait colocalization to uncover cell-specific inflammatory drivers of autoimmune and atopic disease. Hum Mol Genet 28, 3293-3300.

73. Zanin-Zhorov, A., Dustin, M.L., and Blazar, B.R. (2011). PKC-theta function at the immunological synapse: prospects for therapeutic targeting. Trends Immunol 32, 358-363.
74. Anderson, K., Fitzgerald, M., Dupont, M., Wang, T., Paz, N., Dorsch, M., Healy, A., Xu, Y., Ocain, T., Schopf, L., et al. (2006). Mice deficient in PKC theta demonstrate impaired in vivo T cell activation and protection from T cell-mediated inflammatory diseases. Autoimmunity 39, 469-478.
75. Carvalheiro, H., da Silva, J.A., and Souto-Carneiro, M.M. (2013). Potential roles for CD8(+) T cells in rheumatoid arthritis. Autoimmun Rev 12, 401-409.
76. Mitamura, Y., Nunomura, S., Furue, M., and Izuhara, K. (2020). IL-24: A new player in the pathogenesis of pro-inflammatory and allergic skin diseases. Allergol Int.
77. Cohen, G.M. (1997). Caspases: the executioners of apoptosis. Biochem J 326 (Pt 1), 1-16.
78. Boice, A., and Bouchier-Hayes, L. (2020). Targeting apoptotic caspases in cancer. Biochim Biophys Acta Mol Cell Res 1867, 118688.
79. Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet 52, 626-633.

Figure Titles and Legends

Figure 1 - Outline of study workflow
A schematic diagram portraying the analytical pipeline applied in this study.

Figure 2 – Manhattan plots illustrating transcriptome-wide Mendelian randomization results
Transcriptome-wide Mendelian randomization results for A) Asthma, B) Inflammatory bowel disease, C) Breast cancer and D) Hypothyroidism/myxoedema using genetically predicted gene expression from the eQTLGen consortium. Bonferroni corrected thresholds are indicated by the dotted line.

Figure 3 – Locuszoom plots highlighting cell-type dependent effects
Locuszoom plots illustrating colocalization between A) Asthma and PRKCQ expression in T cells, B) Crohn's disease and KSR1 expression in classical monocytes and C) Prostate cancer and C2orf43 expression in non-classical monocytes.

Figure 4 – Heatmaps depicting evidence of enrichment for immune-cell types
Heatmaps portraying evidence of over-representation amongst genetic colocalization results for gene expression derived from immune-cell types and the 14 outcomes analysed in this study.

Figure 5 - **Phenome-wide association study results**
Phenome-wide association study (PheWAS) plots illustrating the genetically predicted effects of A) IL24, B) CASP10 and C) PRKCQ expression using data from the eQTLGen consortium and 311 phenotypes.