Functional genomics implicates natural killer cells as potential key drivers in the pathogenesis of ankylosing spondylitis

Marcos Chiñas[1,3], Daniela Fernandez-Salinas[1,3,4], Vitor R. C. Aguiar[1,2,3], Victor E. Nieto-Caballero[1,3,4], Micah Lefton[5], Peter A. Nigrovic[1,2,5], Joerg Ermann*[2,5], Maria Gutierrez-Arcelus*[1,2,3]

1. Division of Immunology, Boston Children's Hospital, Boston, MA, USA.
2. Harvard Medical School, Boston, MA, USA.
3. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.
4. Licenciatura en Ciencias Genomicas, Centro de Ciencias Genomicas, Universidad Nacional Autónoma de México (UNAM), Morelos 62210, Mexico.
5. Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Boston, MA 02115, USA.

*Co-corresponding authors:

Maria Gutierrez-Arcelus
Boston Children's Hospital
Division of Immunology
Karp Family Research Laboratories, Room 102016.1
1 Blackfan Circle
Boston, MA, 02115
mgutierr@broadinstitute.org
857-559-3229

Joerg Ermann
Brigham and Women's Hospital
Division of Rheumatology, Inflammation and Immunity
Hale Building for Transformative Medicine, Room 6002P
60 Fenwood Road
Boston, MA, 02115
jermann@bwh.harvard.edu
617-525-1227

Abstract

Objective: Multiple lines of evidence indicate that ankylosing spondylitis (AS) is a lymphocyte-driven disease. However, which lymphocyte populations are critical in AS pathogenesis is not known. In this study, we aimed to identify the key cell types mediating the genetic risk in AS using an unbiased integrative functional genomics approach.

Methods: We integrated GWAS data with epigenomic and transcriptomic datasets of immune cells in healthy humans. To quantify enrichment of cell type-specific open chromatin regions or gene expression in AS risk loci, we used three published methods which have identified cell types for other diseases. Additionally, we performed co-localization analyses between GWAS risk loci and genetic variants associated with gene expression (eQTL) to find putative target genes of AS risk variants.

Results: Natural killer (NK) cell-specific open chromatin regions are significantly enriched in heritability for AS, compared to other immune cell types such as T cells, B cells, and monocytes. This finding was consistent between two AS GWAS. Using RNA-seq data, we validated that genes in AS risk loci are enriched in NK cell-specific gene expression. Expression levels of AS-associated genes, such as *RUNX3*, *TBX21*, *TNFRSF1A*, and *NPEPPS*, were found to be highest in NK cells compared to five T cell subsets. Using the human Space-Time Gut Cell Atlas we found significant upregulation of AS-associated genes predominantly in NK cells. Co-localization analysis revealed four AS risk loci affecting regulation of candidate target genes in NK cells: two known loci, *ERAP1 and TNFRSF1A*, and two under-studied loci, *ENTR1* (aka *SDCCAG3*) and *B3GNT2*.

Conclusion: Our results point to NK cells as potential key drivers in the development of AS and highlight four putative target genes for functional follow-up in NK cells.

Introduction

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease characterized by inflammation of the spine and sacroiliac joints, with a proportion of patients also presenting with arthritis in peripheral joints, uveitis, psoriasis or inflammatory bowel disease (1). Historically, most genetic and pathogenetic studies in axSpA have been carried out in ankylosing spondylitis (AS), a severe and well-characterized subtype of axSpA. The heritability of AS is high, with estimates ranging between 40-90% (2). HLA-B27 is the major risk allele for AS (OR = 21.4) (3). Additionally, genome wide-association studies (GWAS) have revealed >100 non-MHC risk loci for AS, most of them implicating non-coding variants (4–7).

Many immune cell-types have been implicated in axSpA (8,9), however, which ones are “driver” cell types involved in the pathogenesis of the disease, versus those that are “bystanders” that participate in inflammation as a consequence of the disease development, remains to be elucidated. Studies leveraging genetic risk variants and their overlap with epigenomic and transcriptomic features suggested CD8+ T cells, CD4+ T cells, NK (natural killer) cells, monocytes, and gastrointestinal cells as mediators of AS genetic risk (9–13). However, these studies have not leveraged new functional genomics datasets that have been generated from human cells or the latest methods designed to integrate functional genomics with GWAS data. This new generation of methods takes advantage of all SNPs tested in a GWAS (not only those that pass the genome-wide significance threshold) and/or robustly control for genomic and linkage disequilibrium biases (14–16).

For other immune-mediated diseases, these integrative functional genomics methods have been successful at identifying cell types that are drivers in disease development. For example, for rheumatoid arthritis (RA), multiple studies have found a significant enrichment of genetic risk in open or active chromatin regions (marking regulatory elements) specific for T cells (10,17–19). Studies in both mouse and human show that T cells are drivers in RA pathogenesis (20–22). For systemic lupus erythematosus (SLE), studies have identified an enrichment of B cell-specific putative regulatory elements and gene expression in SLE risk loci (18,19,23,24) consistent with the well-established role of B cells as a driver of SLE pathogenesis (25,26). Finally, even unexpected findings by integrative functional genomics approaches, such as for Alzheimer's Disease (AD), have been of importance. AD has been traditionally seen as a neurodegenerative disorder characterized by aggregation of amyloid-β plaques in the brain which leads to dementia.

However, GWAS showed that AD genetic risk variants are enriched in regulatory elements specific for the myeloid immune compartment (18,19,27). This has contributed to the recognition of immune dysregulation as a fundamental feature in AD development (28). Hence, integrating GWAS with functional genomics datasets has been instrumental in finding cellular drivers in diseases with complex pathogenesis.

Here we sought to investigate which immune cell populations could be drivers of AS development using established methods that control for biases in genomic enrichment analyses. We used GWAS summary statistics from two different AS cohorts and integrated them with epigenomic and transcriptomic datasets of human leukocytes from peripheral blood and tissue. Our results bring forward NK cells as key candidate drivers in the pathogenesis of AS.

Results

In order to assess which immune cell types might be mediating the genetic susceptibility to AS, we first utilized a dataset of open chromatin profiles (ATAC-seq) of immune cell subsets from peripheral blood of four healthy subjects (18), **Fig. 1A**). Sorted cell subsets were analyzed with and without in vitro stimulation. For our study, we grouped the samples into 7 main immune cell types: T cells, B cells, NK cells, plasma cells, dendritic cells (DCs), plasmacytoid DCs, and monocytes. We identified cell type-specific open chromatin regions and assessed whether these were significantly enriched in AS genetic risk. We removed the MHC region from all of our analyses given the unusually high linkage disequilibrium in this region and the fact that it is coding variants in HLA-B that mostly drive disease risk in this locus. Specifically, we used the LDSC-SEG method (14) to quantify enrichment of partitioned heritability in each of these cell type-specific annotations (conceptual scheme in **Fig. 1B, data in Fig. 1C**) compared to baseline and control annotations, while controlling for linkage disequilibrium. Using the ImmunoChip association study summary statistics from the International Genetics of Ankylosing Spondylitis Consortium (IGAS) (7) (9,069 patients and 13,578 healthy controls), we found that NK-specific open chromatin regions were significantly enriched in genetic risk for AS ($P = 0.026$), while this was not the case for the other six immune cell types (**Fig. 1D**).

To validate this finding in an independent association study where genome-wide genotyping was performed (in contrast to the Immunochip), we used the summary statistics for AS from the UK

Biobank, with 1,185 patients and 419,276 healthy controls. With this GWAS, we confirmed that open chromatin regions specific for NK cells were significantly enriched in AS heritability (P = 0.034, **Fig. 1E**). To evaluate the confidence of our results, we used four "control traits" for which we know what results to expect given they have been often studied in this type of analyses integrating GWAS with functional genomics (14,17–19). We showed as expected that RA presented the highest enrichment for T cell-specific open chromatin regions (P = 0.0018), Alzheimer's for myeloid DC (P = 0.00018), and SLE for B cells (P = 0.0015). We chose body height as a "negative control" for which we do not expect any enrichment for immune cells, and indeed it was the case (all P > 0.1, **Supplementary Fig. 1**). Overall, these analyses at the epigenomic level indicate that AS risk alleles are enriched in regions likely affecting gene regulation in NK cells.

To test whether using orthogonal approaches would support this conclusion, we used our previously published RNA-seq dataset of sorted peripheral CD4+ T, CD8+ T, MAIT, invariant NKT (iNKT), γδ T cells expressing Vδ1 TCR chain (Vd1), γδ T cells expressing Vδ2 TCR chain (Vd2), and NK cells (each in duplicate from 6 healthy donors, **Fig. 2A**) (29). We applied the SNPsea method, which quantifies enrichment of cell type-specific gene expression in risk loci for a given trait (conceptual scheme in **Fig. 2B**) by employing a non-parametric statistical method to calculate empirical P-values through comparison with sets of null SNPs (30). Here we used the AS risk SNPs reported by Brown and Wordsowrth in 2017 which were curated from multiple AS genetic studies (31). SNPsea revealed that NK-specific gene expression (compared to the other T cell subsets in the dataset) presented significant enrichment in AS risk loci (P = 0.01), whereas none of the other lymphocyte subsets were significant (**Fig. 2C**).

Examples of AS-associated genes with highest expression in NK cells are presented in Fig 2D. Two of these genes encode for transcription factors with important roles in lymphocytes: *RUNX3* and *TBX21* (Tbet). *TNFRSF1A* codes for the TNFα Receptor Superfamily Member 1A, the association with AS for this locus has been validated by multiple studies (32–34). *FCGR2A* encodes for the low-affinity activating IgG immune complex receptor FcγR2A (*FCGR2A*) known to bind IgG and mediate immune processes. Less studied genes included *NPEPPS,* which encodes a puromycin-sensitive aminopeptidase, and *LNPEP*, which encodes for a zinc-dependent aminopeptidase (35). Both genes are paralogs of *ERAP1* and belong to the Class I MHC-mediated antigen processing and presentation pathway, along with other known AS risk

genes including *HLA-B*, *ERAP1*, *ERAP2* (36–38). Overall, our second integrative analysis indicates that genes found in risk loci for AS are highly expressed in NK cells compared to T cells, providing additional support for the possibility that AS risk alleles exert their regulatory effects, at least in part, via NK cells.

The transcriptomic phenotype of immune cells commonly differs between blood and tissue (39–42). Hence, to evaluate disease-relevant cell subsets from a tissue relevant for AS (as opposed to peripheral blood as in the datasets above), we used the human Space-Time Gut Cell Atlas (43). This resource includes scRNA-seq data for samples from various locations of fetal (N = 16), pediatric (N = 8) and adult (N = 13, including 6 healthy and 7 Crohn's disease patients) intestine (**Fig. 3A)**. We applied scDRS (15) to this data set. scDRS identifies cells that over-express a significant proportion of genes implicated by GWAS, weighted on their strength of association with disease, compared to null sets of control genes in the dataset (conceptual scheme in **Fig. 3B**). The Space-Time Gut Cell Atlas investigators identified the following broad cell types: mesenchymal, epithelial, endothelial, neuronal, myeloid, red blood cells, B cells, plasma cells, T cells, NK cells and other innate lymphoid cells (**Fig. 3C)**. scDRS revealed 1,852 cells with significant enrichment for expression AS GWAS genes (20% FDR, **Fig. 3D**). Of these, 765 were T cells, 264 myeloid cells, 320 NK cells and 319 other ILCs. Normalized for cell type abundance in the dataset, NK cells showed the highest enrichment (39-fold), followed by other ILCs (34-fold), T cells (5-fold), and myeloid cells (5-fold, **Fig. 3E**). In contrast, non-immune cell types were under-represented disease relevant cells (**Supplementary Fig. 2**). We then used the fine-grained annotations of the Space-Time Gut Cell Atlas to identify the top cell subsets with absolute numbers of cells with significant scDRS. This revealed NK cells as the most abundant (N = 320), followed by Lti-like NCR+ ILC3 cells (N = 147), activated CD8+ T cells (N = 132), macrophages (N = 130), Lti-like NCR- ILC3 cells (N = 112), $\gamma\delta$ T cells (N = 94), and other T cells, ILCs and myeloid subsets (**Fig. 3F**). Some of the AS-associated genes with high expression in NK cells include *RUNX3*, *TBX21* and*TNFRSF1A* (**Fig. 3G**). Using our control traits, we found as expected that the main disease-relevant cell type is T cells for RA and myeloid cells for Alzheimer's disease (**Supplementary Fig. 2**). We did not find any significant disease-relevant cells for height (as expected) and for SLE, which could mean B cells in the gut in a state not relevant for SLE or there wasn't enough power in this dataset for this particular disease (**Supplementary Fig. 2**). Overall, these analyses indicate that NK cells found in tissue exhibit significant expression of AS-associated genes.

Given that our analyses indicate that AS risk loci could be acting through NK cells, we next sought to find putative target genes for AS risk variants in NK cells. To do this we performed co-localization analyses between AS GWAS risk loci and genetic variants associated with gene expression (expression quantitative trail loci, eQTLs) using coloc (44). We utilized eQTL summary statistics made available by the eQTL Catalogue (45) for a study involving transcriptomic profiling of peripheral NK cells from 91 genotyped individuals (46) and another microarray QTL study in which transcriptomic profiling of NK cells was conducted on 245 genotyped individuals (47). We found four AS genome-wide significant risk loci ($P < 5 \times 10^{-8}$) with a posterior probability of a shared causal variant with an NK eQTL greater than 0.8 (PP4, **Table 1, Fig. 4**). An additional 10 loci with suggestive P-values in their association with AS (3.56 × 10-5 < P < 5.40 x 10-8), presented evidence of co-localization with NK eQTL for 18 genes (PP4 > 0.75, **Table 1**). Among the genome-wide significant loci are the known target genes *ERAP1* and *TNFRSF1A*, as well as the putative target genes *ENTR1* (a.k.a. *SDCCAG3*) and *B3GNT2*, which have been less studied.

Discussion

In this study we integrated epigenomic and transcriptomic datasets with AS genetic risk data to find candidate drivers of AS pathogenesis. Our unbiased approach, applying three different methods to datasets from both peripheral blood and tissue, consistently identified NK cells as the dominant disease-relevant cell type. Specifically, we found that NK-specific open chromatin regions and NK-specific gene expression were significantly enriched for non-MHC AS genetic risk. This suggests that a significant portion of AS risk variants likely affect gene regulation in NK cells, pointing to NK cells as potential key drivers of AS pathogenesis.

Previous studies support a role for NK cells in AS. NK cells have the ability to directly destroy infected cells through cell lysis, and in addition they play a significant role in shaping T cell responses by releasing cytokines. AS patients with chronic subclinical intestinal inflammation presented an increased abundance of NKp44+ NK cells in their gut, and these cells were the major producers of IL-22 in the lamina propria, suggesting a possible role in tissue protection (48). Additionally, it has been shown that HLA-B27 can bind to killer cell immunoglobulin-like receptor (KIR) KIR3DL1 and affect NK cell functions, including their ability to lyse cells (49–51). In another study, investigators co-cultured ERAP1-inhibited M1 macrophages with NK cells from AS patients, and found that patients with ERAP1 protective alleles led to decreased CD69 and CD107a on NK cells and a lower number of IFN-γ+ NK cells compared to patients carrying non-

protective alleles (52). Furthermore, multiple risk loci for AS include genes relevant for NK cell function, including KIRs *KIR2DL1*, *KIR3DL1*, *KIR2DS5*, *KIR3DS1* and *KIR2DL5* (53–57).

To note, our findings do not exclude involvement of other cell types in AS pathogenesis. Indeed, in the human Space-Time Gut Cell Atlas, we identified significant expression of AS-associated genes in T cell subsets and ILC subsets, which share transcriptional programs with NK cells (29,58–60). Indeed, it is likely that genetic risk to AS is mediated through multiple cell types, as is the case for other complex diseases such as RA, for which studies have found risk enrichment in open/active chromatin regions specific to activated T cells and B cells (18,19,61,62). We and others have shown that eQTLs often exhibit impact in multiple cell types (63,64). Hence, determining the specific cell type through which a disease risk variant is exerting its pathogenic effects can be challenging.

Our co-localization analyses using two eQTL NK cell datasets identified four putative target genes for AS risk variants: *ERAP1*, *TNFRSF1A*, *ENTR1* (a.k.a. *SDCCAG3*) and *B3GNT2*. The importance of *ERAP1* in AS risk is well established, and polymorphisms affecting its expression have been reported for multiple cell types, including macrophages, monocytes, T cells, and induced pluripotent stem cells, fibroblasts, and immortalized B cells (65–67). Similarly, multiple studies have found significant associations between non-coding polymorphisms at or near *TNFRSF1A* and AS, including in European and East Asian populations (32,33,68,69). While there are multiple genes in this genomic locus, including *PLEKHG6*, *SCNN1A*, and *LTBR*, our co-localization results suggest that *TNFRSF1A*, which encodes for a TNF receptor, is the target gene of the causal variant in this locus, and its dysregulation can happen in NK cells. This is consistent with the therapeutic efficacy of TNF inhibitors in AS and the known function of TNF as a booster of the cytolytic capacity of NK cells (70). Interestingly, TNFRSF1A has been functionally linked to ENTR1, a less extensively studied putative target gene identified in this study. *ENTR1*, which encodes for an endosome-associated trafficking regulator, is needed for the presentation of TNF receptor on the cell surface (71). Lastly, *B3GNT2* encodes an acetylglucosaminyltransferase enzyme that is a type II transmembrane protein. A recent study in a Taiwanese cohort confirmed that a non-coding genetic variant near *B3GNT2* is associated with AS susceptibility, and that *B3GNT2* blood mRNA levels were negatively correlated with C-reactive protein (CRP), erythrocyte sedimentation rate, syndesmophyte formation, and Bath ankylosing spondylitis functional index (BASFI) (72).

While the putative target genes identified here make sense in the context of AS and potential roles in NK cell function, they are not as numerous as we would have expected. This has been a common problem for non-coding risk variants for other complex diseases, where only 25-38% of risk variants co-localize with eQTLs (73–75). We and others have highlighted that these “missing regulatory effects” could be hidden in activation or differentiation cell states that have not been systematically ascertained (76–79). The typical sample size of eQTL studies is likely insufficient to find the regulatory effects of most risk variants identified by GWAS (80). Hence, we believe that better-powered eQTL studies ascertaining multiple activation NK cell states will be needed to find additional target genes for AS risk variants.

Other limitations of our study include not having ascertained all possible immune cell subsets and cell states, especially those that might be present in AS patients, and particularly those found in inflamed sacroiliac joints. Hence, if the real driver for AS pathogenesis is a different cell subset or state that was not present in our interrogated datasets, but has transcriptomic and epigenomic similarities to NK cells, then our results may suffer from “guilt-by-association” bias. To our knowledge, current transcriptomic datasets from AS patients are limited to peripheral blood (13,81–86). When we applied scDRS to a recently published single-cell RNA-seq dataset of 98,884 PBMCs cells from 10 patients and 29 healthy controls (84), we found no significant cells for the disease-relevant score (data not shown), possibly due to lack of power in the study for this type of analysis.

However, our study covered a wide range of immune cell states found in the gastrointestinal tract and peripheral blood of healthy subjects and consistently pointed to NK cells. Since GWAS identifies genetic loci that are involved in disease development, including early phases when future patients are still “healthy”, it is relevant to study samples from healthy subjects, even if not all cell states may be captured. Future larger-scale studies in samples from blood and inflamed joint tissue from AS patients, especially for untreated patients in the early phases of the disease, will be key to understand whether NK cells are indeed drivers of AS pathogenesis.

Materials and Methods

Genome-wide association studies

We used the GWAS ImmunoChip summary statistics from the International Genetics of Ankylosing Spondylitis Consortium (IGAS). The IGAS study, led by (7), performed a high-density genotyping of 9,069 AS cases and 13,578 healthy controls. In addition, we used the GWAS summary statistics from UK Biobank, which involved a case-control design with 1,185 AS cases and 419,276 controls, providing genome-wide coverage for AS susceptibility loci (87).

We lifted the genomic positions of the genetic variants to genome build hg19 or hg38 according to the version compatible with subsequent analyses. Given the complexity and strong genetic association signals within the Major Histocompatibility Complex (MHC) region, we excluded variants located on chr6:25,000,000–34,000,000.

We additionally used GWAS summary statistics from rheumatoid arthritis, Alzheimer's disease, and systemic lupus erythematosus as positive control traits for which we know the disease relevant immune cell types, and height as a negative control for which we do not expect for immune cells to be relevant. These summary statistics were preprocessed by the Alkes Price laboratory, they included HapMap 3 SNPs and SNPs that are in 1000G, and they excluded the MHC region (chr6:25Mb-34Mb). These summary statistics are available at https://alkesgroup.broadinstitute.org/.

Epigenomic and transcriptomic datasets

To identify cell type-specific open chromatin regions in different immune cell types, we used the Calderon *et al.* study *(18)*, in which the authors collected blood from 4 healthy subjects, sorted immune cell types, and generated chromatin accessibility profiles using Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq, GSE118189).

To find AS risk enrichment for cell type-specific expression, we incorporated data from the study conducted by Gutierrez-Arcelus *et al.* (29), which involved low-input mRNA-seq data from sorted NK cells and six T cell subsets isolated from 6 healthy subjects (each with two replicates per cell-type, GSE124731).

We used the Space-Time Gut Cell Atlas to identify cells exhibiting significant upregulation of disease-associated genes. This dataset includes single-cell RNA-seq profiling of 428,000 intestinal cells obtained from fetal (N = 16), pediatric (N = 8), and adult donors (N = 13). The dataset covers up to 11 different intestinal regions (43), https://www.gutcellatlas.org/.

Differential accessibility analysis

We used the counts of open chromatin consensus peaks called by Calderon et al. First, we transformed counts into Reads Per Kilobase per Million mapped reads (RPKM), then normalized by quantiles using the preprocess Core R package and finally scaled to their log2(normalized RPKM+1), thus we account for differences in library size across samples and peak length variability. We pooled sorted samples into 7 main immune cell types: T cells, B cells, plasma cells, dendritic cells, plasma dendritic cells, natural killer cells, monocytes.

Next, we employed linear mixed model regression to identify regions that exhibited differential accessibility between each cell type and the rest of the cell types. To account for potential donor-specific effects, we incorporated the donor ID variable as a random effect in our analysis.
For each cell type comparison, we tested peaks that had counts greater than the mean for that cell type in at least half of the samples, this yielded between 400 to 600 thousand peaks depending on the cell type. We then sorted open chromatin peaks by their P-value and chose the positive top 62,531 peaks, which represented 10% of the maximum number of peaks tested (plasmacytoid DCs 625,312) except for plasma cells, for which we chose 40,307 peaks which was the maximum number of peaks with $P<0.05$), thus selecting the same number of cell type-specific peaks for most cell types.

Partitioned heritability enrichment analysis with LDSC-SEG

Linkage Disequilibrium score regression applied to specifically expressed genes (LDSC-SEG) v1.0.1 method was applied to determine disease-relevant cell types (14,88) for AS.

Cell type-specific open chromatin peaks were extended by 225bp to each side, to match the genomic coverage recommended by the LDSC-SEG authors. These annotations were then utilized as input for the partitioned heritability enrichment analysis by LDSC-SEG. We used the

baseline annotation v1.2 provided by Price Lab for LDSC-SEG, comprising 75 background annotations. Additionally, we used all consensus peaks (N = 829,942) of Calderon et al. as the control annotation. Using other baselines or controls did not affect our results. We utilized SNP weight files derived from the HapMap 3 project (HM3) European population.

Analysis of cell type-specific gene expression enrichment in risk loci using SNPsea

SNPsea analysis aimed to assess the association between risk SNPs and genes expressed specifically for a given cell type (30). We incorporated a curated list of risk SNPs for ankylosing spondylitis (AS), compiled by Brown and Wordsowrth 2017 (31), which includes genetic variants that have been associated with AS susceptibility. This list was derived from multiple AS studies conducted until 2017.

We utilized the expression data obtained from Gutierrez-Arcelus et al. (2019). The gene expression counts in this dataset were normalized to transcripts per million (TPM) and transformed to log2(TPM+1) values. To identify the genes with meaningful expression levels, we included those with log2(TPM+1) > 2 in at least 10 samples. SNPsea was then run for the normalized expression matrix and AS risk SNPs, using recombination intervals from Myers et al. (89), null SNPs from Lango et al. (90), and the following parameters: --score single --slop 10000 --threads 2 --null-snpsets 0 --min-observations 100 --max-iterations 10000000.

Integration of GWAS with single-cell RNA-seq with scDRS

We used the single-cell Disease Relevance Score (scDRS) by combining scRNA-seq and GWAS to identify cells with significant up-regulation of disease-associated genes, which are scored based on their strength of association with disease, and are compared with null sets of genes present in the dataset.

As recommended by scDRS authors, we first created disease-relevant genesets using Multi-marker Analysis of GenoMic Annotation (MAGMA) version 1.10 (91). First, we generated gene annotations with MAGMA setting a window of 10kb using the following parameters: “--annotate window=10,10 --snp-loc ./g1000_eur/g1000_eur.bim --gene-loc ./NCBI37.3/NCBI37.3.gene.loc”. Then we ran MAGMA using GWAS summary statistics for traits of interest with the following

parameters: --bfile ./magma_v1.10/g1000_eur/g1000_eur --pval GWAS.pval use='SNP,P' ncol='N' --gene-annot ./magma_v1.10/out/step1.genes.annot.

We ran scDRS using the disease-relevant gene sets from MAGMA, the expression data obtained from the Space-Time Gut Cell Atlas (43) and corrected for biases by adding as covariates the number of genes expressed per cell and sample batch. Then, for visualization purposes and downstream analysis, we processed the single-cell dataset using Seurat (92) performed integration across batches with Harmony (93), and visualized cells in two dimensions with Uniform Manifold Approximation and Projection (UMAP). We labeled cells plotted in UMAP by the annotations defined by the Space-Time Gut Cell Atlas. Additionally, we colored cells by their scDRS score, when cells passed the 0.20 FDR threshold.

eQTL co-localization analysis

To select regions for colocalization, GWAS summary statistics were sorted by P-values, then starting from the variant with the smallest P-value, variants within a 50 Kb window were removed. The process was repeated with the next most significant variant among the remaining variants until no variant with a P-value below 5×10^{-5} was left. We performed colocalization analysis for GWAS studies against the eQTL Catalogue (45). We imported eQTL summary statistics from RNA-seq and microarray from Schmiedel et al. (46) and Gilchrist et al. (2022). We fetched the summary statistics data using the tabix method with the seqminer R package (v8.5). For each region tested, we included all biallelic SNPs that were ascertained in both the GWAS and eQTL study and performed the analysis only for genes within a window of ±500,000 base pairs from the GWAS top variant, and for which there was at least one eQTL passing the 5×10^{-5} P-value threshold. Before merging GWAS and QTL data, the variant coordinates of the GWAS were lifted to the GRCh38 version of the reference genome using liftOver with the UCSC chain file. We used the coloc v5.1.0.1 package (94) in R v4.1.0 to test for colocalization at each gene and dataset.

Each locus was plotted using plotgardener (95), and we recovered the LD of the top SNP in a given region in the GWAS dataset using the locuscomparer package (96). Then we used plotgardener functions to display the regions near the lead variant and colored the genes tested using the posterior probability that the two traits share a causal variant (PP4).

Data availability

All data and methods are publicly available as specified above.

Acknowledgments

This study was supported by a seed grant from the Spondyloarthritis Research and Treatment Network (SPARTAN) and a microgrant from the Joint Biology Consortium (1P30AR070253-01). PAN was supported by P30AR070253 and R01AR073201. JE was supported by NIH Grant R21 AR076040-01 and an ASPIRE grant from Pfizer. MGA was supported by P30AR070253, the Arthritis National Research Foundation, the Lupus Research Alliance, and the Gilead Sciences Rheumatology Research Scholars Award. We thank Soumya Raychaudhuri and Kamil Slowikowski for guidance on implementing SNPsea, and Steven Gazal for guidance on implementing LDSC-SEG.

Disclosures

The authors declare they have no relevant conflicts of interest.

References

1. Navarro-Compán V, Sepriano A, El-Zorkany B, Heijde D van der. Axial spondyloarthritis. *Ann Rheum Dis* 2021;80:1511–1521.

2. Díaz-Peña R, Castro-Santos P, Durán J, Santiago C, Lucia A. The Genetics of Spondyloarthritis. *J Pers Med* 2020;10. Available at: http://dx.doi.org/10.3390/jpm10040151.

3. Reveille JD, Zhou X, Lee M, Weisman MH, Yi L, Gensler LS, et al. HLA class I and II alleles in susceptibility to ankylosing spondylitis. *Ann Rheum Dis* 2019;78:66–73.

4. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;39:1329–1337.

5. Australo-Anglo-American Spondyloarthritis Consortium (TASC), Reveille JD, Sims A-M, Danoy P, Evans DM, Leo P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet* 2010;42:123–127.

6. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 2011;43:761–767.

7. International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J,

Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 2013;45:730–738.

8. Mauro D, Simone D, Bucci L, Ciccia F. Novel immune cell phenotypes in spondyloarthritis pathogenesis. *Semin Immunopathol* 2021;43:265–277.

9. Li Z, Haynes K, Pennisi DJ, Anderson LK, Song X, Thomas GP, et al. Epigenetic and gene expression analysis of ankylosing spondylitis-associated loci implicate immune cells and the gut in the disease pathogenesis. *Genes Immun* 2017;18:135–143.

10. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015;518:337–343.

11. Costantino F, Breban M, Garchon H-J. Genetics and Functional Genomics of Spondyloarthritis. *Front Immunol* 2018;9:2933.

12. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 2022;376:eabf3041.

13. Brown AC, Cohen CJ, Mielczarek O, Migliorini G, Costantino F, Allcock A, et al. Comprehensive epigenomic profiling reveals the extent of disease-specific chromatin states and informs target discovery in ankylosing spondylitis. *Cell Genom* 2023;3:100306.

14. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* 2018;50:621–629.

15. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* 2022;54:1572–1580.

16. Jagadeesh KA, Dey KK, Montoro DT, Mohan R, Gazal S, Engreitz JM, et al. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet* 2022;54:1479–1492.

17. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013;45:124–130.

18. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* 2019;51:1494–1505.

19. Soskic B, Cano-Gamez E, Smyth DJ, Rowan WC, Nakic N, Esparza-Gordillo J, et al. Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat Genet* 2019;51:1486–1493.

20. Banerjee S, Webber C, Poole AR. The induction of arthritis in mice by the cartilage proteoglycan aggrecan: roles of CD4+ and CD8+ T cells. *Cell Immunol* 1992;144:347–357.

21. Kobezda T, Ghassemi-Nejad S, Mikecz K, Glant TT, Szekanecz Z. Of mice and men: how animal models advance our understanding of T-cell function in RA. *Nat Rev Rheumatol*

2014;10:160–170.

22. Rao DA, Gurish MF, Marshall JL, Slowikowski K, Fonseka CY, Liu Y, et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* 2017;542:110–114.

23. Hu X, Kim H, Stahl E, Plenge R, Daly M, Raychaudhuri S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* 2011;89:496–506.

24. Khunsriraksakul C, Li Q, Markus H, Patrick MT, Sauteraud R, McGuire D, et al. Multi-ancestry and multi-trait genome-wide association meta-analyses inform clinical risk prediction for systemic lupus erythematosus. *Nat Commun* 2023;14:668.

25. Caielli S, Wan Z, Pascual V. Systemic Lupus Erythematosus Pathogenesis: Interferon and Beyond. *Annu Rev Immunol* 2023;41:533–560.

26. Vinuesa CG, Shen N, Ware T. Genetics of SLE: mechanistic insights from monogenic disease and disease-associated variants. *Nat Rev Nephrol* 2023. Available at: http://dx.doi.org/10.1038/s41581-023-00732-x.

27. Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai L-H, et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* 2015;518:365–369.

28. Bettcher BM, Tansey MG, Dorothée G, Heneka MT. Peripheral and central immune system crosstalk in Alzheimer disease - a research prospectus. *Nat Rev Neurol* 2021;17:689–701.

29. Gutierrez-Arcelus M, Teslovich N, Mola AR, Polidoro RB, Nathan A, Kim H, et al. Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nat Commun* 2019;10:687.

30. Slowikowski K, Hu X, Raychaudhuri S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* 2014;30:2496–2497.

31. Brown MA, Wordsworth BP. Genetics in ankylosing spondylitis - Current state of the art and translation into clinical outcomes. *Best Pract Res Clin Rheumatol* 2017;31:763–776.

32. Davidson SI, Liu Y, Danoy PA, Wu X, Thomas GP, Jiang L, et al. Association of STAT3 and TNFRSF1A with ankylosing spondylitis in Han Chinese. *Ann Rheum Dis* 2011;70:289–292.

33. Karaderi T, Pointon JJ, Wordsworth TWH, Harvey D, Appleton LH, Cohen CJ, et al. Evidence of genetic association between TNFRSF1A encoding the p55 tumour necrosis factor receptor, and ankylosing spondylitis in UK Caucasians. *Clin Exp Rheumatol* 2012;30:110–113.

34. Sode J, Bank S, Vogel U, Andersen PS, Sørensen SB, Bojesen AB, et al. Genetically determined high activities of the TNF-alpha, IL23/IL17, and NFkB pathways were associated with increased risk of ankylosing spondylitis. *BMC Med Genet* 2018;19:165.

35. Robinson PC, Brown MA. Genetics of ankylosing spondylitis. *Mol Immunol* 2014;57:2–11.

36. Vitulano C, Tedeschi V, Paladini F, Sorrentino R, Fiorillo MT. The interplay between HLA-B27 and ERAP1/ERAP2 aminopeptidases: from anti-viral protection to spondyloarthritis. *Clin*

Exp Immunol 2017;190:281–290.

37. Tsui FW, Tsui HW, Akram A, Haroon N, Inman RD. The genetic basis of ankylosing spondylitis: new insights into disease pathogenesis. *Appl Clin Genet* 2014;7:105–115.

38. Agrawal N, Brown MA. Genetic associations and functional characterization of M1 aminopeptidases and immune-mediated diseases. *Genes Immun* 2014;15:521–527.

39. Szabo PA, Levitin HM, Miron M, Snyder ME, Senda T, Yuan J, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun* 2019;10:1–16.

40. Meininger I, Carrasco A, Rao A, Soini T, Kokkinou E, Mjösberg J. Tissue-Specific Features of Innate Lymphoid Cells. *Trends Immunol* 2020;41:902–917.

41. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* 2022;376:eabl5197.

42. Mass E, Nimmerjahn F, Kierdorf K, Schlitzer A. Tissue-specific macrophages: how they develop and choreograph tissue biology. *Nat Rev Immunol* 2023:1–17.

43. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. *Nature* 2021;597:250–255.

44. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* 2020;82:1273–1300.

45. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* 2021;53:1290–1299.

46. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* 2018;175:1701–1715.e16.

47. Gilchrist JJ, Makino S, Naranbhai V, Sharma PK, Koturan S, Tong O, et al. Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease associations, highlighting their role in autoimmunity. *Nat Commun* 2022;13:4073.

48. Ciccia F, Accardo-Palumbo A, Alessandro R, Rizzo A, Principe S, Peralta S, et al. Interleukin-22 and interleukin-22-producing NKp44+ natural killer cells in subclinical gut inflammation in ankylosing spondylitis. *Arthritis Rheum* 2012;64:1869–1878.

49. Peruzzi M, Wagtmann N, Long EO. A p70 killer cell inhibitory receptor specific for several HLA-B allotypes discriminates among peptides bound to HLA-B*2705. *J Exp Med* 1996;184:1585–1590.

50. Stewart-Jones GBE, Gleria K di, Kollnberger S, McMichael AJ, Jones EY, Bowness P. Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. *Eur J Immunol* 2005;35:341–351.

51. Malnati MS, Peruzzi M, Parker KC, Biddison WE, Ciccone E, Moretta A, et al. Peptide

specificity in the recognition of MHC class I by natural killer cell clones. *Science* 1995;267:1016–1018.

52. Babaie F, Mohammadi H, Salimi S, Ghanavatinegad A, Abbasifard M, Yousefi M, et al. Inhibition of ERAP1 represses HLA-B27 free heavy chains expression on polarized macrophages and interrupts NK cells activation and function from ankylosing spondylitis. *Clin Immunol* 2023;248:109268.

53. Wang S, Li G, Ge R, Duan Z, Zeng Z, Zhang T, et al. Association of KIR genotype with susceptibility to HLA-B27-positive ankylosing spondylitis. *Mod Rheumatol* 2013;23:538–541.

54. Jiao Y-L, Zhang B-C, You L, Li J-F, Zhang J, Ma C-Y, et al. Polymorphisms of KIR gene and HLA-C alleles: possible association with susceptibility to HLA-B27-positive patients with ankylosing spondylitis. *J Clin Immunol* 2010;30:840–844.

55. Díaz-Peña R, Blanco-Gelaz MA, Suárez-Alvarez B, Martínez-Borra J, López-Vázquez A, Alonso-Arias R, et al. Activating KIR genes are associated with ankylosing spondylitis in Asian populations. *Hum Immunol* 2008;69:437–442.

56. Lopez-Larrea C, Blanco-Gelaz MA, Torre-Alonso JC, Bruges Armas J, Suarez-Alvarez B, Pruneda L, et al. Contribution of KIR3DL1/3DS1 to ankylosing spondylitis in human leukocyte antigen-B27 Caucasian populations. *Arthritis Res Ther* 2006;8:R101.

57. Harvey D, Pointon JJ, Sleator C, Meenagh A, Farrar C, Sun JY, et al. Analysis of killer immunoglobulin-like receptor genes in ankylosing spondylitis. *Ann Rheum Dis* 2009;68:595–598.

58. Sun JC, Lanier LL. NK cell development, homeostasis and function: parallels with $CD8^+$ T cells. *Nat Rev Immunol* 2011;11:645–657.

59. Cohen NR, Brennan PJ, Shay T, Watts GF, Brigl M, Kang J, et al. Shared and distinct transcriptional programs underlie the hybrid nature of iNKT cells. *Nat Immunol* 2013;14:90–99.

60. Robinette ML, Fuchs A, Cortez VS, Lee JS, Wang Y, Durum SK, et al. Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat Immunol* 2015;16:306–317.

61. Terao C, Raychaudhuri S, Gregersen PK. Recent Advances in Defining the Genetic Basis of Rheumatoid Arthritis. *Annu Rev Genomics Hum Genet* 2016;17:273–301.

62. Weinand K, Sakaue S, Nathan A, Jonsson AH, Zhang F, Watts GFM, et al. The Chromatin Landscape of Pathogenic Transcriptional Cell States in Rheumatoid Arthritis. *bioRxiv* 2023. Available at: http://dx.doi.org/10.1101/2023.04.07.536026.

63. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet* 2015;11:e1004958.

64. 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;167:1398–1414.e24.

65. Costantino F, Talpin A, Evnouchidou I, Kadi A, Leboime A, Said-Nahal R, et al. ERAP1

Gene Expression Is Influenced by Nonsynonymous Polymorphisms Associated With Predisposition to Spondyloarthritis. *Arthritis Rheumatol* 2015;67:1525–1534.

66. Chen L, Ridley A, Hammitzsch A, Al-Mossawi MH, Bunting H, Georgiadis D, et al. Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis. *Ann Rheum Dis* 2016;75:916–923.

67. Hanson AL, Cuddihy T, Haynes K, Loo D, Morton CJ, Oppermann U, et al. Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile. *Arthritis Rheumatol* 2018;70:255–265.

68. Zhao S, Chen H, Wu G, Zhao C. The association of NLRP3 and TNFRSF1A polymorphisms with risk of ankylosing spondylitis and treatment efficacy of etanercept. *J Clin Lab Anal* 2017;31. Available at: http://dx.doi.org/10.1002/jcla.22138.

69. Xing-Rong W, Sheng-Qian X, Wen L, Shan Q, Fa-Ming P, Jian-Hua X. Role of TNFRSF1A and TNFRSF1B polymorphisms in susceptibility, severity, and therapeutic efficacy of etanercept in human leukocyte antigen-B27-positive Chinese Han patients with ankylosing spondylitis. *Medicine* 2018;97:e11677.

70. Wang R, Jaw JJ, Stutzman NC, Zou Z, Sun PD. Natural killer cell-produced IFN-γ and TNF-α induce target cell cytolysis through up-regulation of ICAM-1. *J Leukoc Biol* 2012;91:299–309.

71. Neznanov N, Neznanova L, Angres B, Gudkov AV. Serologically defined colon cancer antigen 3 is necessary for the presentation of TNF receptor 1 on cell surface. *DNA Cell Biol* 2005;24:777–785.

72. Wang C-M, Jan Wu Y-J, Lin J-C, Huang L-Y, Wu J, Chen J-Y. Genetic effects of B3GNT2 on ankylosing spondylitis susceptibility and clinical manifestations in Taiwanese. *J Formos Med Assoc* 2022;121:1283–1294.

73. Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* 2017;49:600–605.

74. Mu Z, Wei W, Fair B, Miao J, Zhu P, Li YI. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol* 2021;22:1–28.

75. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol* 2021;22:49.

76. Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. *Nat Rev Genet* 2016;17:160–174.

77. Gutierrez-Arcelus M, Baglaenko Y, Arora J, Hannes S, Luo Y, Amariuta T, et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat Genet* 2020;52:247–253.

78. Umans BD, Battle A, Gilad Y. Where Are the Disease-Associated eQTLs? *Trends Genet* 2021;37:109–124.

79. Connally NJ, Nazeen S, Lee D, Shi H, Stamatoyannopoulos J, Chun S, et al. The missing link between genetic association and regulatory function. *Elife* 2022;11. Available at:

http://dx.doi.org/10.7554/eLife.74970.

80. Mostafavi H, Spence PJ, Naqvi S, Jonathan PK. Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv* 2022. Available at: ;

81. Venken K, Jacques P, Mortier C, Labadia ME, Decruy T, Coudenys J, et al. RORγt inhibition selectively targets IL-17 producing iNKT and γδ-T cells enriched in Spondyloarthritis patients. *Nat Commun* 2019;10:9.

82. Yu H, Wu H, Zheng F, Zhu C, Yin L, Dai W, et al. Gene-regulatory network analysis of ankylosing spondylitis with a single-cell chromatin accessible assay. *Sci Rep* 2020;10:19411.

83. Simone D, Penkava F, Ridley A, Sansom S, Al-Mossawi MH, Bowness P. Single cell analysis of spondyloarthritis regulatory T cells identifies distinct synovial gene expression patterns and clonal fates. *Commun Biol* 2021;4:1395.

84. Alber S, Kumar S, Liu J, Huang Z-M, Paez D, Hong J, et al. Single Cell Transcriptome and Surface Epitope Analysis of Ankylosing Spondylitis Facilitates Disease Classification by Machine Learning. *Front Immunol* 2022;13:838636.

85. Ren C, Li M, Zheng Y, Cai B, Du W, Zhang H, et al. Single-cell RNA-seq reveals altered NK cell subsets and reduced levels of cytotoxic molecules in patients with ankylosing spondylitis. *J Cell Mol Med* 2022;26:1071–1082.

86. Yi K, Jo S, Song W, Lee H-I, Kim H-J, Kang J-H, et al. Analysis of Single-Cell Transcriptome and Surface Protein Expression in Ankylosing Spondylitis Identifies OX40-Positive and Glucocorticoid-Induced Tumor Necrosis Factor Receptor-Positive Pathogenic Th17 Cells. *Arthritis Rheumatol* 2023;75:1176–1186.

87. Pan-UKB team. Pan-UK Biobank. 2020. Available at: https://pan.ukbb.broadinstitute.org. Accessed 2023.

88. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015;47:1228–1235.

89. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 2005;310:321–324.

90. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010;467:832–838.

91. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;11:e1004219.

92. Yuhan H, Stephanie H, Erica A-N, Mauck WM, Shiwei Z, Andrew B, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–3587.e29.

93. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;16:1289–1296.

94. 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:e1004383.

95. Kramer NE, Davis ES, Wenger CD, Deoudes EM, Parker SM, Love MI, et al. Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics* 2022;38:2042–2045.

96. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet* 2019;51:768–769.

Table 1. Putative target genes identified by co-localization analysis between AS-associated loci and eQTLs in NK cells.

Lead_GWAS_Variant	P-value	GWAS	Putative_Target_Gene	Posterior probability of shared causal variant	Quantification_method	eQTL_study
rs27529	1.24E-40	IGAS	ERAP1	0.99186	microarray	Gilchrist et al. 2021
rs6759298	2.07E-38	IGAS	B3GNT2	0.97245	RNA-seq	Schmiedel et al. 2018
rs1128905	3.17E-10	IGAS	ENTR1	0.8207	microarray	Gilchrist et al. 2021
rs1860545	8.66E-10	IGAS	TNFRSF1A	0.99673	RNA-seq	Schmiedel et al. 2018
rs11065898	5.41E-08	IGAS	TMEM116	0.87563	microarray	Gilchrist et al. 2021
rs9619386	4.42E-07	IGAS	UBE2L3	0.96397	microarray	Gilchrist et al. 2021
rs1250542	2.07E-06	IGAS	ZMIZ1	0.87272	microarray	Gilchrist et al. 2021
rs1250542	2.07E-06	IGAS	ZMIZ1	0.83268	RNA-seq	Schmiedel et al. 2018
rs952594	2.08E-06	IGAS	APEH	0.75689	microarray	Gilchrist et al. 2021
rs952594	2.08E-06	IGAS	RBM6	0.91981	RNA-seq	Schmiedel et al. 2018
rs952594	2.08E-06	IGAS	UBA7	0.94237	microarray	Gilchrist et al. 2021
rs952594	2.08E-06	IGAS	UBA7	0.8851	RNA-seq	Schmiedel et al. 2018

rs6565217	2.82E-06	IGAS	AC135050.3	0.91556	RNA-seq	Schmiedel et al. 2018
rs6565217	2.82E-06	IGAS	STX4	0.94653	RNA-seq	Schmiedel et al. 2018
rs7191548	3.13E-06	IGAS	EIF3CL	0.7822	microarray	Gilchrist et al. 2021
rs7191548	3.13E-06	IGAS	NPIPB8	0.97545	microarray	Gilchrist et al. 2021
rs7191548	3.13E-06	IGAS	SGF29	0.85225	microarray	Gilchrist et al. 2021
rs7191548	3.13E-06	IGAS	TUFM	0.82246	microarray	Gilchrist et al. 2021
rs7191548	3.13E-06	IGAS	TUFM	0.95694	RNA-seq	Schmiedel et al. 2018
rs6583441	3.84E-06	IGAS	IKZF1	0.96244	RNA-seq	Schmiedel et al. 2018
rs4690326	6.49E-06	IGAS	DGKQ	0.91883	microarray	Gilchrist et al. 2021
rs4690326	6.49E-06	IGAS	DGKQ	0.89989	RNA-seq	Schmiedel et al. 2018
rs4690326	6.49E-06	IGAS	IDUA	0.98988	microarray	Gilchrist et al. 2021
rs4690326	6.49E-06	IGAS	SLC49A3	0.801	RNA-seq	Schmiedel et al. 2018
rs26481	9.17E-06	UK Biobank	CAST	0.95638	microarray	Gilchrist et al. 2021
rs26481	9.17E-06	UK Biobank	ERAP1	0.95565	microarray	Gilchrist et al. 2021
rs2236167	3.57E-05	IGAS	PPP2R3C	0.9491	microarray	Gilchrist et al. 2021

Figure legends

Figure 1. Human NK-specific open chromatin regions are enriched in AS genetic risk. (**A**) Cartoon depicting the Calderon et al. study design. Peripheral blood cells from 4 healthy subjects were sorted into immune cell populations that we grouped in silico into seven main cell types. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) was performed with and without prior in vitro activation. (**B**) Graphical representation of LDSC-SEG: identification of cell type-specific annotations (in our case open chromatin regions), followed by the integration with GWAS summary statistics to obtain a risk enrichment coefficient and p-value. (**C**) Volcano plots showing results of differential accessibility analyses for each cell type compared to the other cell types. Colored dots indicate open chromatin peaks in the top decile of the t-statistic for each

cell type, which were used for LDSC-SEG analysis. (**D-E**) Bar graphs display the AS genetic risk enrichment coefficient (y-axis) and standard error for cell type-specific open chromatin accounting for control peaks and baseline annotations. Summary statistics from the International Genetics of Ankylosing Spondylitis Consortium (IGAS) (**D**) and UK Biobank (**E**) GWAS were used. Bars marked with "*" indicate a P-value less than 0.05.

Supplementary Figure 1. Heritability enrichment results for control traits. (**A**) Bar graphs display the genetic risk enrichment coefficient (y-axis) and standard error for cell-type specific open chromatin accounting for control peaks and baseline annotations. Open chromatin data were taken from the Calderon et al. study. Risk enrichment was assessed using GWAS summary statistics for the positive control traits rheumatoid arthritis, Alzheimer's disease, systemic lupus erythematosus, and the negative control trait height. Bars marked with "*" indicate $P < 0.05$, "**" indicates $P < 0.01$, "***" indicates $P < 0.001$.

Figure 2. Within human lymphocyte populations, NK cells have enrichment of cell type-specific expression of AS-associated genes. (**A**) Cartoon depicting the Gutierrez-Arcelus et al. study. Peripheral blood cells from 6 healthy subjects were sorted into NK cells and six T cell populations: CD4+ T, CD8+ T, MAIT, iNKT, and two γδ T cell populations. Bulk RNA sequencing was performed on two replicates per sample. (**B**) Graphical representation of the SNPsea method illustrating the integration of gene expression profiles with risk loci obtained from GWAS. (**C**) Bar graphs showing -log10(P-value) for enrichment of cell type-specific expression of genes in AS risk loci using SNPsea. (**D**) Box and whisker plots showing the distribution of normalized expression values of AS-associated genes for each cell type. Data from NK cells colored in orange, data from T cells in purple. Tpm: transcripts per million.

Figure 3. Human gut single-cell atlas reveals significant upregulation of AS-associated genes in NK cells. (**A**) Schema depicting the generation of the Space-Time Gut Cell Atlas study with samples from fetal, pediatric and adult subjects. (**B**) Graphical representation of the scDRS method, which integrates GWAS risk genes with single cell data to identify disease-relevant cells. (**C**) Visualization of the Space-Time Gut Cell Atlas data using UMAP on the top 20 principal components from 1997 variable genes from the scRNA-seq expression matrix. (**D**) Same UMAP visualization as in C, where cells with significant scDRS score (20% FDR) are colored in red. (**E**) Bar graph showing enrichment of scDRS significant cells per cell type. (**F**) Bar graph showing the number of significant scDRS cells for each cell type using the fine-grained annotations from the Space-Time Gut Cell Atlas. Cell populations with at least 15 significant scDRS cells are shown. (**G**) Scaled average expression levels and percent of cells expressing a given gene for 50 genes associated with AS that had significant upregulation (5% FDR) in NK cells compared to the other cell types. Genes are sorted by multiplying their MAGMA score (strength of association with AS) by their average level of expression in NK cells.

Supplementary Figure 2. Single-cell disease relevant score results for control traits. (**A**) Visualization of the Space-Time Gut Cell Atlas using Uniform Manifold Approximation and Projection (UMAP) on the top 20 principal components from 1997 variable genes from the single-cell RNA-seq expression matrix. Cells are colored based on the coarse cell type annotations from the Space-Time Gut Cell Atlas. (**B**) Barplots shows the cell type proportions within the whole Space-Time Gut Cell Atlas and within cells with significant disease relevant score (20% FDR) for AS (using IGAS GWAS), Alzheimer's disease (AD) and rheumatoid arthritis (RA). (**C**) Same UMAP visualization as in A, where cells with significant scDRS score (20% FDR) are colored in red and non-significant cells are colored in gray, for each control trait.

Figure. 4. Co-localization of AS risk loci and NK eQTLs points to putative target genes for AS risk variants. (**A-D**) Manhattan plots showing AS GWAS and NK eQTL -log10(P-values) for SNPs within 500 kb of a lead GWAS SNP. Legend indicates the Linkage Disequilibrium (LD) between the lead GWAS SNP (purple diamond) and other SNPs in the region. Genes in the region are colored according to their posterior probability of the same causal variant being shared between two traits (PP4). (**A**) Manhattan plot identifying putative target gene *ERAP1* using AS IGAS GWAS and NK microarray gene expression QTL (eQTL) data obtained from Gilchrist et al. (**B**) Manhattan plot identifying putative target gene *TNFRSF1A* using AS IGAS GWAS and NK gene expression QTL (eQTL) data obtained from Schmiedel et al. (**C**) Manhattan plot identifying putative target gene *ENTR1* using AS IGAS GWAS and NK microarray gene expression QTL (eQTL) data obtained from Gilchrist et al. (**D**) Manhattan plot identifying putative target gene *B3GNT2* using AS IGAS GWAS and NK gene expression QTL (eQTL) data obtained from Schmiedel et al. All QTL summary statistics taken from eQTL Catalogue.

Figure 1

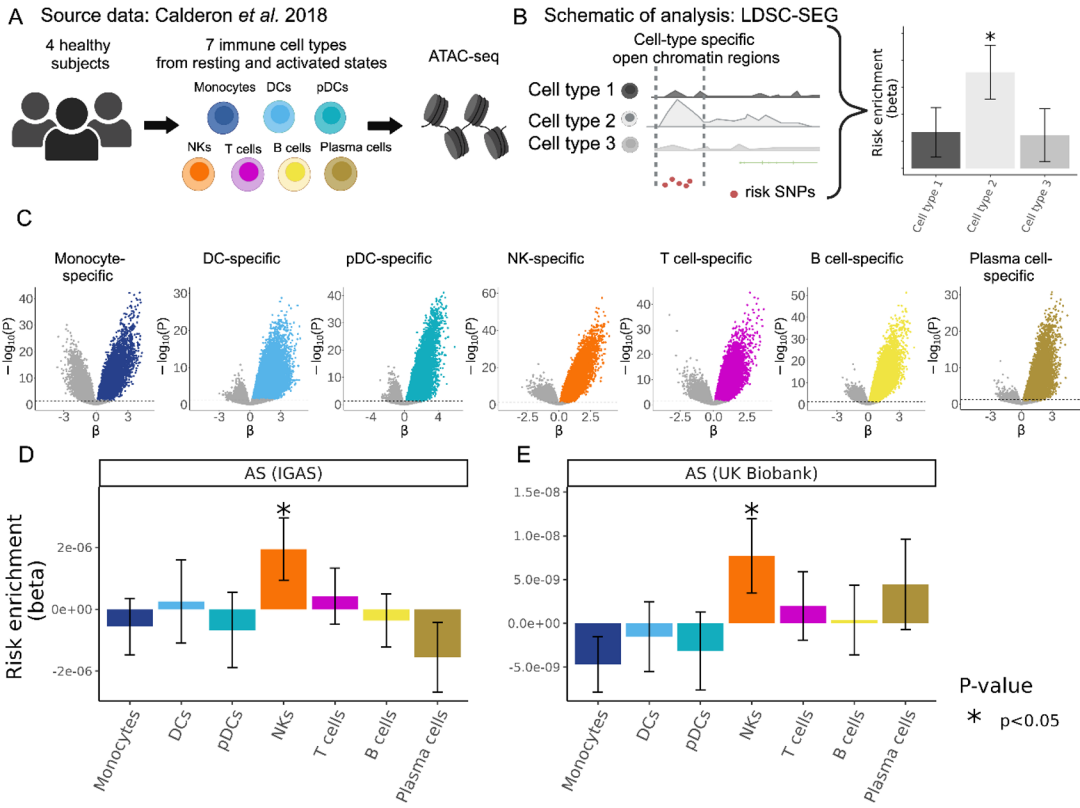

A
Source data: Calderon et al. 2018
4 healthy subjects
7 immune cell types from resting and activated states
Monocytes
DCs
pDCs
NKs
T cells
B cells
Plasma cells
ATAC-seq
B
Schematic of analysis: LDSC-SEG
Cell-type specific open chromatin regions
Cell type 1
Cell type 2
Cell type 3
risk SNPs
Risk enrichment (beta)
Cell type 1
Cell type 2
Cell type 3
C
Monocyte-specific
DC-specific
pDC-specific
NK-specific
T cell-specific
B cell-specific
Plasma cell-specific
− log10(P)
β
D
AS (IGAS)
Risk enrichment (beta)
Monocytes
DCs
pDCs
NKs
T cells
B cells
Plasma cells
E
AS (UK Biobank)
P-value
* p<0.05

Figure 2

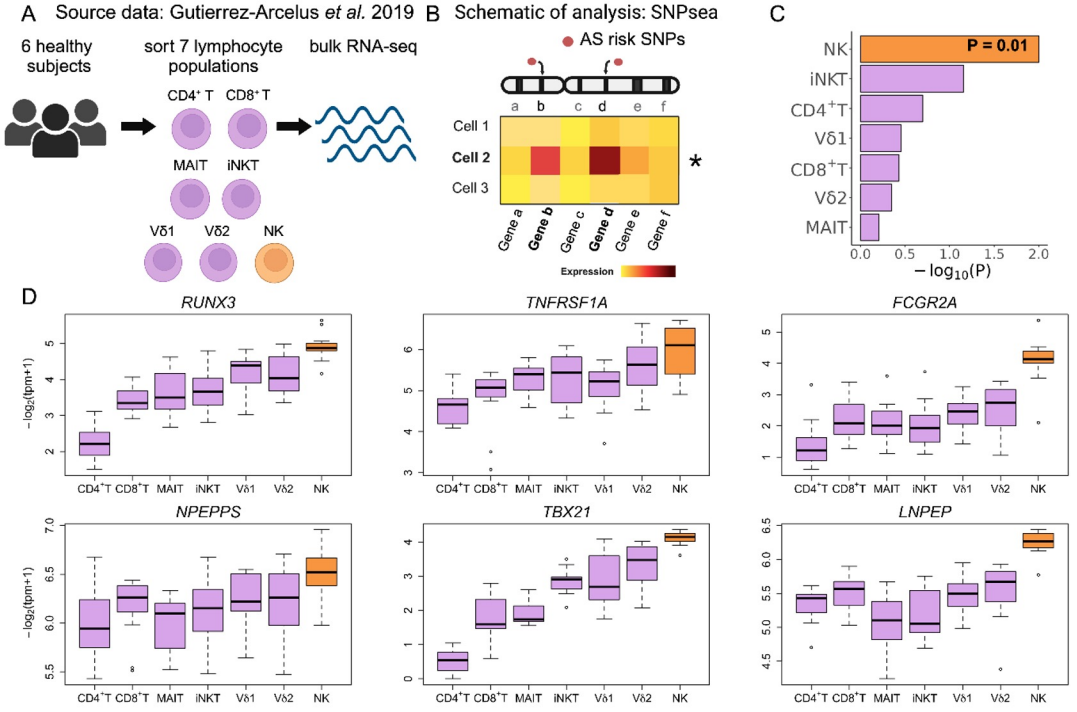

A Source data: Gutierrez-Arcelus et al. 2019
6 healthy subjects
sort 7 lymphocyte populations
bulk RNA-seq
CD4+ T
CD8+ T
MAIT
iNKT
Vδ1
Vδ2
NK
B Schematic of analysis: SNPsea
AS risk SNPs
a b c d e f
Cell 1
Cell 2
Cell 3
Gene a
Gene b
Gene c
Gene d
Gene e
Gene f
Expression
*
C
P = 0.01
NK
iNKT
CD4+T
Vδ1
CD8+T
Vδ2
MAIT
0.0 0.5 1.0 1.5 2.0
− log10(P)
D
RUNX3
TNFRSF1A
FCGR2A
NPEPPS
TBX21
LNPEP
−log2(tpm+1)
CD4+T CD8+T MAIT iNKT Vδ1 Vδ2 NK

Figure 3

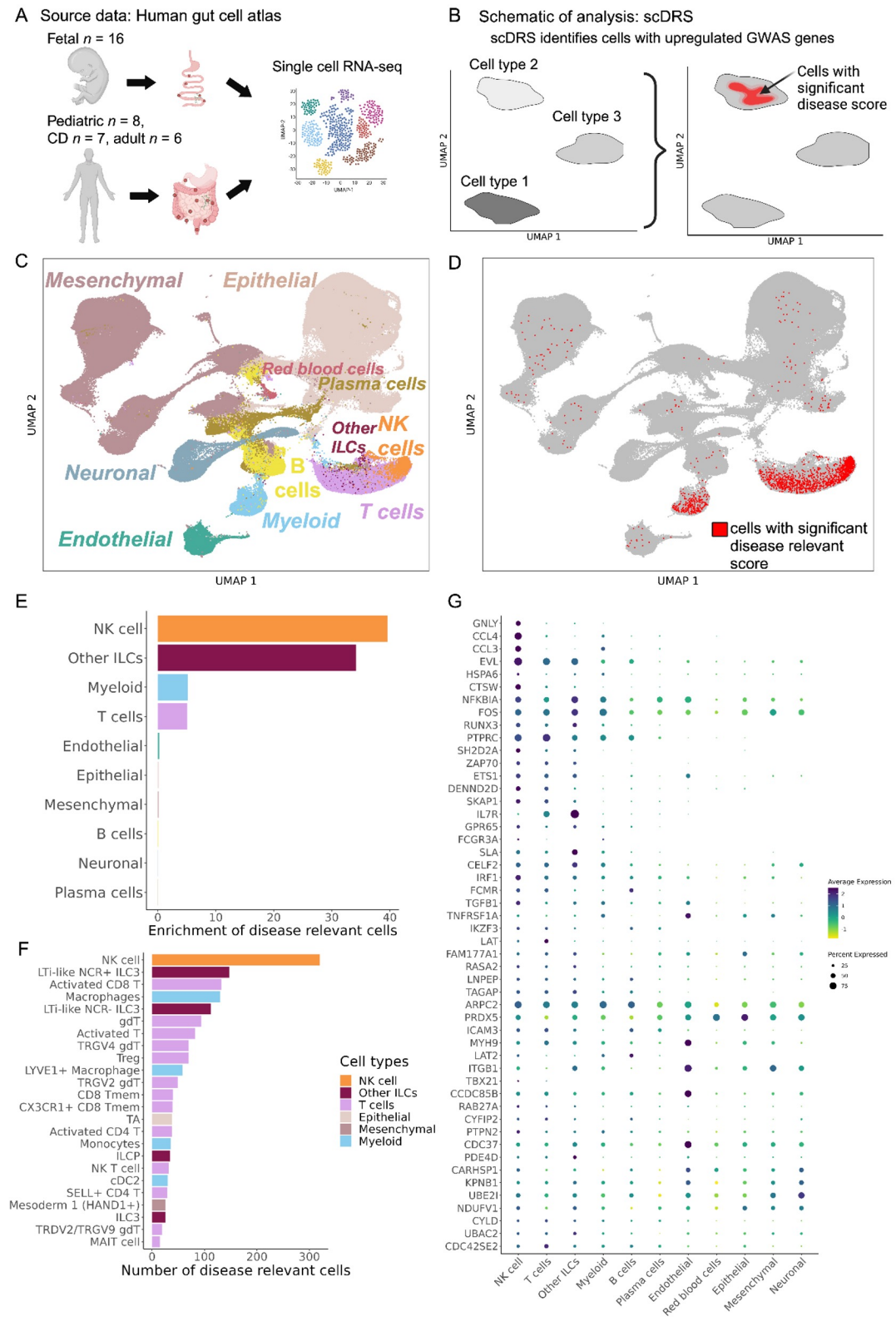

A
Source data: Human gut cell atlas
Fetal n = 16
Pediatric n = 8,
CD n = 7, adult n = 6
Single cell RNA-seq
B
Schematic of analysis: scDRS
scDRS identifies cells with upregulated GWAS genes
Cell type 2
Cell type 3
Cell type 1
UMAP 2
UMAP 1
Cells with significant disease score
C
Mesenchymal
Epithelial
Red blood cells
Plasma cells
Other ILCs
NK cells
Neuronal
B cells
Myeloid
T cells
Endothelial
D
cells with significant disease relevant score
E
NK cell
Other ILCs
Myeloid
T cells
Endothelial
Epithelial
Mesenchymal
B cells
Neuronal
Plasma cells
Enrichment of disease relevant cells
F
Cell types
NK cell
Other ILCs
T cells
Epithelial
Mesenchymal
Myeloid
Number of disease relevant cells
G
Average Expression
Percent Expressed
NK cell
T cells
Other ILCs
Myeloid
B cells
Plasma cells
Endothelial
Red blood cells
Epithelial
Mesenchymal
Neuronal

Figure 4

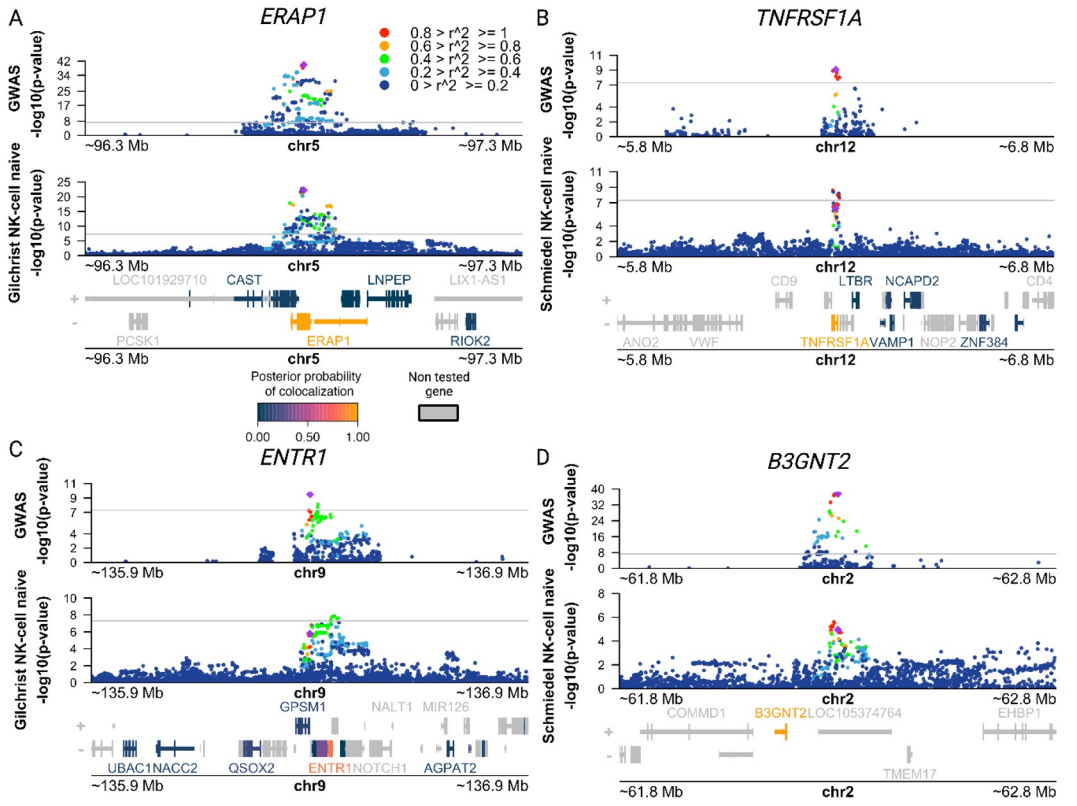

A
ERAP1
0.8 > r^2 >= 1
0.6 > r^2 >= 0.8
0.4 > r^2 >= 0.6
0.2 > r^2 >= 0.4
0 > r^2 >= 0.2
GWAS -log10(p-value)
Gilchrist NK-cell naive -log10(p-value)
~96.3 Mb
chr5
~97.3 Mb
LOC101929710
CAST
LNPEP
LIX1-AS1
PCSK1
ERAP1
RIOK2
Posterior probability of colocalization
0.00
0.50
1.00
Non tested gene
B
TNFRSF1A
Schmiedel NK-cell naive -log10(p-value)
~5.8 Mb
chr12
~6.8 Mb
CD9
LTBR
NCAPD2
CD4
ANO2
VWF
TNFRSF1A
VAMP1
NOP2
ZNF384
C
ENTR1
~135.9 Mb
chr9
~136.9 Mb
GPSM1
NALT1
MIR126
UBAC1
NACC2
QSOX2
ENTR1
NOTCH1
AGPAT2
D
B3GNT2
~61.8 Mb
chr2
~62.8 Mb
COMMD1
B3GNT2
LOC105374764
EHBP1
TMEM17

Figure S1

A

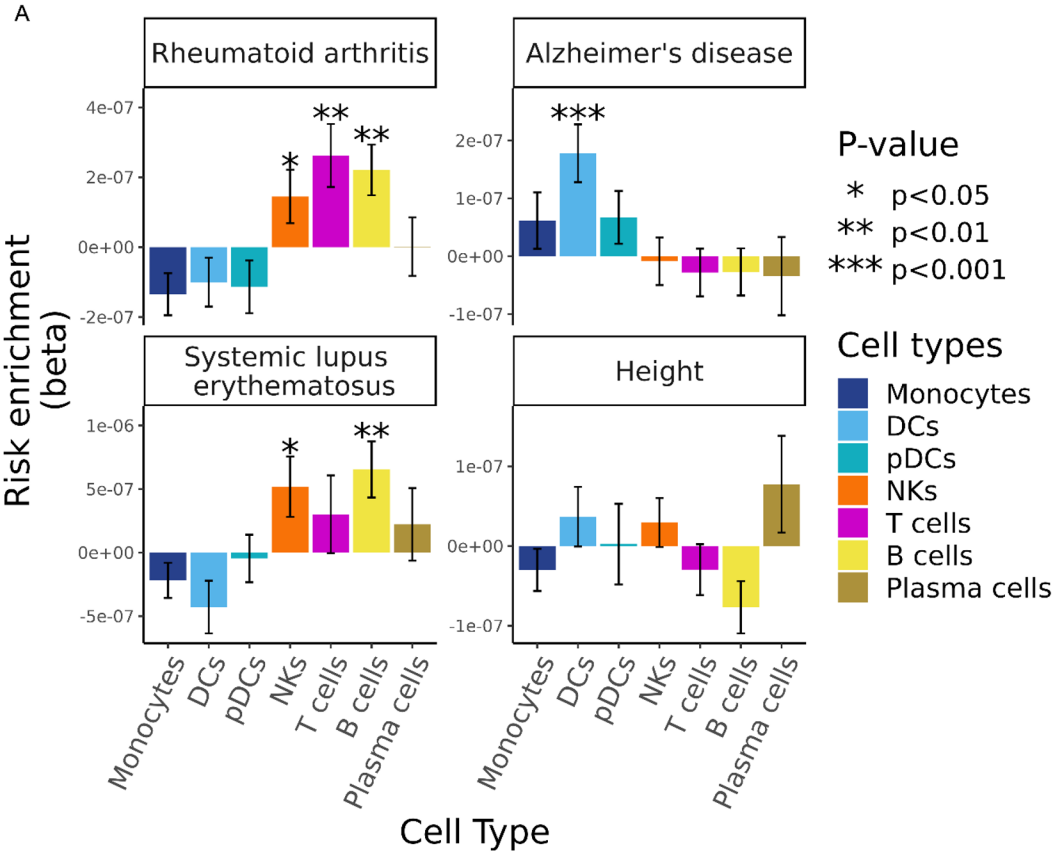
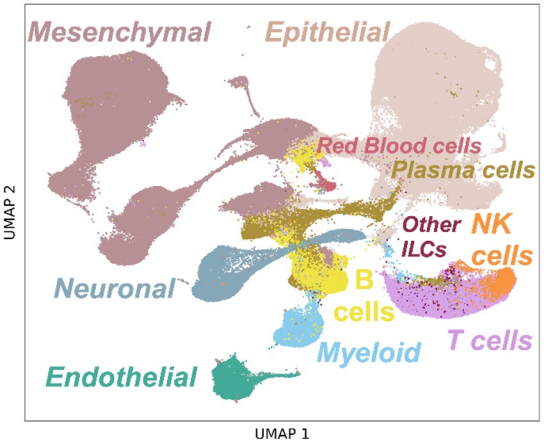
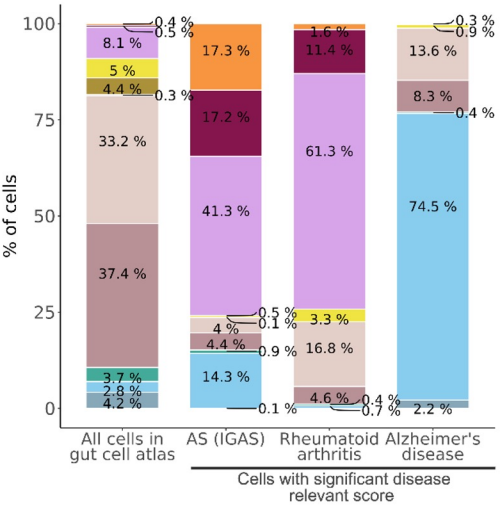

Rheumatoid arthritis
Alzheimer's disease
Systemic lupus erythematosus
Height
Risk enrichment (beta)
4e-07
2e-07
0e+00
-2e-07
1e-06
5e-07
-5e-07
1e-07
-1e-07
*
**
***
Monocytes
DCs
pDCs
NKs
T cells
B cells
Plasma cells
Cell Type
P-value
* p<0.05
** p<0.01
*** p<0.001
Cell types

Figure S2

A

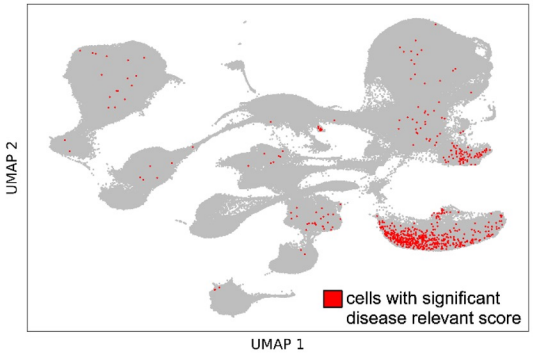


B

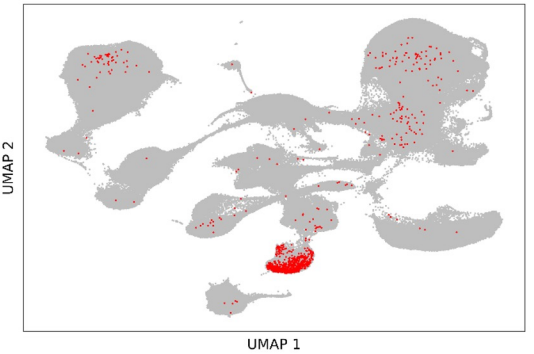


C

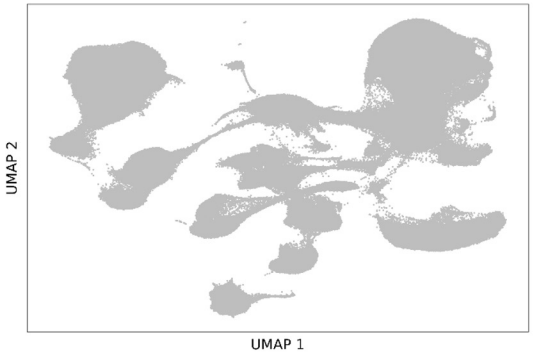
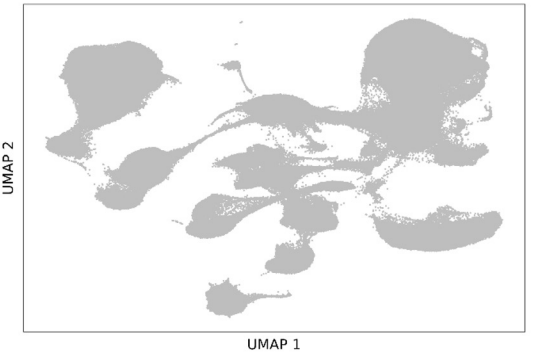
Rheumatoid arthritis



Alzheimer's disease



Systemic lupus erythematosus



Height