Single-cell genetics identifies cell type-specific causal mechanisms in complex traits and diseases

Albert Henry[1,2,3,4,*], Anne Senabouth[1,2,*], Rika Tyebally[1,2], Blake Bowen[1,2,3], Peter C. Allen[1,2], Eleanor Spenceley[1,2,3], Eszter Sagi-Zsigmond[1], Anna S.E. Cuomo[1,2,3,5,6], Jianan Fan[1,2], Hao Lawrence Huang[1,2,3], Hope A. Tanudisastro[2,5,6,7], Angli Xue[1,2,3], Katrina M. de Lange[1,2,3], Gemma A. Figtree[8,9], Alex W. Hewitt[10,11,12] , Daniel G. MacArthur[2,5,6], Joseph E. Powell[1,2,3,**]

[1] Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia.
[2] Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia
[3] UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia
[4] Institute of Cardiovascular Science, University College London, London, UK
[5] Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia
[6] Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia
[7] Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia
[8] Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia
[9] Kolling Institute of Medical Research, Royal North Shore Hospital, Sydney, NSW, Australia
[10] Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia
[11] Department of Ophthalmology, Royal Hobart Hospital, Hobart, TAS, Australia
[12] Centre for Eye Research Australia, University of Melbourne, East Melbourne, VIC, Australia

* Equal contribution
** Correspondence: j.powell@garvan.org.au

Abstract

Genome-wide association studies (GWAS) have been instrumental in uncovering the genetic basis of complex traits. When integrated with expression quantitative trait loci (eQTL) mapping, they can elucidate how risk loci influence traits through gene regulatory mechanisms. Recent single-cell eQTL (sc-eQTL) studies suggest that genetic effects on gene expression are often cell type- and subtype-specific, but such datasets have so far been underpowered for causal inference. Here, we leverage results from sc-eQTL mapping in the TenK10K project, comprising 154,932 common variant sc-eQTL across 28 immune cell types derived from matched whole-genome sequencing (WGS) and single-cell RNA-sequencing (scRNA-seq) of over 5 million peripheral blood mononuclear cells (PBMCs) from 1,925 individuals. We present a catalogue of cell type-specific causal effects of gene expression on 53 diseases (spanning 58,058 causal associations across 8,672 genes and 28 cell types), and 31 biomarker traits (spanning 681,480 causal associations across 16,085 genes and 28 cell types). By quantifying polygenic enrichment at both the single-cell and cell-type levels, we identify distinct immune cell contributions to both immune-related and systemic conditions. We demonstrate differential polygenic enrichment of Crohn's disease and COVID-19 amongst dendritic cell subtypes, and high activity of B cell interferon II response in SLE. Integration with clinical drug development data reveals that therapeutic compounds targeting gene-trait associations identified in this study are three times more likely to have secured regulatory approval. Using Crohn's disease as a motivating example, we demonstrate how population-based sc-eQTL data can pinpoint risk loci, effector genes and cell types, complementing findings from disease-focused tissue samples. Our findings provide a foundational resource for understanding the cell type-specific genetic architecture of disease and for guiding therapeutic discovery.

Introduction

Genome-wide association studies (GWAS) have identified thousands of loci associated with complex traits and diseases; yet, the majority of these signals reside in non-coding regions, complicating efforts to pinpoint causal genes and mechanisms[1]. Many non-coding variants act through gene regulatory mechanisms[2], influencing gene expression in a manner that is often highly cell type-specific[3–5]. This presents a significant challenge in pinpointing molecular mechanisms of diseases in heterogeneous tissues, where regulatory effects are often obscured in bulk transcriptomic analyses.

Single-cell genomics has transformed our understanding of gene regulation, revealing that the effects of genetic variation on gene expression can vary across different cell types and states. Consequently, single-cell expression quantitative trait loci (sc-eQTL) data have emerged as a promising avenue for identifying the cellular contexts in which genetic variants exert their effects. However, existing sc-eQTL studies have been underpowered[5], limiting their utility in causal inference of disease mechanisms.

Recent efforts to integrate GWAS data with single-cell transcriptomic atlases have highlighted disease-relevant cell types and programs. These approaches can be broadly categorised into two main types: those based solely on cell atlases[6] without consideration of genotypes, and those that incorporate genetic effects on single-cell molecular phenotypes[3,7–9]. Methods such as cell type enrichment scoring and topic modelling of gene expression patterns have linked regulatory programs to complex traits[10,11]. However, such approaches primarily establish correlation and do not directly assess whether genetically driven changes in gene expression causally contribute to disease risk. This distinction is particularly important in the immune system and complex tissues, where co-expression and linkage disequilibrium can confound signal interpretation.

In contrast, causal inference frameworks such as Mendelian randomisation (MR) test whether genetically-predicted gene expression has a directional effect on disease[12–14]. Owing to random allocation of genetic variants at conception and the fact that they remain fixed over the life course, sc-eQTL variants can be used as genetic instruments to predict gene expression. Under specific assumptions[12–14], these genetically-predicted gene expressions can be integrated with disease-specific genetic association estimates from GWAS through MR analysis, providing a platform to distinguish correlation from causation at cell type-specific and disease-specific resolution.

Here, we present a comprehensive framework for inferring causal gene–trait relationships by integrating sc-eQTL data from Phase 1 of the TenK10K project, which comprises matched whole-genome sequencing and single-cell RNA sequencing of over 5 million peripheral blood mononuclear cells across 28 immune cell types in 1,925 individuals. We applied MR using multiple sc-eQTL instruments to test the causal effects of 17,446 genes on 69 complex diseases

and 31 biomarker traits. This revealed gene–trait associations that are highly cell type-specific, including many not detectable by comparable methods using GWAS data only and bulk-tissue MR with whole blood eQTL instruments.

We further quantified polygenic enrichment across individual cells and cell types, identifying both immune-related and systemic disease associations, including cancer, neuropsychiatric, and cardiometabolic traits, with distinct cellular origins. We intersected these results with deep learning-derived cell function scores to prioritise biological processes involved in the pathogenesis of systemic lupus erythematosus (SLE). To explore the translational relevance of these findings, we intersected our causal gene–trait map with clinical drug development data from the Open Targets Platform[5]. We found that drugs whose targets matched our predicted gene–trait pairs were three times more likely to achieve regulatory approval.

Finally, we demonstrate how these findings can be used to identify effector genes and new therapeutic targets for Crohn's disease through triangulation of evidence from our study and an independent study measuring changes in gene expression using single-cell dataset from patients. Together, this study provides a population-scale, cell type-resolved map of causal effects of gene expression on complex traits, offering a valuable resource for advancing mechanistic understanding and therapeutic development.

Results

Landscape of causal associations between gene expression and complex traits across immune cells

We recently profiled the transcriptomes of 5,920,025 peripheral blood mononuclear cells (PBMCs) collected from 1,925 donors as a part of the first phase of the TenK10K project[15]. These cells encompassed 28 immune cell types and integration with matching whole genome sequencing (WGS) data revealed 154,932 common variant sc-eQTLs (minor allele frequency > 1%) for 17,674 eGenes (genes with at least one *cis*-eQTL). The majority of eGenes exhibited cell-type restricted effects, with only half of all identified eGenes being shared amongst at most a quarter of the cell types[15], highlighting the importance of cell-resolved approaches for linking regulatory variants to complex trait risk.

To quantify causal relationships between cell type-specific gene expression and complex traits, we applied a multi-instrument summary data-based Mendelian randomisation (MR) framework using sc-eQTL as proxies for gene expression[14] at the cell type level (**Figure 1a**). We compared this to an alternative method that prioritises trait-associated genes directly from GWAS data (hereafter referred to as GWAS genes) using gene-level aggregation of genetic association estimates with MAGMA[16]. Analysis was restricted to genes expressed in at least one cell type and with at least one significant *cis*-eQTL, for which corresponding variants were available across harmonised GWAS summary statistics (**Methods**).

We tested 17,446 eGenes (including 12,873 protein-coding genes) across 28 immune cell types, with a median of 5,948 eGenes (4,693 protein-coding) per cell type (**Figure 1b**, **Supplementary Table 1**). For each gene-cell type pair, we estimated causal effects on 69 complex diseases and 31 biomarker traits spanning 10 categories (**Figure 1c**, **Supplementary Table 2**). This yielded 15,092,351 gene–cell type–trait combinations, of which 739,538 (4.9%) passed the discovery threshold ($Q_{MR} < 0.05$ and heterogeneity in dependent instrument / HEIDI test $P_{HEIDI} > 0.01$). Across disease traits, we identified 58,058 unique causal associations from combinations of 53 diseases, 28 cell types, and 8,672 genes (6,801 protein-coding). The median number of unique genes with causal associations was 1,579 per cell type, 48 per disease, and 9 per cell type-disease pair (**Supplementary Figure 1, Supplementary Table 3**). Across biomarker traits, 681,480 causal associations were identified from combinations of 16,085 unique genes (12,078 protein-coding), 31 traits, and 28 cell types. The median number of unique genes with causal associations was 4,891 per cell type, 6,078 per disease, and 710 per cell type-disease pair (**Supplementary Figure 2, Supplementary Table 3**). Across the 28 immune cell types, the number of significant MR gene-trait associations ranged from 2,133 for ILC cells to 51,149 for NK cells (median = 29,186). This count scaled linearly with the total number of MR tests performed per cell type (**Supplementary Figure 3**), with a consistent proportion of significant results across cell types (mean = 5.1%, SD = 0.3%). These findings establish a high-resolution, cell-type-specific map of gene–trait relationships across the circulating immune system.

Trait-level insights from cell type-specific causal gene mapping

We identified 190,449 unique gene-trait associations in at least one cell type for 53 diseases and 31 biomarker traits. The number of significant MR-based gene associations per trait was strongly correlated with the number of GWAS-prioritised genes identified using MAGMA (Pearson's r = 0.97), with an approximately exponential relationship between the two (**Supplementary Figure 4**). Of the 190,449 unique gene-trait associations identified by single-cell MR, 65,457 (34%) were not identified by using MAGMA with GWAS data only, with a median of 44 (18%) per trait (**Supplementary Figure 5**, **Supplementary Table 4**). Importantly, MR-identified associations not detected by GWAS were more likely to be restricted to a small number of cell types (**Figure 2a**). A similar pattern was observed in comparison with bulk-tissue MR using genetic instruments derived from whole blood eQTL from eQTLgen[17], wherein 115,762 (61%) gene-trait associations were not identified (median = 209 / 64% per trait) (**Supplementary Figure 6**, **Supplementary Table 4**). These comparisons underscore the added value of eQTL mapping at single-cell resolution for identifying effector genes with cell type-specific roles in complex traits.

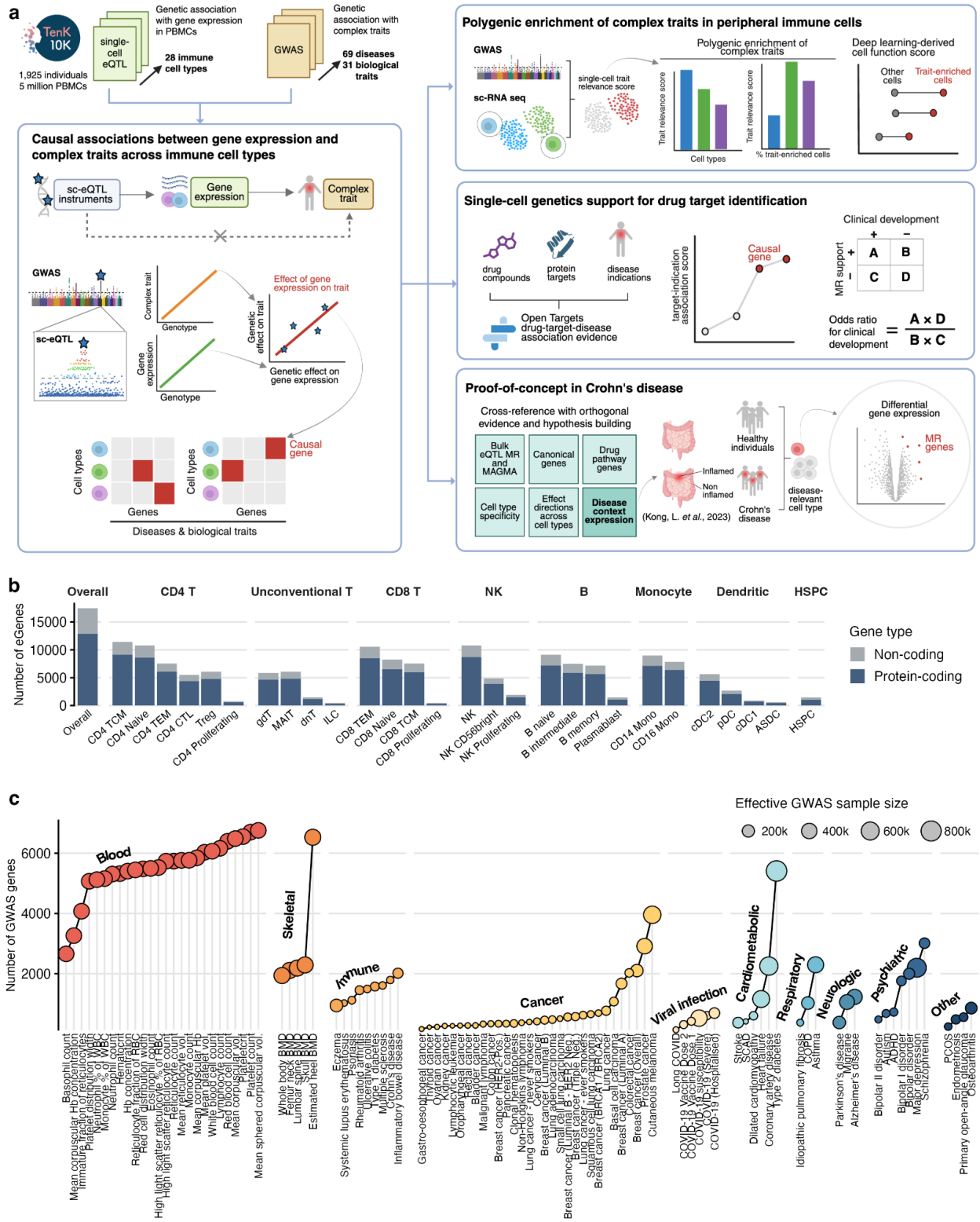


Figure 1 | Study design and causal inference using single-cell eQTLs.
a) Overview of study design. We integrated single-cell eQTL data from 5 million peripheral blood mononuclear cells (PBMCs) (covering 28 immune cell types) profiled in the TenK10K Phase 1

cohort (1,925 donors) with GWAS summary statistics for 100 complex traits (69 diseases and 31 biomarkers) using a multi-instrument Mendelian randomisation (MR) framework. **b**) Number of eGenes (genes with at least one significant cis-eQTL) included in the MR analysis per cell type, stratified by gene type (protein-coding vs non-coding). Median of 5,948 eGenes per cell type. **c**) Number of trait-associated genes identified using GWAS summary statistics alone via MAGMA (FDR < 5%) across the 100 traits. Each point represents a trait; the point size reflects the effective sample size of the GWAS, and the colour indicates the trait category (e.g., immune, cancer, cardiometabolic).

In addition to identifying causal genes, MR also helps in resolving the causal effect direction on trait outcomes, providing critical information to distinguish between targeting upregulation and downregulation of gene expression in drug design. Across all significant MR associations in our study, the direction of effect was evenly split between positive (*n*=369,155) and negative (*n*=370,383), indicating a balanced contribution of activating and repressing regulatory effects on trait outcomes (**Figure 2b**). This symmetry was consistent across both phenotype categories and cell types, implicating that disease-relevant gene networks involve a balance of activation and inhibition.

Across trait categories, traits closely related to peripheral blood composition yielded the highest number of associations, including 163,383 unique gene-trait associations across 26 blood biomarker traits. Substantial numbers of associations were also detected for cancer-related traits (5,375 associations across 20 traits), and immune-mediated diseases (2,943 associations across nine traits) (**Figure 2b**). Polygenic traits with large GWAS sample sizes also showed strong MR signals, including type 2 diabetes mellitus (5,608 associations), estimated heel bone mineral density (7,982 associations), and schizophrenia (1,033 associations) (**Figure 2c**, **Supplementary Table 3**). These associations were spread across immune cell types, but correlated with the number of eGenes identified from the sc-eQTL mapping (**Supplementary Figures 1-2**). Similar patterns of the relative number of MR genes across cell types were observed amongst top-associated traits in each category (**Figure 3c**). These findings imply that context-specific gene expression in PBMCs may play a causal role in complex traits beyond immune-related traits, or that these cells share a similar genetic regulation of gene expression with trait-relevant cell types in other tissues.

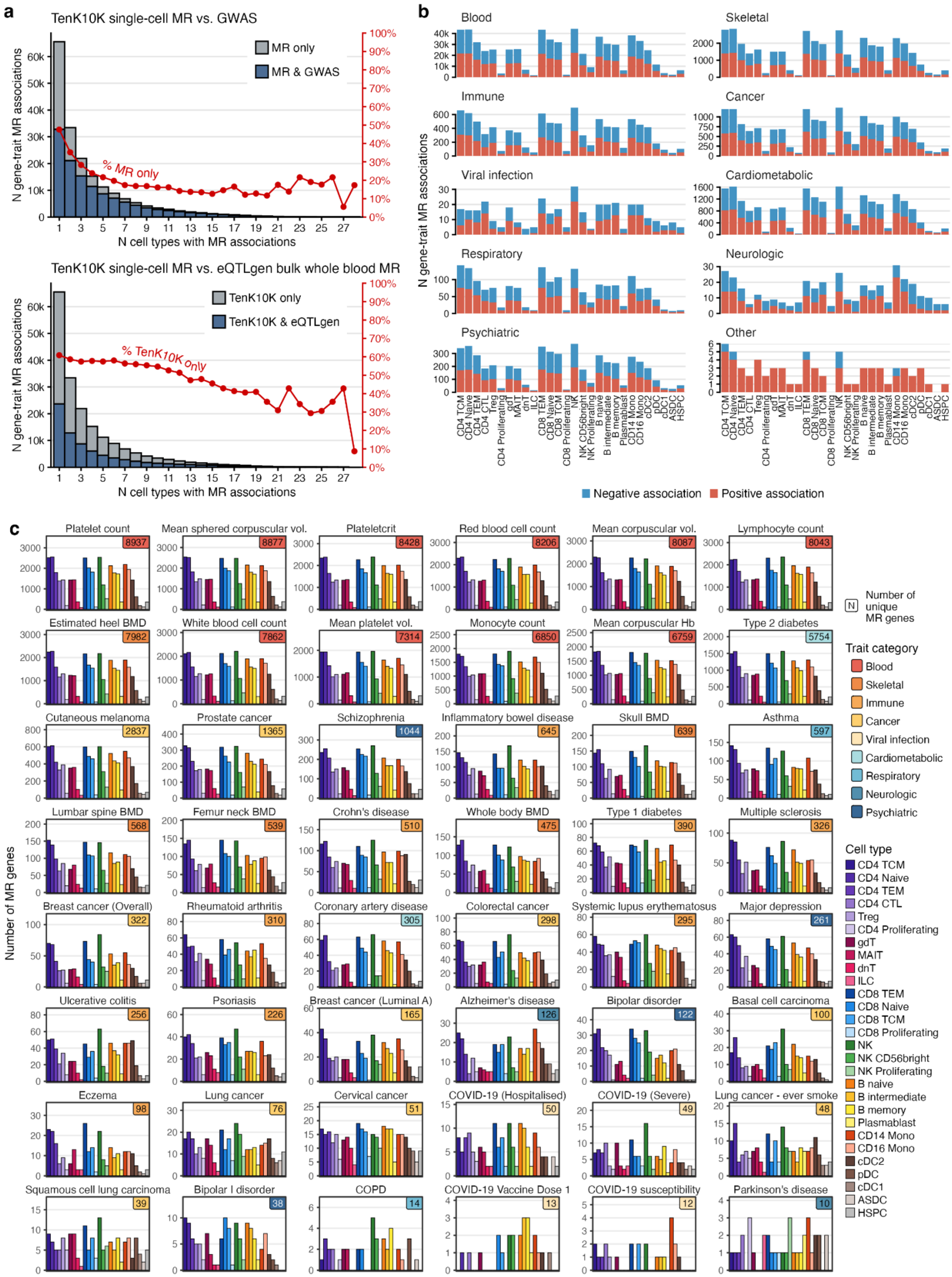


Figure 2 | Cell type-specific causal gene-trait associations across complex traits.

a) Distribution of gene-trait associations with MR evidence by the number of cell types in which the association is detected. Left y-axis: total number of gene–trait pairs (bars). Colours distinguish associations identified by MR only versus those also supported by GWAS ($Q_{MAGMA} < 0.05$). Right y-axis: proportion of MR-significant associations also supported by GWAS (red line and points). **b)** Number of significant gene-trait pairs identified by Mendelian randomisation (MR) across trait categories. Associations are defined as $Q_{MR} < 0.05$, $P_{HEIDI} \geq 0.01$. Bar color indicates the direction of association (positive or negative) with an increased gene expression. **c)** Same as panel b, but for comparison against results from MR using eQTL derived from bulk whole blood expression data from eQTLgen consortium. **d)** Bar charts showing number of genes with MR associations (MR genes) by cell types across selected 48 traits (out of 100 tested). The traits are displayed in subpanels within a grid (8 rows, 6 columns) and ordered row-wise in descending order by number of unique MR genes as displayed on the top-right corner of each trait subpanel, with a background coloured by trait category. Only traits with the top 10 highest number of unique MR genes per category and with 10 or more unique MR genes are displayed. The full results are presented in **Supplementary Figures 8-9**.

Polygenic enrichment reveals trait-relevant immune cell types and gene programs

To identify the immune cells involved in the genetic architecture of complex traits, we performed a polygenic enrichment analysis using the single-cell Disease Relevance Score (scDRS) framework[10]. We analysed 100 traits across the 28 immune cell types profiled in Phase 1 of the TenK10K study by computing cell-level scDRS score at cell-level, and statistics for polygenic enrichment at both cell-level and cell type-level using the default scDRS permutation scheme. To control for differences in cell abundance, we randomly sampled up to 10,000 cells per cell type, yielding a dataset of 229,239 cells.

We quantified the proportion of phenotype-enriched cells (cell-level $P_{scDRS} < 0.05$) within phenotype-enriched cell types (cell-type-level $P_{scDRS} < 0.05$) for each trait (**Figure 3a, Supplementary Table 5**). All 28 cell types showed enrichment for at least one phenotype (mean = 14.6 traits per cell type), with cDC1 (40 traits), cDC2 (37 traits), and CD4 proliferating T cell (33 traits) showing the strongest polygenic enrichment (**Supplementary Figure 7**). In total, 95 of 100 traits were enriched in at least one cell type (mean = 4.1, max = 10). Immune-mediated diseases showed the highest number of enriched cell types, as expected; for example, multiple sclerosis was enriched in 10 cell types. We also observed enrichment of non-immune polygenic traits, including estimated heel bone mineral density (9 cell types), Alzheimer's disease (7), and schizophrenia (7).

At single-cell resolution, the proportion of enriched cells per trait ranged from 8% and 79% (median = 18.9%) (**Figure 3a**). The strongest enrichment was observed for a negative antibody response to the first dose of the COVID-19 vaccine (no antibody detected post-vaccination), where 79% of cells were significantly associated with the trait, but predominantly enriched in antibody-producing B cells. As expected, monocyte count traits were enriched in CD14+ and CD16+ monocytes, and lymphocyte count traits in T cell subsets. Notable examples also included

enrichment of systemic lupus erythematosus in B cells; lung and breast cancers in proliferating cells, and coronary artery disease and Alzheimer's disease in monocytes (**Figure 3a**). These findings suggest that complex trait-associated variants, including those not traditionally linked to immune function, can affect gene expression patterns across diverse immune cell types.

Using Crohn's disease as a case study, we identified 10,741 out of 46,920 sampled cells (23%) as enriched for polygenic signals across seven cell types (**Figure 3b**). The most enriched cell types, ranked by Z_{scDRS} statistics, were cDC2, MAIT, cDC1, ILC, ASDC, CD4 TEM, and NK CD56^{bright} cells. Enrichment proportions ranged from 9.2% in CD4+ TEM to 54.3% in cDC2 (**Figure 3c**). Dendritic cells, particularly cDC2 and cDC1, showed the strongest enrichment. These results are consistent with prior reports of dendritic cell infiltration and activation in Crohn's-affected intestinal tissue[18], including overexpression of chemokine receptors involved in T cell recruitment. We also observed enrichment in ASDCs, a less-characterised dendritic subset with unclear roles in autoimmunity[19]. In contrast, plasmacytoid dendritic cells (pDCs) did not show enrichment for Crohn's disease but were the only dendritic subtype enriched for COVID-19 hospitalisation (**Figure 3d)**, consistent with their established role in antiviral immunity and type I interferon production[20].

The polygenic enrichment analysis also facilitates the identification of trait-related cell functions through integration with scDeepID, an in-house deep learning method to quantify cell function as cell states at single-cell resolution (see **Methods**). Using B cells in Systemic Lupus Erythematosus (SLE) as an example, we leveraged scDeepID to compute 400 representative cell function scores in 539,820 B cells (including B naive, B intermediate, and B memory cell types) identified in the TenK10K phase 1 study. We extracted cell function scores for the 30,000 sampled B cells used in the scDRS analysis, and compared the average cell function scores between 17,264 SLE-enriched cells (cell-level $P_{\text{scDRS-SLE}} < 0.05$) and 12,736 non-SLE-enriched cells using a one-sided *T*-test. We identified 202 cell functions with a higher average score in the SLE-enriched cells compared to non-SLE-enriched cells at $Q < 0.05$ (**Supplementary Figure 8, Supplementary Table 6**). The cell functions with the highest mean score difference based on Q value include cellular response to type II interferon, positive regulation of inflammatory response, tumour necrosis factor-mediated signalling pathway, biological process involved in interaction with host, and leukocyte-mediated cytotoxicity (**Supplementary Figure 9a-b)**. The scores of these five cell functions also showed a trend of linear increase across deciles of $\text{scDRS}_{\text{SLE}}$ (**Supplementary Figure 9c**), implicating involvement of these pathways and their gene members in the pathogenesis of SLE.

To assess the robustness of our sampling strategy, which oversampled rare cell types, we conducted a sensitivity analysis using proportional sampling (10% of cells per cell type) (**Methods**). Both the distribution of cell-level scDRS scores and the proportion of phenotype-enriched cells remained consistent (**Supplementary Figures 10-11**), confirming that enrichment patterns were not artefacts of oversampling.

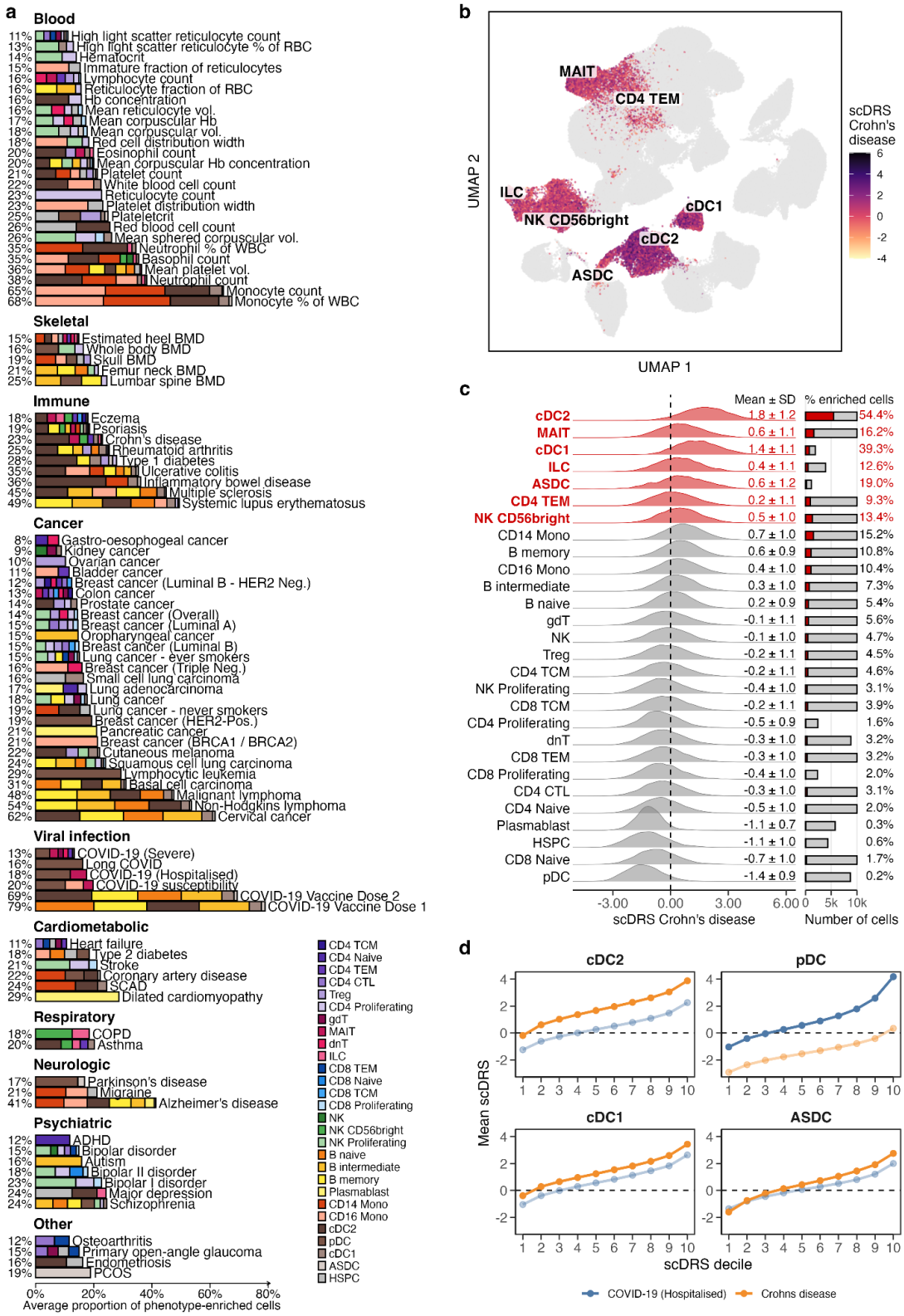


Figure 3 | Polygenic enrichment of complex traits across immune cell types.

a) Average proportion of phenotype-enriched cells across 100 complex traits, stratified by immune cell types. Enrichment was calculated using the single-cell disease relevance score (scDRS); cells were considered enriched if their scDRS $P < 0.05$ and belonged to cell types with significant trait enrichments (cell-type-level $P_{\text{scDRS}} < 0.05$). **b.** UMAP embedding of 229,236 sampled peripheral immune cells, coloured by scDRS score for Crohn's disease. Only cells from the seven cell types with significant enrichment (cell-type level $P_{\text{scDRS}} < 0.05$) are coloured; all other cells are shown in grey. **c.** Distribution of scDRS scores for Crohn's disease across the 28 immune cell types. Cell types are ordered by scDRS *Z*-score, and those with significant enrichment ($P_{\text{scDRS}} < 0.05$) are highlighted in red. **d.** Mean scDRS score across deciles of cell-level scDRS for COVID-19 and Crohn's disease across four dendritic cell subtypes (cell type-level $P_{\text{scDRS}} < 0.05$). Cell types without evidence of enrichment ($P_{\text{scDRS}} \geq 0.05$) are faded out, demonstrating contrasting enrichment patterns across diseases and dendritic cell subsets.

Single-cell MR supports drug target discovery and clinical translation

Recent studies have shown that drugs supported by human genetic evidence are more likely to advance through the drug development pipeline and gain regulatory approval[21,22]. Single-cell transcriptomics can further refine target validation by resolving cell type specificity within disease-relevant contexts[23]. Since most drug targets are proteins encoded by genes, the cell-resolved causal associations uncovered by our sc-eQTL MR framework may offer novel insights into tractable therapeutic mechanisms.

To evaluate this, we cross-referenced our MR results with curated drug development data from the Open Targets Platform[24] (version 25.06). First, we focused on 12,873 protein-coding genes tested in the MR analysis and 57 diseases with at least one matching indication available in the Open Targets, yielding 733,761 possible gene-indication pairs, hereafter the target-indication universe. Of these, 13,196 pairs (1.8%) were supported by MR evidence in at least one cell type (**Methods**). According to drug development activity data (from ChEMBL[25]), 6,712 target–indication pairs (involving 853 targets and 56 indications) have been investigated in clinical trials. Among these, 174 pairs (2.6%) were supported by our MR analysis. Notably, 473 of the 853 clinically tested targets (55%) had MR support for at least one indication in our dataset, including 151 targets that have received regulatory approval (**Figure 4**, **Supplementary Table 7**). MR-supported target–indication pairs were more likely to be observed across all stages of clinical development (**Figure 4b**, **Supplementary Table 8**), with approved indications showing the highest enrichment. Overall, MR-supported pairs were 3.3 times more likely to reach regulatory approval than unsupported pairs (odds ratio = 3.3; lower 95% CI: 2.41; $P = 7.1 \times 10^{-9}$, one-sided Fisher's exact test).

To further validate this association, we compared Open Targets evidence scores for drug–indication pairs with and without MR support. The Open Targets score is a normalized metric (0–1) derived from diverse evidence sources, including genetic associations, somatic mutations, expression data, animal models, and literature mining (**Methods**). Pairs supported by MR showed significantly higher evidence scores overall (mean = 0.37 vs. 0.26; $P = 2.3 \times 10^{-9}$; **Figure 4c, Supplementary Table 9**). Stratifying by evidence type, we found stronger support for MR-backed

336 pairs in both genetic evidence (mean = 0.52 vs. 0.42; $P = 7.7 \times 10^{-3}$) and non-genetic sources (mean
337 = 0.38 vs. 0.33; $P = 6.3 \times 10^{-3}$), with a more pronounced effect in the genetic category.
338
339 Next, we assessed the concordance between the direction of MR effect and the mechanism of
340 action (MoA) of drugs at the most advanced clinical development stage for each target–indication
341 pair (**Figure 4d, Supplementary Table 10**). On average, MR-supported pairs more often showed
342 concordant MoA, for instance, risk-increasing (positive effect) MR results aligning with inhibitors,
343 or protective (negative effect) results aligning with agonists. This pattern was especially
344 pronounced for drugs in Phase III trials or with regulatory approval. Moreover, MR-supported
345 targets in later development stages were associated with stronger evidence of causal effect (**Figure
346 4e, Supplementary Table 10**).
347
348 Finally, we examined selected examples from the 174 MR-supported target–indication pairs
349 (covering 131 genes and 25 diseases) in more detail (**Figure 4f, Supplementary Figure 12,
350 Supplementary Table 10**). These included several known immune targets of approved therapies:
351 JAK2 and TNF for Crohn's disease, IL6R for rheumatoid arthritis, and IL23A and IL17RA for
352 psoriasis. Importantly, our analysis also identified high-confidence targets in diseases not
353 classically associated with immune dysfunction, including APP (amyloid precursor protein) for
354 Alzheimer's disease and GLP1R (glucagon-like peptide-1 receptor) for type 2 diabetes.
355
356 APP encodes a precursor to Amyloid-β peptides, whose aggregation forms amyloid plaques linked
357 to synaptic dysfunction and cognitive decline in Alzheimer's disease[26]. GLP-1R, a receptor
358 primarily expressed in the pancreas and brain, modulates insulin secretion and glucose metabolism
359 through binding to GLP-1 proteins[27]. Initially developed for diabetes, GLP-1R agonists (e.g.,
360 semaglutide) have recently shown efficacy in reducing cardiovascular risk[28] and promoting weight
361 loss[29] in individuals without diabetes. These examples highlight how cell type-specific MR results
362 can identify therapeutically actionable targets across complex diseases including cases where the
363 encoding genes are primarily expressed in other tissues, potentially due to shared genetic
364 regulation.

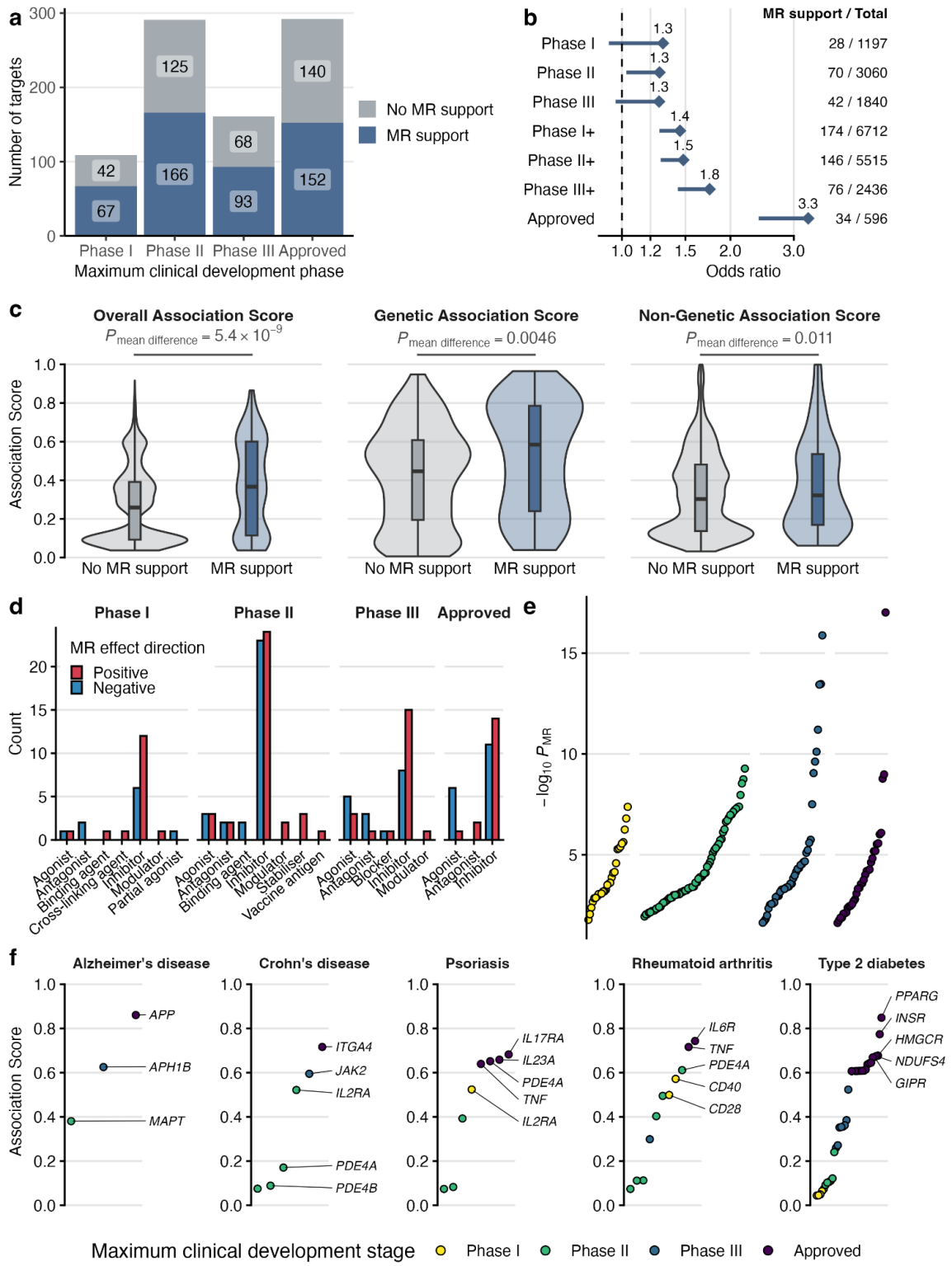


Figure 4 | Single-cell MR supports drug target validation and clinical development success. **a)** Number of protein-coding gene targets with MR support ($Q_{MR} < 0.05$; $P_{HEIDI} \geq 0.01$) across clinical development stages, as defined by target–indication pairs in the Open Targets Platform (OTP). Bars are grouped by the maximum development phase for each pair. **b)** Enrichment of MR-

supported target–indication pairs among clinically investigated indications. Odds ratios and lower 95% confidence intervals are shown for pairs with versus without MR support, stratified by maximum clinical development phase. Right-side annotations indicate counts of MR-supported target–indication pairs over the total number of clinically investigated pairs per phase. **c)** Distribution of OTP evidence scores for target–indication pairs with and without MR support. Scores are shown for all evidence sources (overall), as well as genetic and non-genetic sources separately. **d)** Number of MR-supported target–indication pairs, stratified by predicted direction of MR effect (positive or negative) and mechanism of action (MoA) of drugs with the highest OTP evidence score. Data are grouped by maximum clinical development stage. **e)** Strength of MR evidence ($-\log_{10} P_{MR}$) for target–indication pairs across clinical development stages. **f)** MR-supported targets for five selected diseases. Each bar represents a target–indication pair, ordered by OTP overall association score and coloured by maximum clinical development phase. The top five targets by OTP score are labelled.

Single-cell genetics reveals causal cell-type-specific mechanisms in Crohn's disease

To illustrate the added resolution of single-cell MR, we present a focused case study of Crohn's disease—a chronic inflammatory disorder of the gastrointestinal tract characterised by disrupted epithelial integrity, microbial dysbiosis, and dysregulated immune signalling[30,31]. Genetic studies have implicated immune-related genes and pathways involved in autophagy, cytokine signalling, and endoplasmic reticulum stress in disease pathogenesis[32].

Using our single-cell MR framework, we identified 1,791 gene-cell type pairs and 510 unique genes (407 protein coding) with evidence of a causal relationship with Crohn's disease across 28 immune cell types (**Supplementary Figure 13**). Of these, 88 were not detected using MAGMA-based gene prioritisation from GWAS data alone, and 339 were not identified using bulk whole blood-based MR (**Supplementary Figure 14**). A total of 209 genes (163 protein-coding) exhibited cell-type-specific MR associations, defined as significant in only one cell type. Notably, 146 (71%) of these were not detected in bulk-tissue whole blood MR, highlighting the unique resolution provided by single-cell approaches (**Supplementary Table 11**). Of the 1,791 gene-cell type pairs with MR evidence, between 66 to 1,076 pairs were also causally associated with other immune diseases (in descending order of shared signals: inflammatory bowel disease, ulcerative colitis, psoriasis, type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, and eczema), potentially revealing shared causal molecular pathways. Of the 1,791 gene-cell type pairs with MR evidence, 232 (13%) demonstrate discordant directions of effect between diseases, highlighting disease context-dependent causal effects (**Supplementary Table 12**).

As a proof-of-principle, we recovered known therapeutic targets for Crohn's disease, including *TNF*, *IL23R*, and canonical risk genes such as *ATG16L*, *NOD2*, and *IRGM* (**Supplementary Table 13**, **Figure 5a**). Interestingly, we observed divergent directions of causal effect for 61 genes (20%), where the same gene exhibited opposing MR directions in different cell types. For example, *NCF4* (which encodes a component of an enzyme complex in phagocytes), showed a positive effect in B memory cells but a negative effect in B naive cells (**Figure 5b**), suggesting context-specific

412 regulatory roles. The negative effect of *NCF4* in B naive cells is also observed in eczema.
413 Interestingly, while the positive effect of *NCF4* in B memory cells is also observed in psoriasis, it
414 has a negative effect in multiple sclerosis (**Supplementary Table 12**).
415
416 To validate the biological relevance of our findings, we compared MR-identified genes with
417 differentially expressed genes from single-cell RNA-seq of colon biopsies from 46 patients with
418 Crohn's disease and 25 healthy controls[33]. While some intestinal immune cells are tissue-resident,
419 many cell type lineages, such as B cells, T cells, dendritic cells, and NK cells, are shared between
420 peripheral and tissue compartments. We harmonised cells between datasets by major cell types,
421 excluding tissue-specific subsets such as macrophages (**Supplementary Figure 15**), and
422 examined the consistency of MR directionality with differential expression.
423
424 We reasoned that genes with a positive MR effect (*i.e.*, risk-increasing) would be upregulated in
425 disease tissue, whereas those with a negative MR effect (protective) would be downregulated. We
426 identified a total of 116 genes with matched major cell types. Of these, 62 genes had concordant
427 direction of effect between MR and DEG, 47 with discordant effects and 7 with both concordant
428 and discordant effects, depending on concordance between and within major cell types
429 (**Supplementary Table 14, Supplementary Figure 16**). We highlight *ZBTB38* as an interesting
430 example with convergent : it showed a positive MR effect and was upregulated in B cells, CD4+
431 T cells, and CD8+ T cells from Crohn's disease colon samples (**Figure 5c**). Conversely,
432 *TNFRSF18* (a known immune modulator) showed a protective MR effect but was also upregulated
433 in CD4+ T cells from Crohn's disease colon, suggesting potential compensatory mechanisms due
434 to reverse causation (**Supplementary Table 14, Supplementary Figure 16**).
435
436 Together, these results demonstrate how single-cell MR can identify both established and
437 uncharacterised causal genes with cell-type-specific effects, refine interpretation of differential
438 expression studies, and inform mechanistic hypotheses for complex disease biology.

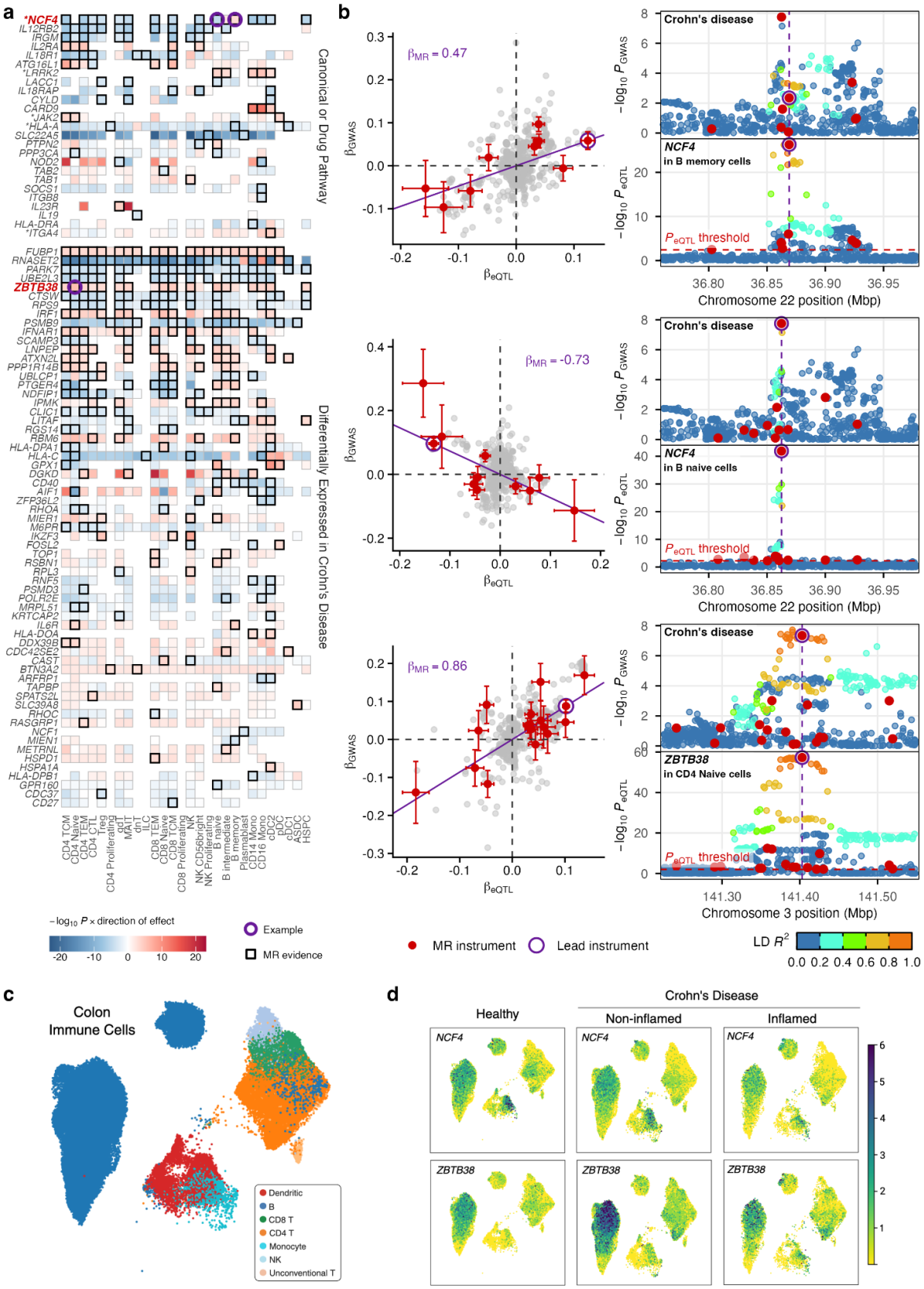


Figure 5 | Cell-type-specific causal and transcriptional signatures in Crohn's disease.

a) Selected MR-significant genes in Crohn's disease, shown as $-\log_{10} P_{MR}$ multiplied by direction of MR estimate of the lead instrument. Only protein coding genes are shown. A positive direction denotes increased disease risk with increased expression, and negative direction denotes reduced disease risk for increased expression. Gene categories include canonical risk genes, approved drug targets and their pathways (combined), and differentially expressed genes in matched major cell types of colon immune cells of Crohn's disease patients (DEGs) with concordant direction of effect (See **Methods** and **Supplementary Table 12**). Genes denoted with an asterisk are both canonical or drug target genes and DEGs. DEGs were identified from colon biopsy scRNA-seq data (Kong et al., 2023)[33] by matching genes and major immune cell types with MR results. Empty grids indicate the gene was not an eGene in that cell type and was therefore not tested. Black outlines denote MR associations ($Q_{MR} < 0.05$; $P_{HEIDI} > 0.01$). **b)** Causal effect plots showing GWAS effect sizes (*y*-axis) against cell type-specific sc-eQTL SNP effect sizes (*x*-axis), paired with locus zoom plots for selected genes. Data points in both plots represent common genetic variants in the *cis*-region of the gene that are available in both GWAS and sc-eQTL datasets. Variants used as instruments for gene expression in the MR analysis are highlighted in red, and lead instruments (lowest sc-eQTL *P*) are circled in purple. The slope in the effect plot corresponds to the point estimate of the MR effect size (coefficient) of the lead instrument. Row 1 and 2, respectively: positive causal effect of *NCF4* in B memory cells and negative causal effect of *NCF4* in B naive cells; Row 3: positive causal effect of *ZBTB38* in CD4+ naive cells. **c)** Colon immune cells from Kong et al., 2023, colour-coded by major cell type annotations derived by mapping major cell types with the TenK10K study (see **Methods, Supplementary Table 16**). Unmatched major cell types are excluded (original annotations of the full datasets are presented in **Supplementary Figure 16**). **d)** Expression of *NCF4* and *ZBTB38*, which both have causal effects on Crohn's disease in circulating immune cells and differential expression in colon immune cells. Colon immune cells are projected on the UMAP coordinates from panel c, stratified by disease and inflammation status.

Discussion

By integrating population-scale sc-eQTL data from 1,925 individuals in the TenK10K cohort with genome-wide association study (GWAS) results, we generated a comprehensive map of cell type-specific causal gene–trait relationships across 28 peripheral immune cell types and 100 complex traits. Using a multi-instrument Mendelian randomisation (MR) framework, we identified over 8,000 genes with evidence of causal association with at least one disease, and over 16,000 with at least one biomarker. In total, this resulted in 190,343 unique gene–trait associations, spanning 739,538 gene–trait–cell type combinations. These findings align with the postulated genetic architecture of complex traits, where trait-associated variants are numerous, context-specific, and distributed across diverse regulatory programs[34,35]. Given that genes and variants exhibit pleiotropic effects[36], and that diseases may share causal molecular aetiology[37], this cell type-resolved causal association map provides a platform for prioritising fundamental mechanisms underlying complex traits.

Approximately one-third (34%) of gene–trait associations identified through MR were not recovered using conventional gene-based GWAS approaches (*e.g.*, MAGMA), and the majority

485 of these were specific to a single or small number of cell types. This underscores the value of cell-
486 resolved sc-eQTL data for uncovering regulatory effects that are masked in bulk tissue analyses.
487 Notably, we found that 61% of the gene-trait associations identified in our study were undetectable
488 in a comparable MR analysis using genetic instruments derived from bulk-tissue eQTL mapping
489 of whole blood sample from eQTLgen[17], despite having 16 times larger sample size for eQTL
490 discovery. These findings affirm the hypothesis that cell-specific genetic regulations are
491 undetectable in aggregated datasets, further reinforcing the utility of single-cell resolved eQTL
492 mapping.
493
494 While our analysis focused on PBMCs, we identified causal associations not only for immune-
495 related diseases and blood traits, but also for systemic traits such as type 2 diabetes, schizophrenia,
496 cancer, and bone mineral density. These results potentially suggest that peripheral immune cell
497 dysregulation is directly involved in the pathogenesis of systemic diseases, or that they may share
498 cell-level genetic regulation with other cell types in disease-relevant tissues. These two hypotheses
499 can be observed in the polygenic enrichment patterns observed in the single-cell disease relevance
500 score (scDRS) analysis, which shows that polygenic signatures of systemic disease were indeed
501 observed across PBMC cells, but to a lesser extent compared to traits directly related to immune
502 dysregulation and peripheral blood composition. Given that nearly half of cis-eQTLs are shared
503 across tissues[20], these findings implicate that large-scale sc-eQTL maps may provide some insights
504 into causal biology of systemic diseases in other tissues with similar expression of disease-relevant
505 genes or with shared genetic regulation.
506
507 At the cell type level, we found striking enrichment of immune-mediated disease associations in
508 dendritic cell (DC) populations. In Crohn's disease, for example, multiple DC subsets (cDC1,
509 cDC2, ASDC) showed strong polygenic enrichment and MR-supported causal associations,
510 consistent with prior evidence that DCs accumulate and become activated in inflamed intestinal
511 tissue[9]. In contrast, COVID-19 genetic risk was specifically enriched in plasmacytoid DCs (pDCs),
512 aligning with their established role in antiviral immunity and rapid interferon production[11].
513 Focusing on B cells in the context of SLE, we found SLE-enriched cells display significantly a
514 higher cell function score for cellular response to type II interferon, a B cell-intrinsic mechanism
515 implicated in SLE pathogenesis via autoimmune response due to dysregulated germinal centre[38].
516 These findings suggest that large-scale sc-eQTL data, when integrated with GWAS and functional
517 annotation of relevant biological context, can resolve the cell type-specific contributions of
518 immune populations to disease even in the absence of disease-context tissue sampling.
519
520 Translationally, our data show that causal gene–trait relationships derived from single-cell MR are
521 predictive of clinical development success. By intersecting MR results with drug target and trial
522 data from the Open Targets Platform, we found that target–indication pairs supported by single-
523 cell MR were 3.3 times more likely to reach regulatory approval. These pairs also showed higher
524 evidence scores across both genetic and non-genetic data sources and were more likely to display

525 concordant direction of effect with the mechanism of action of approved or late-phase drugs. Our
526 framework successfully recapitulated known therapeutic targets, including *JAK2* and *TNF* for
527 Crohn's disease, *IL6R* for rheumatoid arthritis, *APP* for Alzheimer's disease, and *GLP1R* for type
528 2 diabetes, while nominating cell type-specific genes with strong genetic support for new
529 indications. In addition, our data show more than half of targets tested in clinical trials (including
530 151 that are already approved) had single-cell MR support for at least one tested indication,
531 implicating opportunities for drug repurposing.
532
533 To highlight the utility of cell-type-resolved MR for identifying both canonical and novel
534 therapeutic targets, we further explored 510 genes associated with Crohn's disease as a case study.
535 We identified 510 MR-significant genes, including those with well-established roles such as
536 *NCF4*, *ATG16L1*, and *IL23R*, as well as less characterised candidates such as *ZBTB38*. The
537 transcriptional repressor *ZBTB38* exhibited a positive causal effect in T and B cell subsets and was
538 independently upregulated in disease-affected colon tissue. *ZBTB38* has been linked to Crohn's
539 disease[39] and other inflammatory diseases[40,41], however its functional role in T and B cells has not
540 been explored. Our findings here offer a testable hypothesis that *ZBTB38* may modulate cell type-
541 specific programs relevant to disease pathogenesis, supported by differential expression in disease.
542 A second example, *NCF4,* is a known risk gene that encodes a subcomponent of the NADPH
543 complex, playing a critical role in oxidative stress and antimicrobial function[42]. In Crohn's disease,
544 *NCF4* exhibited a positive effect in B memory cells, in contrast to a negative effect in other cell
545 types including B naive cells. Interestingly, independent eQTL data show that the top eSNP in B
546 memory cells, which has a uniquely positive effect, is a fine-mapped SNP with a negative eQTL
547 effect in monocytes, macrophages, neutrophils, but a positive splicing QTL (sQTL) effect in B-
548 cell derived lymphoblastoid cell lines[24,43]. Many eQTL effects are explained by splicing QTLs[44,45],
549 and these results highlight the need for long-read sequencing to resolve cell type specific sQTLs
550 and explain novel causal associations. In this case study, we form links between MR results within
551 diseases as well as independent, orthogonal sources of evidence such as differential gene
552 expression[46].
553
554 We also identified MR-supported associations that may inform mechanisms of action for existing
555 therapies. For example, for Crohn's disease, we found a positive causal association of *IL23R* in γδ
556 T cells and a negative (protective) association of *IL12Rβ2* in multiple immune cell types, while
557 *IL12Rβ1* showed no significant effect. These observations are consistent with clinical evidence
558 that ustekinumab (targeting the shared *IL12*/*IL23* p40 subunit, blocking the *IL23R*/*IL12Rβ1*
559 receptor complex) acts predominantly through *IL23* blockade, and may help explain the increased
560 efficacy of *IL23*-specific therapies like guselkumab[47–49]. Our data suggest that targeted modulation
561 of *IL23R* and sparing of *IL12RB2* may optimise therapeutic efficacy while reducing off-target
562 effects.
563

564 Our study has several limitations. First, MR estimates are subject to bias from overestimation in
565 sc-eQTL discovery (winner's curse) and residual pleiotropy not fully accounted for by the HEIDI
566 test. Second, PBMCs may not capture causal regulation in disease-relevant tissues or tissue-
567 resident immune compartments. Third, the statistical power of our study may be limited for rare
568 cell types genes with low expression. While we downsampled abundant cell types to reduce bias
569 in the polygenic enrichment analysis, sensitivity analyses suggest this had minimal impact.
570 Furthermore, OpenTargets may not capture all active drug development programs, and many
571 apparently negative results may reflect the absence of evidence rather than evidence of absence.
572 Depending on the particular disease, signals may not be detected if their key causal effects are
573 mediated in non-peripheral immune cells, or tissue-specific cells such as epithelial or stromal cells.
574 Finally, functional validation of hypotheses is essential to confirm the causal relevance of newly
575 implicated genes and their therapeutic potential.
576
577 In summary, we present a large-scale, cell type-resolved atlas of causal gene–trait associations by
578 integrating single-cell eQTL mapping with GWAS through Mendelian randomisation. Our results
579 highlight extensive and cell-specific polygenic contributions to complex traits, identify new
580 regulatory mechanisms in disease, and demonstrate that MR-informed cell type-specific targets
581 are predictive of drug development success. This resource provides a powerful foundation for
582 understanding the molecular underpinnings of complex disease and for guiding precision
583 therapeutic strategies.
584
585 **Methods**
586
587 *Data sources*
588 *Single-cell eQTL*
589 To perform single-cell MR analysis, we used genetic instruments derived from a large-scale single-
590 cell expression quantitative trait loci (sc-eQTL) association analysis from Phase 1 of the TenK10K
591 project[15]. This dataset comprises matched whole-genome sequencing (WGS) and single-cell RNA-
592 sequencing (scRNA-seq) data from 1,925 individuals of European ancestry from Tasmanian
593 Ophthalmic Biobank (TOB) and the BioHEART study cohort, covering over 5 million peripheral
594 blood mononuclear cells (PBMCs) across 28 immune cell types. The cell types were identified
595 using a combined approach that uses scPred[50] and sc-HPL[51] with Azimuth PBMC reference
596 dataset[52].
597
598 The sc-eQTL association test was performed separately in each of the 28 cell types using a Poisson
599 mixed model for single-cell counts implemented in SAIGE-QTL[53] with the following covariates:
600 sex and age of the individuals, 7 genotype PCs, 5 cell type-specific expression PCs, cohort status
601 (TOB or BioHEART), percentage of counts from mitochondrial genes and total counts. For each
602 gene, common genetic variants (minor allele frequency >1%) within *cis*-region of the gene, defined
603 as 100 kbp window upstream and downstream site of the tested eGene bodies based on ENCODE

604 v44 hg38 coordinates, were tested for association with expression level of the gene measured in
605 raw counts. A total of 12,874,224 variants were included in the analysis, which yielded 154,932
606 common variant sc-eQTL across 17,674 unique eGenes across cell types at FDR 5%.
607
608 *Genome-wide association study of complex traits*
609 To perform MR, MAGMA and scDRS, analyses, we used publicly accessible genome-wide
610 association results of 100 complex diseases and biological traits across 10 phenotype categories:
611 Blood (26 traits)[54], Skeletal traits (5 traits)[55], Cancer (31 diseases)[56–65], Immune (9 diseases)[66–72],
612 Viral infection (6 diseases)[73–75], Cardiometabolic (6 diseases)[76–82], Neurologic (3 diseases)[83–85],
613 Psychiatric (7 diseases)[86–90], Respiratory (3 diseases)[91–93], and Other (4 diseases)[94–97]. For each
614 trait, we obtained GWAS summary statistics from public repositories.
615
616 Summary statistics were harmonised to the TenK10K study by matching variants' genomic
617 coordinates and alleles, performing conversion to hg38 with liftOver and harmonisation of minor
618 and major alleles based on allele frequency in individual-level genotype from the TenK10K study.
619 To ensure consistency of variants' allele frequency and linkage disequilibrium pattern with
620 TenK10K sc-eQTL dataset, analysis was limited to GWAS performed in European ancestries. If
621 unreported in the GWAS summary statistics, values such as the beta value as log(OR), standard
622 error and p-values were derived from other reported values. The effective GWAS sample size for
623 binary disease traits presented in **Figure 1c** was calculated as 4 / (1/number of cases + 1/number
624 of controls). More details of each trait, GWAS data source, reference, and number of samples are
625 provided in **Supplementary Table 2**.
626
627 *Mendelian randomisation analysis*
628 To identify genes associated with disease and biomarker traits in PBMC cell types, we performed
629 a multi-instrument Mendelian randomisation (MR) analysis using the SMR approach as
630 implemented in the SMR v1.3.1 software package[13]. Genetic instruments for MR were selected
631 from common variants within the *cis*-region of the tested gene bodies passing cell type-specific
632 thresholds. The cell type-specific thresholds were derived from the maximum *P*-value across sc-
633 eQTL for eGenes identified at FDR <5% in the TenK10K phase I study[15]. Following the default
634 parameter of the SMR software[14], variants that are not available in the GWAS summary statistics
635 of tested phenotype or those with allele frequency difference > 0.2 are excluded prior to the
636 instrument selection steps. In the MR analysis, we pruned variants that are highly correlated
637 (linkage disequilibrium (LD) $r^2 > 0.9$) with the top sc-eQTL within the region, and used set-based
638 approximation accounting for residual LD to combine MR estimates from the remaining variants,
639 with *P*-value derived using Saddlepoint approximation. MR analysis was performed by modelling
640 the effect of increased gene expression level on disease risk or trait levels so that a positive
641 association implies a trait-increasing (or risk-increasing) effect and a negative association implies
642 a trait-decreasing (or protective) effect.

643 To adjust for multiple testing, MR *P*-values were adjusted for false discovery rate (FDR) by
644 calculating *Q*-values based on *P*-values distribution across cell type-gene combinations per trait
645 with the *qvalue* R package[98]. To minimise confounding by pleiotropic effects on other genes, we
646 further performed a test for heterogeneity in dependent instruments (HEIDI)[13] for each MR test
647 we performed. An MR estimate with a $Q < 0.05$ and a $P_{\mathrm{HEIDI}} \geq 0.01$ were considered as evidence
648 of a causal association.
649
650 The comparison with eQTLGen was performed using publicly available pre-formatted SMR
651 summary statistics from eQTLgen Consortium[17], downloaded from https://eqtlgen.org/cis-
652 eqtls.html. Consistent with the original publication, we used a threshold of $P_{\mathrm{eQTL}} < 2.02 \times 10^{-5}$ to
653 select a genetic instrument for MR. Multiple testing correction for MR was performed consistently
654 with the TenK10K study.
655
656 *Identification of GWAS genes*
657 To identify GWAS genes for comparison with MR results, we performed a gene-level aggregation
658 of single-variant estimates from GWAS summary statistics using MAGMA[16]. The aggregation for
659 each gene was performed across variants within the transcription start and end sites of the gene
660 (based on ENCODE v44 hg38 coordinates) using a SNP-wise model with linkage disequilibrium
661 between variants modeled using reference data TenK10K phase I individual-level genotype. *P*-
662 value for a gene-trait association was derived using mean chi-squared statistics by approximation
663 of known sampling distribution. To account for multiple testing, we calculated Q-values across all
664 analytical sets. Genes associated with a trait at $Q < 0.05$ were considered as putative GWAS genes
665 for the trait.
666
667 *Single-cell polygenic enrichment of complex traits*
668 We utilized the single-cell disease relevance score (scDRS)[10] method to quantify polygenic
669 enrichment of complex traits at both individual-cell and cell-type levels. scDRS integrates single-
670 cell gene expression profiles with GWAS summary statistics to associate individual cells to the
671 GWAS trait, by testing overexpression of trait-relevant genes derived from GWAS in a given cell
672 relative to other genes with similar expression profile across all cells. To run scDRS for each trait,
673 we first extracted the top 1,000 trait-associated genes based on *Z*-score derived from the MAGMA
674 analysis. Next, we used scDRS to compute the aggregate expression of the associated genes in
675 each cell, weighted by the GWAS MAGMA *Z*-score and inversely weighted by gene-specific
676 technical covariates in the single-cell data, using the mean-variance model. To compute cell-level
677 *P*-value for overexpression of trait-relevant genes, we followed scDRS method to first generate
678 1,000 sets of cell-specific raw control scores computed using Monte Carlo (MC) samples of
679 matched control gene sets (matching gene set size, mean expression and expression variance of
680 the selected genes). The *P*-value was then derived by comparing normalised trait scores against
681 normalised control scores based on the empirical distribution of the pooled normalised control
682 scores across all control gene sets and all cells.

683 Single-cell RNA sequencing data to perform scDRS analysis were extracted from ~5 million
684 PBMCs that passed through the processing and quality control pipeline in the TenK10K phase I
685 study[15]. The cell-level scoring in scDRS was performed using sex, age, number of genes, and
686 cohort (TOB and BioHEART) as covariates. To manage computational efficiency and account for
687 variations in cell type proportions, we randomly sampled a maximum of 10,000 cells for each of
688 the 28 immune cell types in the TenK10K phase 1 study, resulting in 229,239 total sampled cells.
689 To test the robustness of this sampling strategy, we repeated this procedure using Crohn's disease
690 as an example with an alternative sampling strategy by taking 10% of cells from each cell type.
691
692 *Quantification of cell function score in B cells*
693 We computed cell function scores in B cells as cell states using scDeepID, a flexible multi-task
694 transformer that leverages different biological pathway databases as guides to train the model.
695 First, we trained a scDeepID model with 539,820 B cells (including B naive, B intermediate, and
696 B memory cells) from TenK10K phase 1 single-cell RNA sequencing data. The training used
697 expression features of 2,351 highly variable genes selected with Scanpy function
698 highly_variable_genes (min_mean=0.0123, max_mean=3, min_disp=0.5) with human Biological
699 Process Gene Ontology (GOBP) database as a guide, filtered to the top 400 most informative
700 GOBP pathways based on the presence of pathway genes in the highly variable gene list. We then
701 extracted the attention-based latent space for UMAP representation, and extracted the cell function
702 scores followed by transformation of each cell function score to approximate a normal distribution
703 with a zero mean and one unit variance.
704
705 As a proof-of-concept to explore the function of cells that are enriched for specific phenotypes, we
706 intersected scDeepID cell function score with scDRS score for systemic lupus erythematosus
707 ($scDRS_{SLE}$) amongst 30,000 subsampled B cells used in the scDRS analysis. We first grouped
708 these cells into SLE-enriched cells (cell-level $P_{scDRS\text{-}SLE}$) and non-SLE-enriched cells, and
709 compared mean cell function scores between the two groups using a one-sided *T*-test for each
710 pathway under analysis followed by FDR correction.
711
712 *Estimating the support of single-cell genetics for drug target validation*
713 To evaluate the utility of single-cell genetics for drug target identification, we cross-referenced our
714 cell-type resolved MR results with a target-evidence association score and drug development
715 activity dataset sourced from Open Targets Platform (OTP) [24] version 25.06. Analysis were limited
716 to a set of 12,873 protein-coding genes tested in the MR analysis and 57 diseases with at least one
717 matching indication available in the Open Targets Platform database, resulting in a total of 733,761
718 possible pairs of target (gene) and indication (disease) – the target-indication universe in this study.
719 We used Ensembl gene ID to match tested genes to OTP targets, and used Experimental Factor
720 Ontology (EFO) ID to match tested phenotypes to OTP indications.

721 First, we extracted drug development activity data in OTP (sourced from ChEMBL[25]) and
722 calculated the maximum evidence score in OTP for every target-indication pair. Following
723 ChEMBL classification, we categorised the evidence score into the following:
724

Maximum Clinical Development Stage	**Evidence Score**
Phase I	0.1
Phase II	0.2
Phase III	0.7
Phase I+	0.1 or more
Phase II+	0.2 or more
Phase III+	0.7 or more
Approved	1

725
726 For every category, we computed the 2x2 intersection between target-indication inside or outside
727 the category and target-indication with or without MR support, resulting in four possible
728 combinations: in the category & with MR support, in the category & without MR support, outside
729 the category & with MR support, and outside the category & without MR support. Using this 2x2
730 contingency table, we then computed the odds ratio for being in the category, comparing between
731 target-indication pairs with and without MR support. *P*-value and 95% lower confidence interval
732 were calculated using a one-sided Fisher's exact test implemented in *fisher.test()* R function. For
733 this analysis, MR support was defined as a significant MR association ($Q_{MR} < 0.05$ and $P_{HEIDI} \geq$
734 0.01) in at least one cell type under analysis.
735
736 Second, for every matched target-indication pair under analysis, we extracted the overall OTP
737 evidence score which quantifies aggregated evidence from data sources of genetic association,
738 somatic mutation, known drug, affected pathway, literature, RNA expression, and animal model.
739 We took the maximum of direct (evidence referring directly to indication ID) and indirect
740 association score (evidence referring to indication ID or any of its descendant IDs in EFO). The
741 association score takes a value between 0 and 1, where a higher score indicates more evidence
742 supporting the target-indication association relative to a theoretical maximum. We computed the
743 mean score of target-indication with and without MR support, and tested the difference using a
744 two-sample Welch's *T*-test implemented in *t.test()* R function. We then extracted the evidence
745 score stratified by data score, computed the genetic evidence score (unweighted mean of scores
746 from genetic association sources) and non-genetic evidence (unweighted mean of scores from non-
747 genetic association sources), and repeated the test procedures for the two evidence sources.

748 Third, we assessed the concordance between MR-predicted effect direction and drug mechanism
749 across clinical development phases. The MR-predicted effect direction for a target-indication pair
750 was derived by taking the majority of effect directions across cell types with significant MR
751 associations. To remove ambiguity, target-indication with equal number of positive and negative
752 effects across cell types were excluded from the analysis. The drug mechanism effects were
753 extracted from ChEMBL drug data in OTP, and matched with ChEMBL ID of drug with maximum
754 clinical development phase for a given target-indication pair.
755
756 *Crohn's disease case study*
757 We extracted all MR associations for Crohn's disease which passed the discovery threshold
758 of FDR-adjusted $Q < 5\%$ and $P_{\text{HEIDI}} > 0.01$.. For each unique MR gene identified, we constructed
759 a Boolean annotation (TRUE / FALSE) based on membership in: 1) the canonical gene list, 2) the
760 drug target gene list 3) the drug target pathways, 4) cell type-specific MR association (whether the
761 MR association is only identified in one cell type), 5) discordance of MR effect direction in at least
762 two cell types, and 6) differentially expressed in Crohn's disease colon tissue where the major cell
763 types are matched (**Supplementary Table 13**).
764
765 Canonical genes from Crohn's disease were downloaded from a systematic review of genes
766 associated with Crohn's disease[99] The full list (n = 1,172) was filtered to belong to Categories
767 "Experimental evidence of variant", "Other evidences of genetic alterations", "Treatment
768 Response" , "Biologically related but no evidence of mutation", with a Document Score >= 8 and
769 Abstract Number >= 10, resulting in a list of 49 genes (**Supplementary Table 15**). The drug target
770 genes were extracted from review papers[100,101], and the drug pathway genes (IL23/12, ITGA4,
771 JAK/STAT, TNF) were extracted from GSEA platform[102,103] following the approach described
772 previously[70] (**Supplementary Table 16**). The list of differentially expressed genes was derived
773 from an external dataset of single-cell RNA sequencing in intestinal tissues from 46 Crohn's
774 disease patients and 25 healthy controls, comparing both inflamed and non-inflamed disease tissue
775 with healthy tissue.[33] To derive this annotation, we extracted differentially expressed genes in the
776 colon (pre-computed Discrete FDR < 0.05) across immune cell types. We harmonised the original
777 cell type annotation with the TenK10K cell type annotation used in this study by matching major
778 cell types (**Supplementary Table 17**). We then identified MR and DEG genes in matched major
779 cell types, and checked for concordance of direction of effect by comparing the beta sign between
780 MR and the Discrete model of DEG (**Supplementary Table 14**). This generated multiple
781 combinations of gene-major cell type pairs, based on unique combinations of original MR cell
782 types and DEG cell types. 'All concordant' corresponds to cases where every gene-MR cell type-
783 DEG cell type set has a matched direction of effect between MR and DEG. 'All discordant'
784 corresponds to cases where every gene-MR cell type-DEG cell type set has an unmatched direction
785 of effect between MR and DEG. 'Mixed concordance' corresponds to genes which have discordant
786 directions between or within major cell types.

To evaluate expression of the genes of interest in colon tissue (**Figure 5d**), we extracted sc-RNAseq data from colon tissue of Crohn's disease patients[33], downloaded from CELLxGENE[104]. We present the gene expression as normalised counts on the original Uniform Manifold Approximation and Projection (UMAP) embedding using Scanpy, stratified by 'Type' into "Healthy", "Non-inflamed Crohn's disease", and "Inflamed Crohn's disease" and filtered to immune cell types with matched major cell types annotation in the TenK10K. The original cell annotation and distribution of cells before and after filtering is presented in **Supplementary Figure 15**.

Author contribution

A.H., A.S., R.T., and J.E.P. designed the overall study. A.H., A.S., R.T., B.B., and J.E.P. contributed to the initial writing of the manuscript. A.H. and A.S. performed single-cell MR analysis. A.H. designed and performed MAGMA, scDRS, and drug target support analyses. R.T. designed and performed the proof-of-concept analysis in Crohn's disease. B.B. and A.S.E.C. performed the sc-eQTL mapping in the TenK10K led by J.E.P., with genotype data processing led by K.M.dL. and D.G.M. A.S., H.A.T., and A.H. processed GWAS summary statistics. E.S. and E.S.Z. contributed to the sample collection, processing, and generation of single-cell RNA sequencing data. A.X, J.F., and P.C.A. provided input for single-cell MR and scDRS analysis. H.L.H. designed and performed the cell function analysis with scDeepID. A.H., A.S., B.B., R.T., and H.L.H. contributed to the design and generation of figures, tables, and the computational pipeline for analysis. G.A.F. led the BioHEART study and A.W.H. led the Tasmanian Ophthalmic Biobank study, of which samples were included in the sc-eQTL mapping. J.E.P. contributed to funding acquisition and supervision for the study. All authors reviewed and approved the final version of the manuscript.

Competing interests

D.G.M. is a paid advisor to Insitro and GSK, and receives research funding from Google and Microsoft, unrelated to the work described in this manuscript. G.A.F reports grants from National Health and Medical Research Council (Australia), Abbott Diagnostic, Sanofi, Janssen Pharmaceuticals, and NSW Health, and honorarium from CSL, CPC Clinical Research, Sanofi, Boehringer-Ingelheim, Heart Foundation, and Abbott. G.A.F serves as Board Director for the Australian Cardiovascular Alliance (past President), Executive Committee Member for CPC Clinical Research, Founding Director and CMO for Prokardia and Kardiomics, and Executive Committee member for the CAD Frontiers A2D2 Consortium. In addition, G.A.F serves as CMO for the non-profit, CAD Frontiers, with industry partners including, Novartis, Amgen, Siemens Healthineers, ELUCID, Foresite Labs LLC, HeartFlow, Canon, Cleerly, Caristo, Genentech, Artyra, and Bitterroot Bio, Novo Nordisk and Allelica. G.A.F also reports the following patents: "Patent Biomarkers and Oxidative Stress" awarded USA May 2017 (US9638699B2) issued to Northern Sydney Local Health District, "Use of P2X7R antagonists in cardiovascular disease" PCT/AU2018/050905 licensed to Prokardia, "Methods for treatment and prevention of vascular disease" PCT/AU2015/000548 issued to The University of Sydney/Northern Sydney Local Health District, "Methods for predicting coronary artery disease" AU202290266 issued to The University of Sydney, and the patent "Novel P2X7 Receptor Antagonists" PCT/AU2022/051400 (23.11.2022), International App No: WO/2023/092175 (01.06.2023), issued to The University of Sydney.

Data availability

The main single-cell MR association results will be made publicly available via Zenodo website prior to acceptance.

Code availability

Analysis code will be made publicly available at https://github.com/powellgenomicslab/tenk10k-causal-paper

842 **References**
843

844 1. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs:
845 illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
846 2. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: From
847 polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
848 3. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of
849 autoimmune disease. *Science* **376**, eabf3041 (2022).
850 4. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of
851 disease loci. *Nature* **606**, 120–128 (2022).
852 5. Cuomo, A. S. E., Nathan, A., Raychaudhuri, S., MacArthur, D. G. & Powell, J. E. Single-
853 cell genomics meets human genetics. *Nat. Rev. Genet.* (2023) doi:10.1038/s41576-023-00599-5.
854 6. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time.
855 *Nature* **597**, 250–255 (2021).
856 7. Perez, R. K. *et al.* Single-cell RNA-seq reveals cell type-specific molecular and genetic
857 associations to lupus. *Science* **376**, eabf1970 (2022).
858 8. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify
859 novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* **25**, 1104–1112
860 (2022).
861 9. Natri, H. M. *et al.* Cell-type-specific and disease-associated expression quantitative trait
862 loci in the human lung. *Nat. Genet.* **56**, 595–604 (2024).
863 10. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual
864 cells in single-cell RNA-seq data. *Nat. Genet.* **54**, 1572–1580 (2022).
865 11. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by
866 integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 (2022).
867 12. Porcu, E. *et al.* Mendelian randomization integrating GWAS and eQTL data reveals
868 genetic determinants of complex and clinical traits. *Nat. Commun.* **10**, 3300 (2019).
869 13. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts
870 complex trait gene targets. *Nature Genetics* **48**, 481–487 (2016).
871 14. Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms
872 underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
873 15. Cuomo, A. S. E. *et al.* Impact of Rare and Common Genetic Variation on Cell Type-
874 Specific Gene Expression. *Genetic and Genomic Medicine* (2025).
875 16. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-
876 set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
877 17. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic
878 loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310
879 (2021).
880 18. Middel, P., Raddatz, D., Gunawan, B., Haller, F. & Radzun, H.-J. Increased number of
881 mature dendritic cells in Crohn's disease: evidence for a chemokine mediated retention
882 mechanism. *Gut* **55**, 220–227 (2006).
883 19. Collin, M. & Bigley, V. Human dendritic cell subsets: an update. *Immunology* **154**, 3–20
884 (2018).
885 20. Swiecki, M. & Colonna, M. The multifaceted biology of plasmacytoid dendritic cells.
886 *Nat. Rev. Immunol.* **15**, 471–485 (2015).
887 21. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic

evidence on clinical success. *Nature* **629**, 624–629 (2024).
22. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nature Genetics* **47**, 856–860 (2015).
23. Dann, E. *et al.* Estimating the impact of single-cell RNA sequencing of human tissues on drug target validation. *medRxiv* 2024.04.04.24305313 (2024) doi:10.1101/2024.04.04.24305313.
24. Buniello, A. *et al.* Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res* **53**, D1467–D1475 (2025).
25. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024).
26. Hampel, H. *et al.* The amyloid-β pathway in Alzheimer's disease. *Mol. Psychiatry* **26**, 5481–5503 (2021).
27. Zheng, Z. *et al.* Glucagon-like peptide-1 receptor: mechanisms and advances in therapy. *Signal Transduct. Target. Ther.* **9**, 234 (2024).
28. Lincoff, A. M. *et al.* Semaglutide and cardiovascular outcomes in obesity without diabetes. *N. Engl. J. Med.* **389**, 2221–2232 (2023).
29. Ryan, D. H. *et al.* Long-term weight loss effects of semaglutide in obesity without diabetes in the SELECT trial. *Nat. Med.* **30**, 2049–2057 (2024).
30. Roda, G. *et al.* Crohn's disease. *Nat. Rev. Dis. Primers* **6**, 22 (2020).
31. Torres, J., Mehandru, S., Colombel, J.-F. & Peyrin-Biroulet, L. Crohn's disease. *Lancet* **389**, 1741–1755 (2017).
32. de Lange, K. M. & Barrett, J. C. Understanding inflammatory bowel disease via immunogenetics. *J. Autoimmun.* **64**, 91–100 (2015).
33. Kong, L. *et al.* The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. *Immunity* **56**, 444–458.e5 (2023).
34. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19**, 110–124 (2018).
35. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173**, 1573–1580 (2018).
36. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat. Rev. Genet.* **25**, 639–657 (2024).
37. Stankey, C. T. *et al.* A disease-associated gene desert directs macrophage inflammation through ETS2. *Nature* **630**, 447–456 (2024).
38. Jackson, S. W. *et al.* B cell IFN-γ receptor signaling promotes autoimmune germinal centers via cell-intrinsic induction of BCL-6. *J. Exp. Med.* **213**, 733–750 (2016).
39. Burdett, T. *et al.* GWAS Catalog. https://www.ebi.ac.uk/gwas/genes/ZBTB38.
40. Ocskó, T. *et al.* Transcription factor Zbtb38 downregulates the expression of anti-inflammatory IL1r2 in mouse model of rheumatoid arthritis. *Biochim. Biophys. Acta Gene Regul. Mech.* **1861**, 1040–1047 (2018).
41. Yan, M. *et al.* Dynamic regulatory networks of T cell trajectory dissect transcriptional control of T cell state transition. *Mol. Ther. Nucleic Acids* **26**, 1115–1129 (2021).
42. He, C. *et al.* NCF4 dependent intracellular reactive oxygen species regulate plasma cell formation. *Redox Biol.* **56**, 102422 (2022).
43. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).

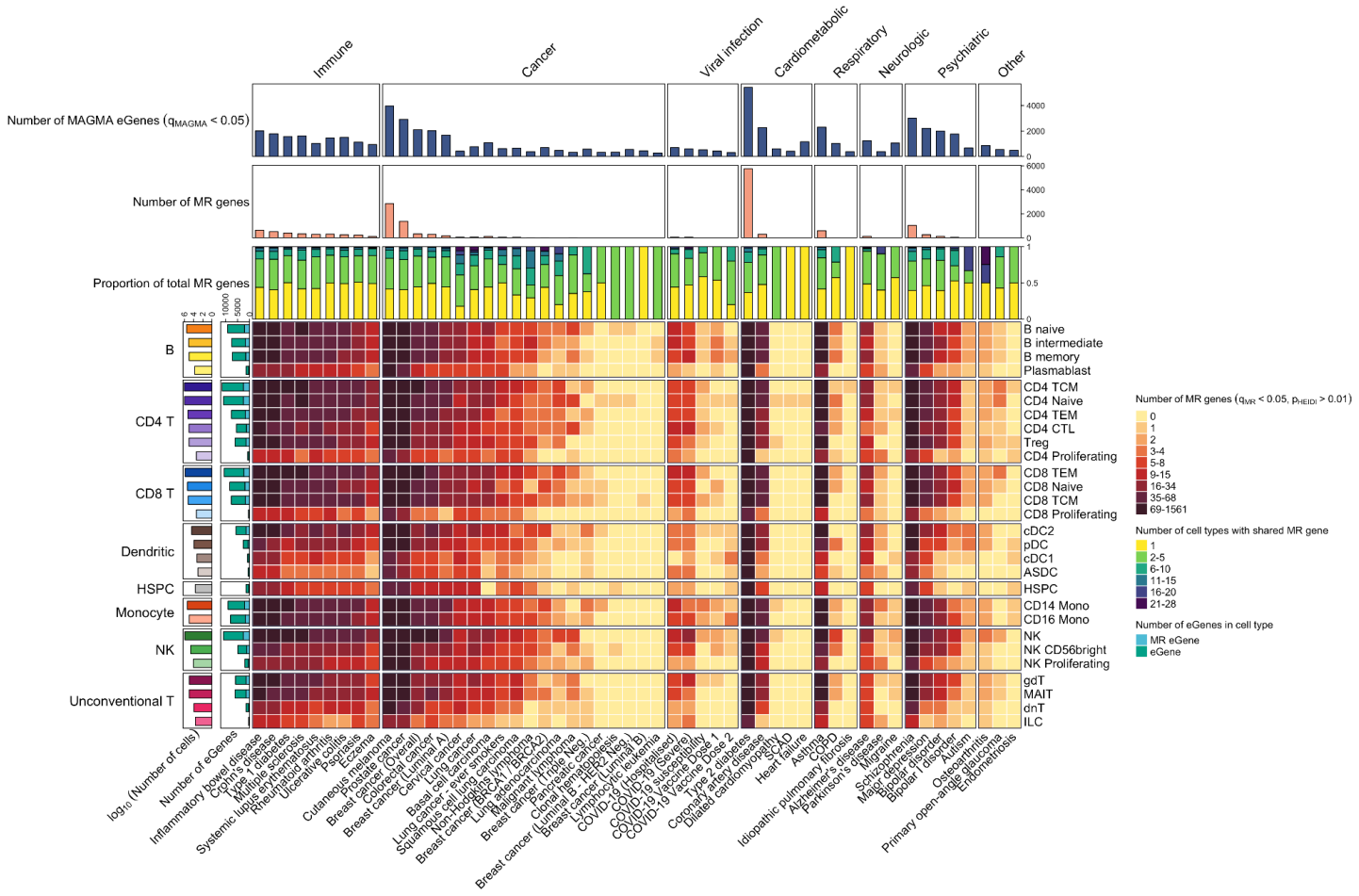
934 44. Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F. & Guigó, R. Identification and
935 analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat.*
936 *Commun.* **12**, 727 (2021).
937 45. Yamaguchi, K. *et al.* Splicing QTL analysis focusing on coding sequences reveals
938 mechanisms for disease susceptibility loci. *Nat. Commun.* **13**, 4659 (2022).
939 46. Munafò, M. R., Higgins, J. P. T. & Smith, G. D. Triangulating evidence through the
940 inclusion of genetically informed designs. *Cold Spring Harb. Perspect. Med.* **11**, a040659
941 (2021).
942 47. Luo, J. *et al.* Structural basis for the dual recognition of IL-12 and IL-23 by ustekinumab.
943 *J. Mol. Biol.* **402**, 797–812 (2010).
944 48. Ma, C., Panaccione, R., Khanna, R., Feagan, B. G. & Jairath, V. IL12/23 or selective
945 IL23 inhibition for the management of moderate-to-severe Crohn's disease? *Best Pract. Res.*
946 *Clin. Gastroenterol.* **38-39**, 101604 (2019).
947 49. Verstockt, B. *et al.* IL-12 and IL-23 pathway inhibition in inflammatory bowel disease.
948 *Nat. Rev. Gastroenterol. Hepatol.* **20**, 433–446 (2023).
949 50. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. scPred: accurate
950 supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**,
951 264 (2019).
952 51. Michielsen, L., Reinders, M. J. T. & Mahfouz, A. Hierarchical progressive learning of
953 cell identities in single-cell data. *Nat. Commun.* **12**, 2799 (2021).
954 52. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–
955 3587.e29 (2021).
956 53. Zhou, W. *et al.* Efficient and accurate mixed model association tool for single-cell eQTL
957 analysis. *Genetic and Genomic Medicine* (2024).
958 54. Vuckovic, D. *et al.* The polygenic and monogenic basis of blood traits and diseases. *Cell*
959 **182**, 1214–1231.e11 (2020).
960 55. Chai, R. C. *et al.* Multiscale analysis and functional validation of the cellular and genetic
961 determinants of skeletal disease. *Genetics* (2024).
962 56. Rashkin, S. R. *et al.* Pan-cancer study detects genetic risk variants and shared genetic
963 basis in two large cohorts. *Nat. Commun.* **11**, 4423 (2020).
964 57. Fernandez-Rozadilla, C. *et al.* Deciphering colorectal cancer genetics through multi-omic
965 analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat.*
966 *Genet.* **55**, 89–99 (2023).
967 58. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer
968 susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat.*
969 *Genet.* **49**, 1126–1132 (2017).
970 59. Adolphe, C. *et al.* Genetic and functional interaction network analysis reveals global
971 enrichment of regulatory T cell genes influencing basal cell carcinoma susceptibility. *Genome*
972 *Med.* **13**, 19 (2021).
973 60. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
974 population. *Nature* **613**, 508–518 (2023).
975 61. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million
976 Veteran Program. *Science* **385**, eadj1182 (2024).
977 62. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer
978 susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).
979 63. Wang, A. *et al.* Characterizing prostate cancer risk through multi-ancestry genome-wide

980 discovery of 187 novel risk variants. *Nat. Genet.* **55**, 2065–2074 (2023).
981 64. Kar, S. P. *et al.* Genome-wide analyses of 200,453 individuals yield new insights into the
982 causes and consequences of clonal hematopoiesis. *Nat. Genet.* **54**, 1155–1166 (2022).
983 65. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human
984 phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
985 66. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell
986 epigenomics. *Nature* **594**, 398–402 (2021).
987 67. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and
988 adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**,
989 1457–1464 (2015).
990 68. Ishigaki, K. *et al.* Multi-ancestry genome-wide association analyses identify novel
991 genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).
992 69. Stuart, P. E. *et al.* Transethnic analysis of psoriasis susceptibility in South Asians and
993 Europeans enhances fine-mapping in the MHC and genomewide. *HGG Adv.* **3**, 100069 (2022).
994 70. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of
995 multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
996 71. International Multiple Sclerosis Genetics Consortium (IMSGC) *et al.* Analysis of
997 immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.*
998 **45**, 1353–1360 (2013).
999 72. Budu-Aggrey, A. *et al.* European and multi-ancestry genome-wide association meta-
1000 analysis of atopic dermatitis highlights importance of systemic immune regulation. *Nat.*
1001 *Commun.* **14**, 6172 (2023).
1002 73. COVID-19 Host Genetics Initiative. A second update on mapping the human genetic
1003 architecture of COVID-19. *Nature* **621**, E7–E26 (2023).
1004 74. Lammi, V. *et al.* Genome-wide association study of long COVID. *Nat. Genet.* **57**, 1402–
1005 1417 (2025).
1006 75. Bian, S. *et al.* Genetic determinants of IgG antibody response to COVID-19 vaccination.
1007 *Am. J. Hum. Genet.* **111**, 181–199 (2024).
1008 76. Suzuki, K. *et al.* Multi-ancestry genome-wide study in >2.5 million individuals reveals
1009 heterogeneity in mechanistic pathways of type 2 diabetes and complications. *medRxiv* (2023)
1010 doi:10.1101/2023.03.31.23287839.
1011 77. Zhao, C., Ma, T., Cheng, X., Zhang, G. & Bai, Y. Genome-wide association study of
1012 cardiometabolic multimorbidity in the UK Biobank. *Clin. Genet.* **106**, 72–81 (2024).
1013 78. Adlam, D. *et al.* Genome-wide association meta-analysis of spontaneous coronary artery
1014 dissection identifies risk variants and genes related to artery integrity and tissue-mediated
1015 coagulation. *Nat. Genet.* **55**, 964–972 (2023).
1016 79. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes
1017 for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
1018 80. Zheng, S. L. *et al.* Genome-wide association analysis provides insights into the molecular
1019 etiology of dilated cardiomyopathy. *Nat. Genet.* **56**, 2646–2658 (2024).
1020 81. Henry, A. *et al.* Genome-wide association study meta-analysis provides insights into the
1021 etiology of heart failure and its subtypes. *Nat. Genet.* **57**, 815–828 (2025).
1022 82. Surakka, I. *et al.* Multi-ancestry meta-analysis identifies 2 novel loci associated with
1023 ischemic stroke and reveals heterogeneity of effects between sexes and ancestries. *bioRxiv*
1024 (2022) doi:10.1101/2022.02.28.22271647.
1025 83. Kim, J. J. *et al.* Multi-ancestry genome-wide association meta-analysis of Parkinson's

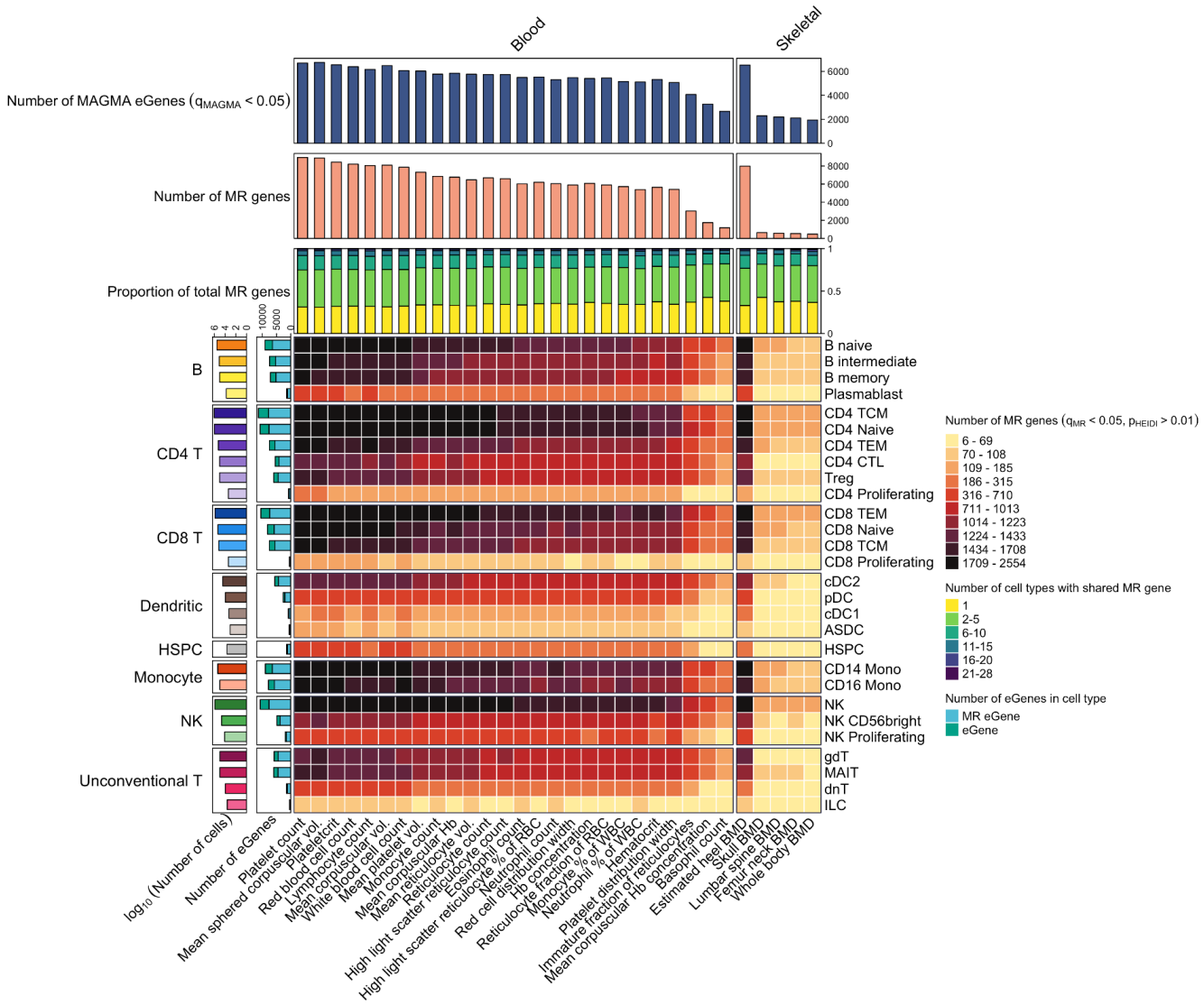
1026 disease. *Nat. Genet.* **56**, 27–36 (2024).
1027 84. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and
1028 related dementias. *Nat. Genet.* **54**, 412–436 (2022).
1029 85. Hautakangas, H. *et al.* Genome-wide analysis of 102,084 migraine cases identifies 123
1030 risk loci and subtype-specific risk alleles. *Nat. Genet.* **54**, 152–160 (2022).
1031 86. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the
1032 genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208 (2023).
1033 87. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
1034 disorder. *Nat. Genet.* **51**, 431–444 (2019).
1035 88. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
1036 schizophrenia. *Nature* **604**, 502–508 (2022).
1037 89. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid
1038 psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).
1039 90. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder
1040 cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
1041 91. Tsuo, K. *et al.* Multi-ancestry meta-analysis of asthma identifies novel associations and
1042 highlights the value of increased power and diversity. *Cell Genom.* **2**, 100212 (2022).
1043 92. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery
1044 across human disease. *Cell Genom.* **2**, 100192 (2022).
1045 93. Allen, R. J. *et al.* Genome-wide association study of susceptibility to idiopathic
1046 pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **201**, 564–574 (2020).
1047 94. McDonald, M.-L. N. *et al.* Novel genetic loci associated with osteoarthritis in multi-
1048 ancestry analyses in the Million Veteran Program and UK Biobank. *Nat. Genet.* **54**, 1816–1826
1049 (2022).
1050 95. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome
1051 suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**, e1007813
1052 (2018).
1053 96. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain
1054 and inflammatory conditions. *Nat. Genet.* **55**, 423–436 (2023).
1055 97. Lo Faro, V. *et al.* Novel ancestry-specific primary open-angle glaucoma loci and shared
1056 biology with vascular mechanisms and cell proliferation. *Cell Rep. Med.* **5**, 101430 (2024).
1057 98. John D. Storey with contributions from Andrew J. Bass, A. A. D. R. *Qvalue*.
1058 (Bioconductor, 2017). doi:10.18129/B9.BIOC.QVALUE.
1059 99. Garza-Hernandez, D. *et al.* A systematic review and functional bioinformatics analysis of
1060 genes associated with Crohn's disease identify more than 120 related genes. *BMC Genomics* **23**,
1061 302 (2022).
1062 100. Feng, Z. *et al.* Breaking through the therapeutic ceiling of inflammatory bowel disease:
1063 Dual-targeted therapies. *Biomed. Pharmacother.* **158**, 114174 (2023).
1064 101. Tyebally, R., Xue, A. & Powell, J. E. The potential clinical impact of cell type-specific
1065 genetic regulation: Crohn's disease. *Clin. Transl. Med.* **13**, e1474 (2023).
1066 102. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1067 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550
1068 (2005).
1069 103. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation
1070 are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
1071 104. CZI Single-Cell Biology Program *et al.* CZ CELL×GENE Discover: A single-cell data

platform for scalable exploration, analysis and modeling of aggregated data. *bioRxiv* (2023) doi:10.1101/2023.10.30.563174.

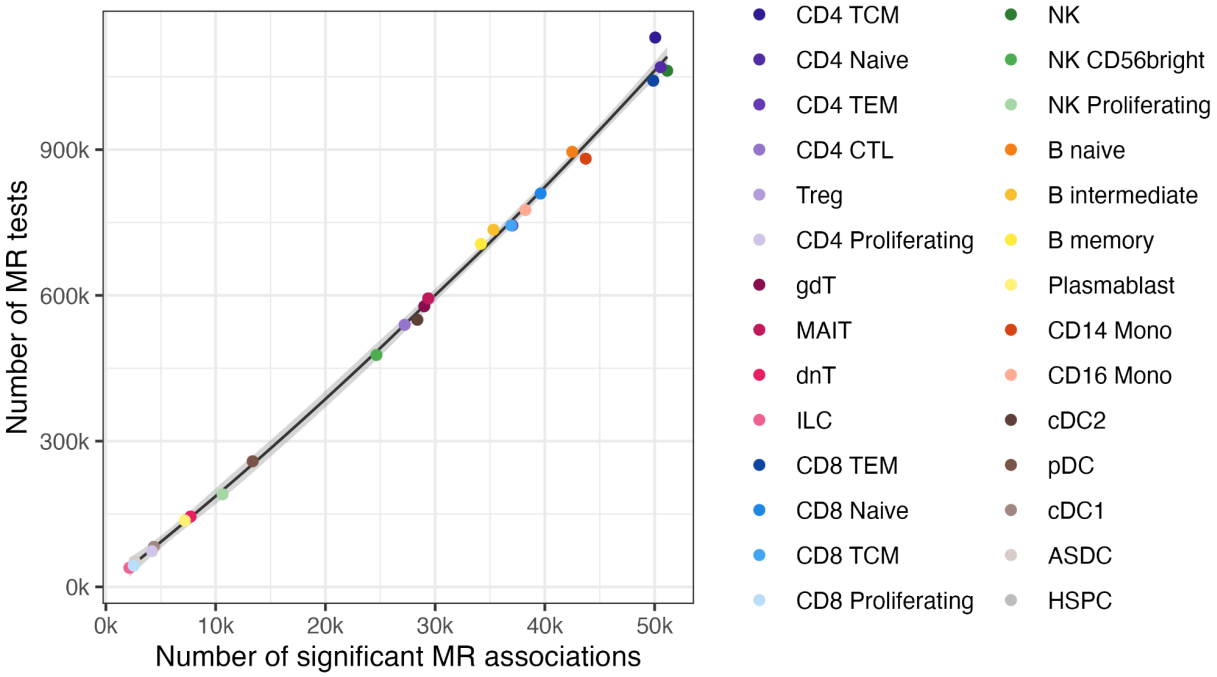
Supplementary Figures



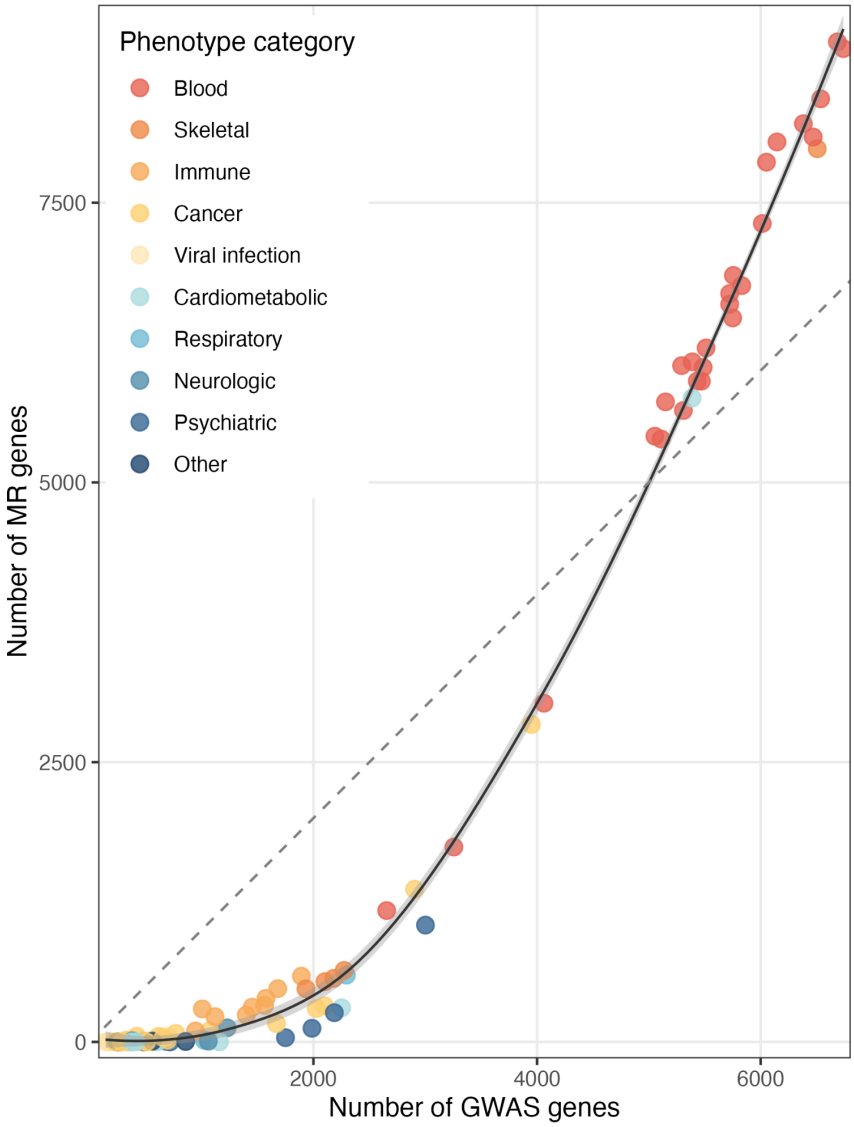
Supplementary Figure 1. Summary of MR associations across 53 diseases and 28 immune cell types. The heatmap shows the number of MR-significant genes per cell type (rows) and traits (columns), binned into deciles with the first two deciles containing zero-only values merged into one. Left-side bar plots indicate (i) the $\log_{10}$-transformed number of single cells per cell type (left) and (ii) the number and proportion of eGenes with at least one MR-significant disease association (right). Top annotations display the number of MAGMA-significant GWAS genes per trait that are also eGenes (top), the total number of MR-significant genes per trait (middle), and the number of cell types in which each gene is MR-significant (bottom). All MR associations meet the threshold of $Q_{MR} < 0.05$ and $P_{HEIDI} \geq 0.01$.



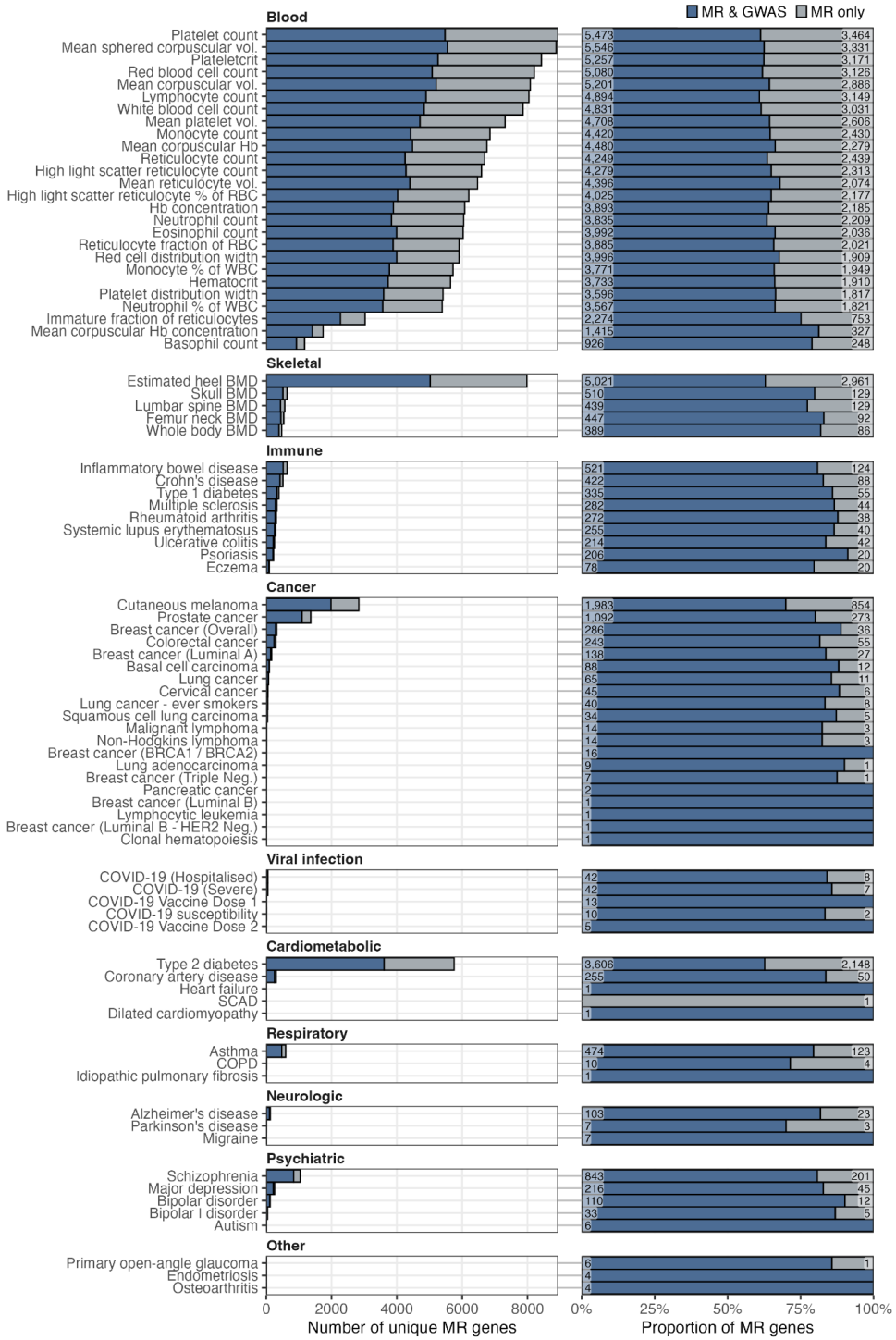
Supplementary Figure 2. Summary of MR associations across 31 biomarker traits and 28 immune cell types. The heatmap shows the number of MR-significant genes per cell type (rows) and traits (columns), binned into deciles. Left-side bar plots indicate (i) the $\log_{10}$-transformed number of single cells per cell type (left) and (ii) the number and proportion of eGenes with at least one MR-significant disease association (right). Top annotations display the number of MAGMA-significant GWAS genes per trait that are also eGenes (top), the total number of MR-significant genes per trait (middle), and the number of cell types in which each gene is MR-significant (bottom). All MR associations meet the threshold of $Q_{MR} < 0.05$ and $P_{HEIDI} \geq 0.01$.



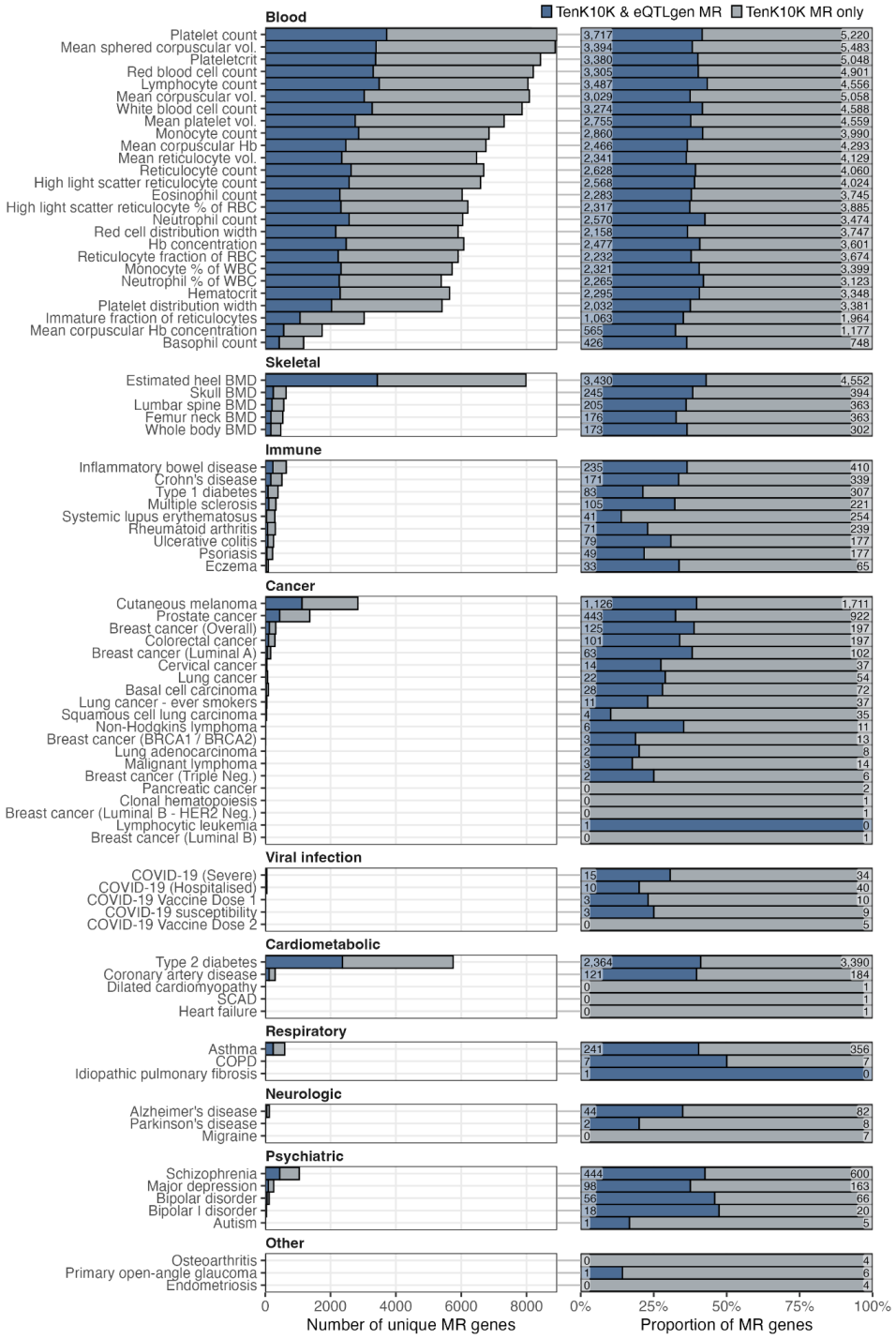
Supplementary Figure 3. Number of significant MR associations and number of MR tests across cell types



Supplementary Figure 4. Number of MR and GWAS genes across 100 complex traits.

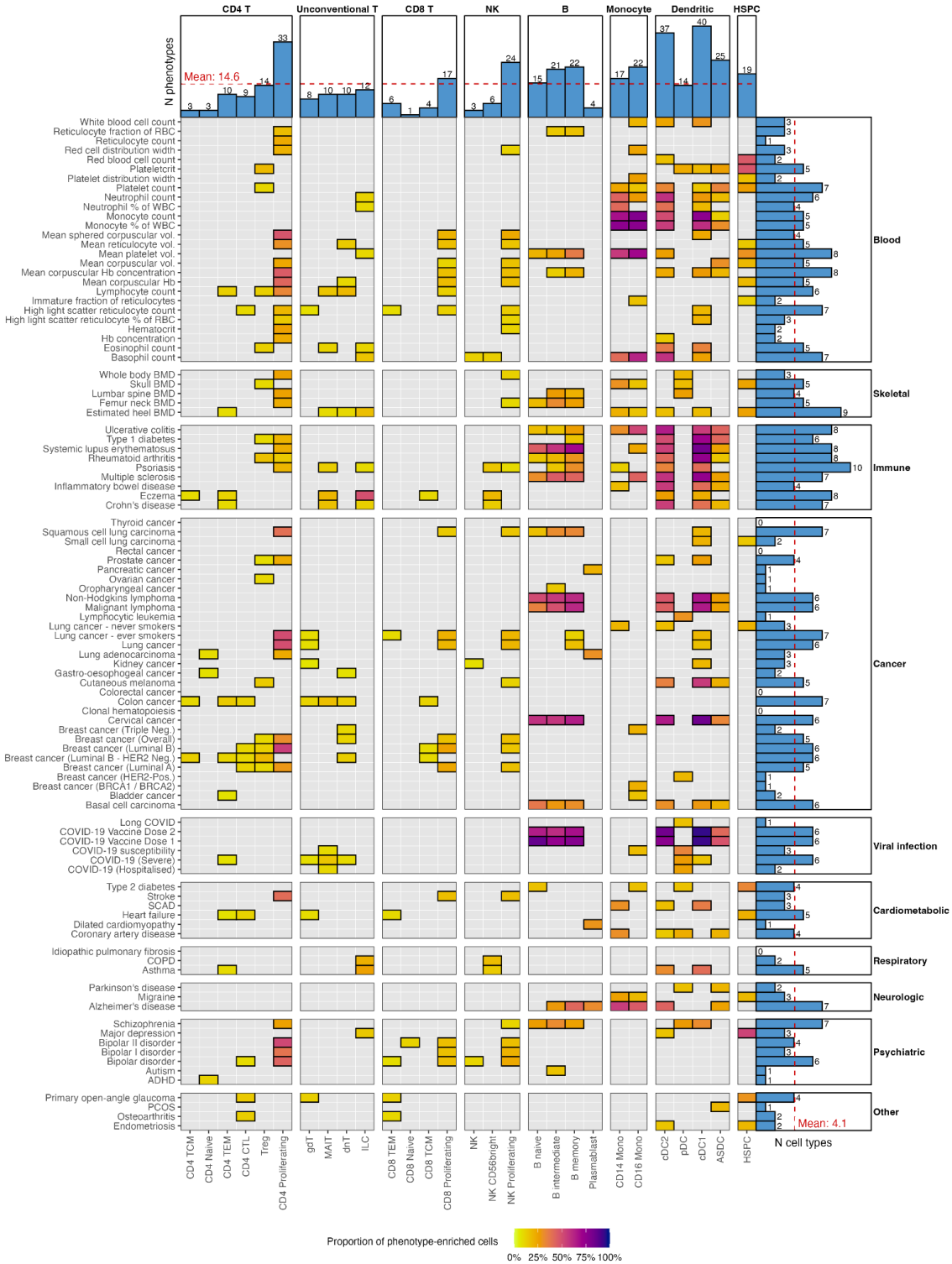


Supplementary Figure 5. Overlaps between MR and GWAS genes across phenotypes. For a given trait, MR genes are defined as genes showing MR association ($Q_{MR} < 0.05$, $P_{HEIDI} \geq 0.01$) in at least one cell type and GWAS genes are defined as associated genes with Q_{MAGMA} <0.05. Bar charts denote numbers of unique MR genes (left panel) and proportion of MR genes (right panel), categorised into whether the gene is also identified as GWAS genes. Numbers of MR & GWAS genes / GWAS only genes are labelled on the left / right hand side of the proportion bar charts.

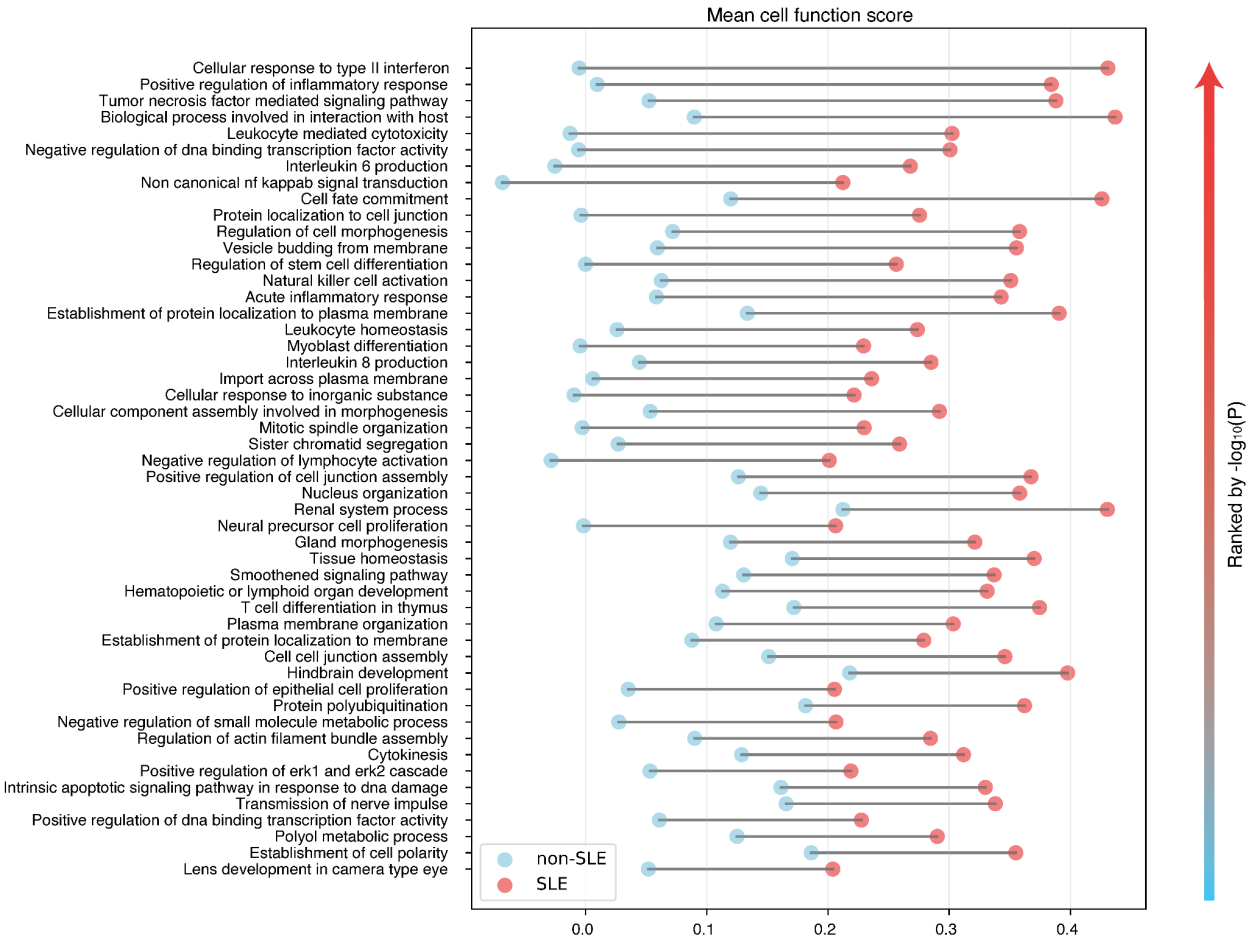


Supplementary Figure 6. Overlaps between genes identified in MR using single-cell eQTL instruments from TenK10K and using bulk-tissue eQTL instruments from eQTLgen across phenotypes. For a given trait, MR genes are defined as genes showing MR association ($Q_{MR} < 0.05$, $P_{HEIDI} \geq 0.01$), in at least one cell type for the TenK10K analysis. Bar charts denote numbers of unique MR genes (left panel) and proportion of MR genes (right panel) identified from the TenK10K analysis, categorised into whether the gene is also identified in the eQTLgen analysis. Numbers of TenK10K & eQTLgen MR

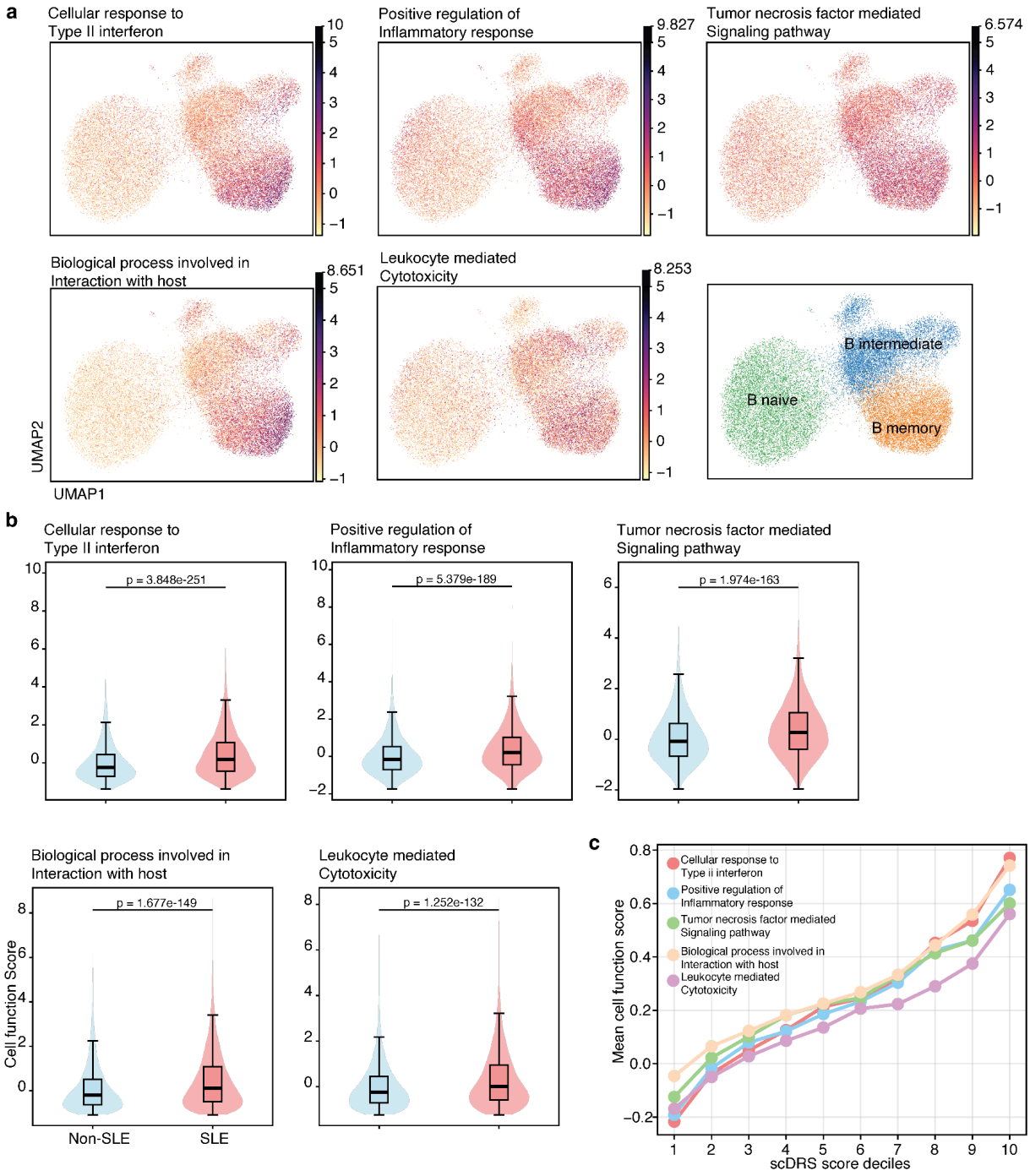
genes / TenK10K only MR genes are labelled on the left / right hand side of the proportion bar charts.



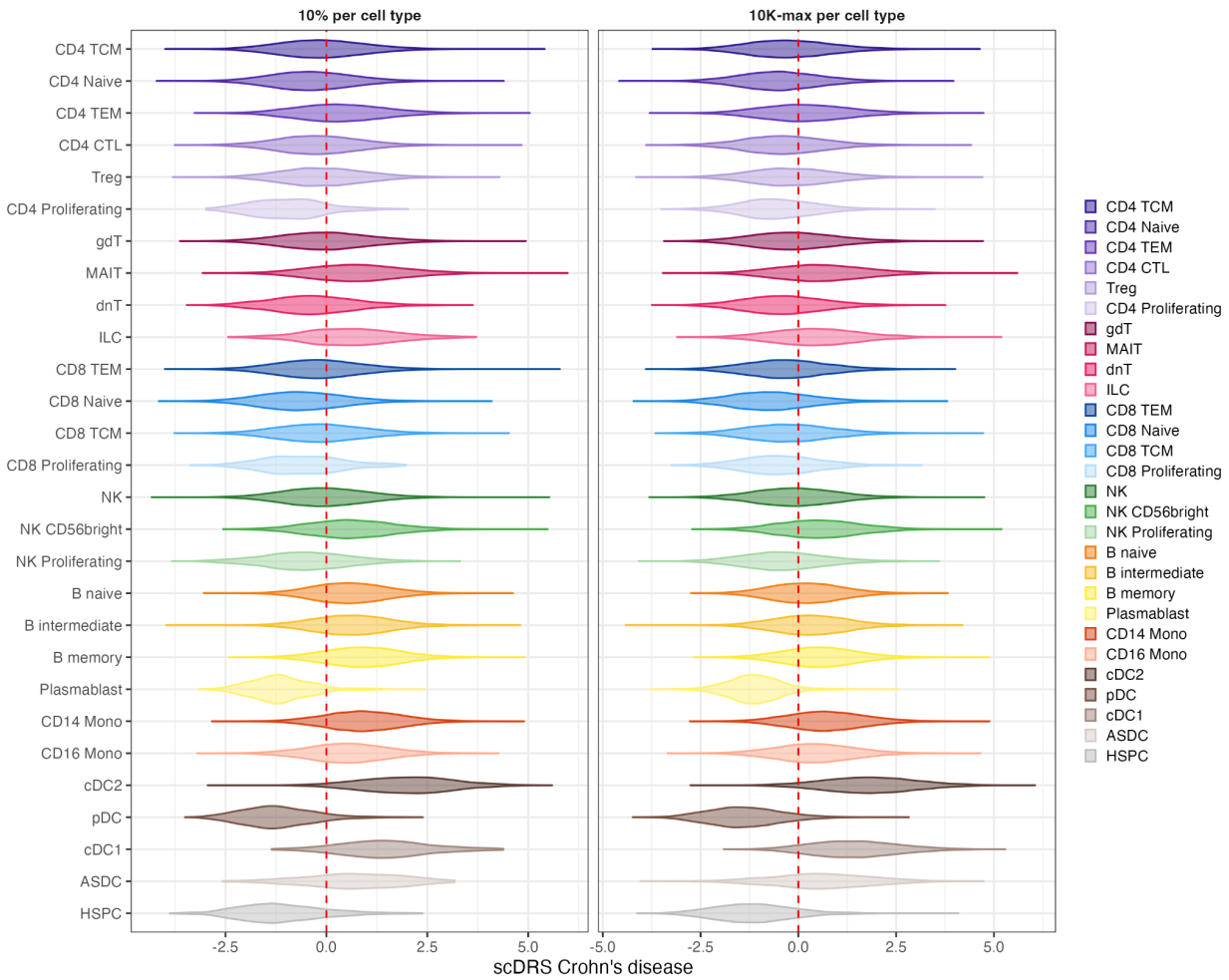
Supplementary Figure 7. Polygenic enrichment of phenotypes across immune cell types



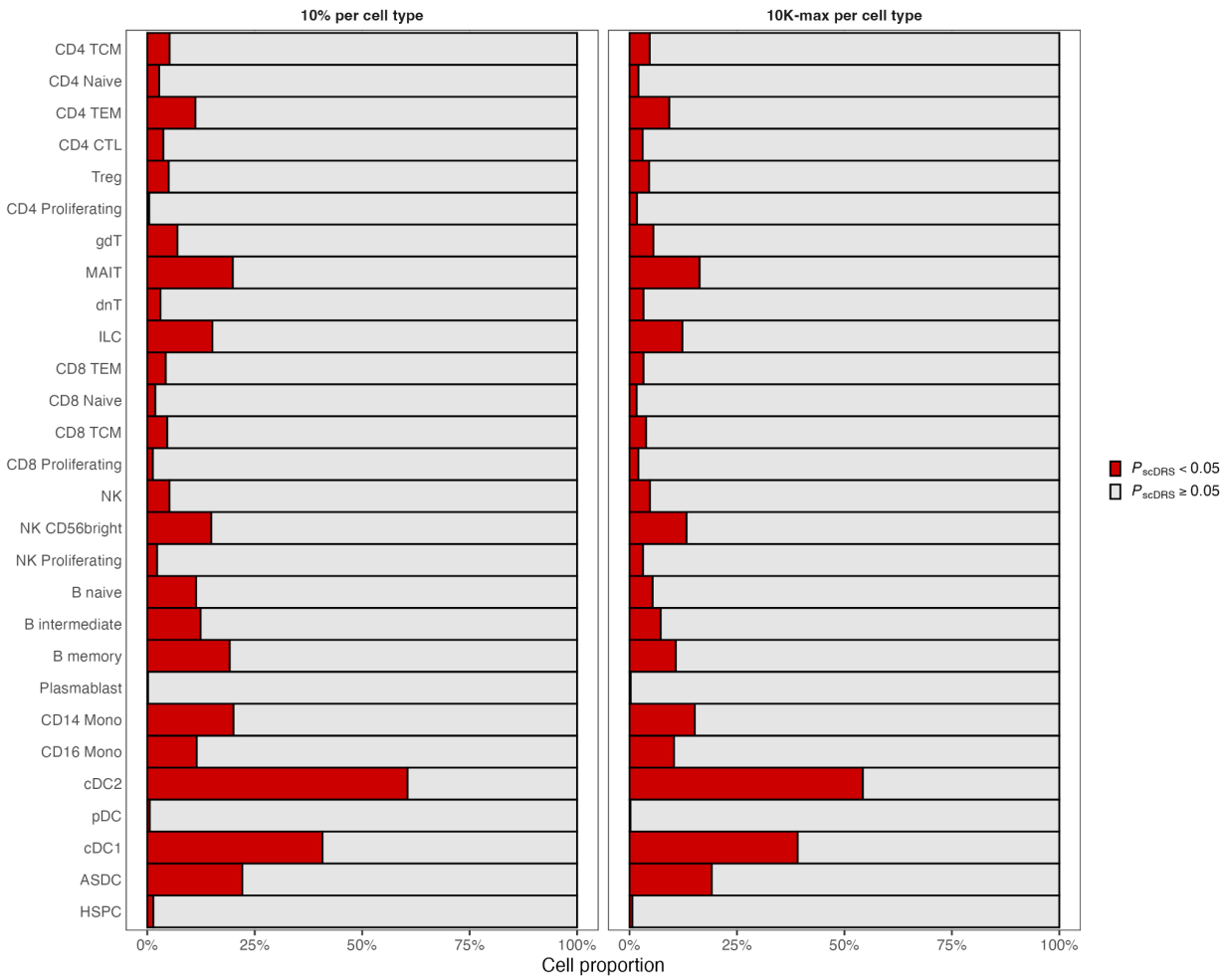
Supplementary Figure 8. The mean scDeepID cell function scores of the top 50 cell functions, based on the FDR-adjusted one-side *T*-test between SLE-enriched (cell-level scDRS *P*-value for SLE polygenic enrichment < 0.05) and non-SLE-enriched cells (cell-level scDRS P-value ≥ 0.05). To pinpoint the most relevant cell functions, we only visualise cell functions with mean score > 0.2 in SLE-enriched cells.



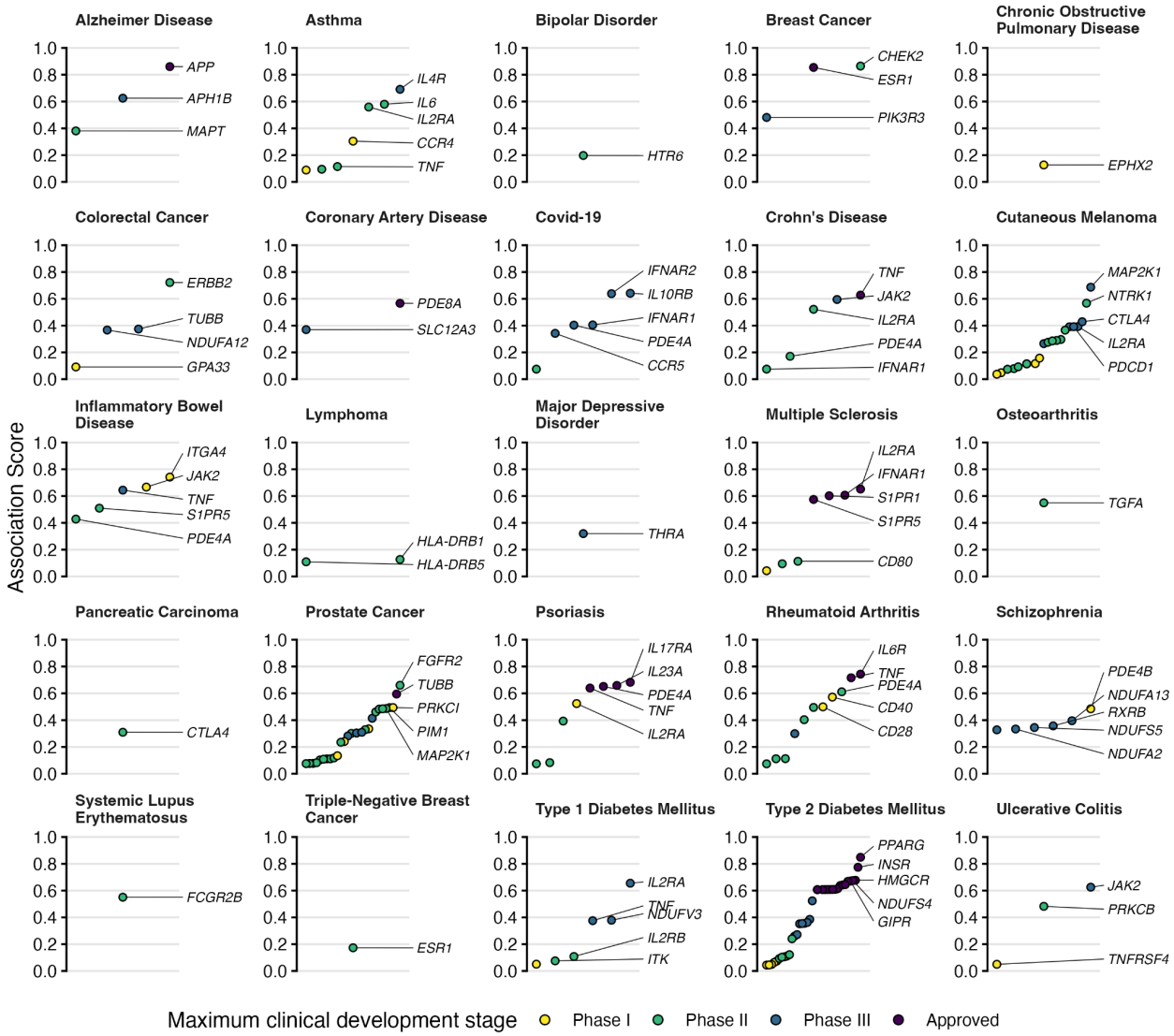
Supplementary Figure 9. a) UMAP representation of the 30,000 sampled B cells used in scDRS analysis. The UMAP was generated from scDeepID attention embedding, with scores of 5 selected cell functions, as well as the cell type labels. **b)** Distribution of the top 5 cell functions scores from (A). **c)** Mean scores of the top 5 cell functions across deciles of scDRS SLE score.



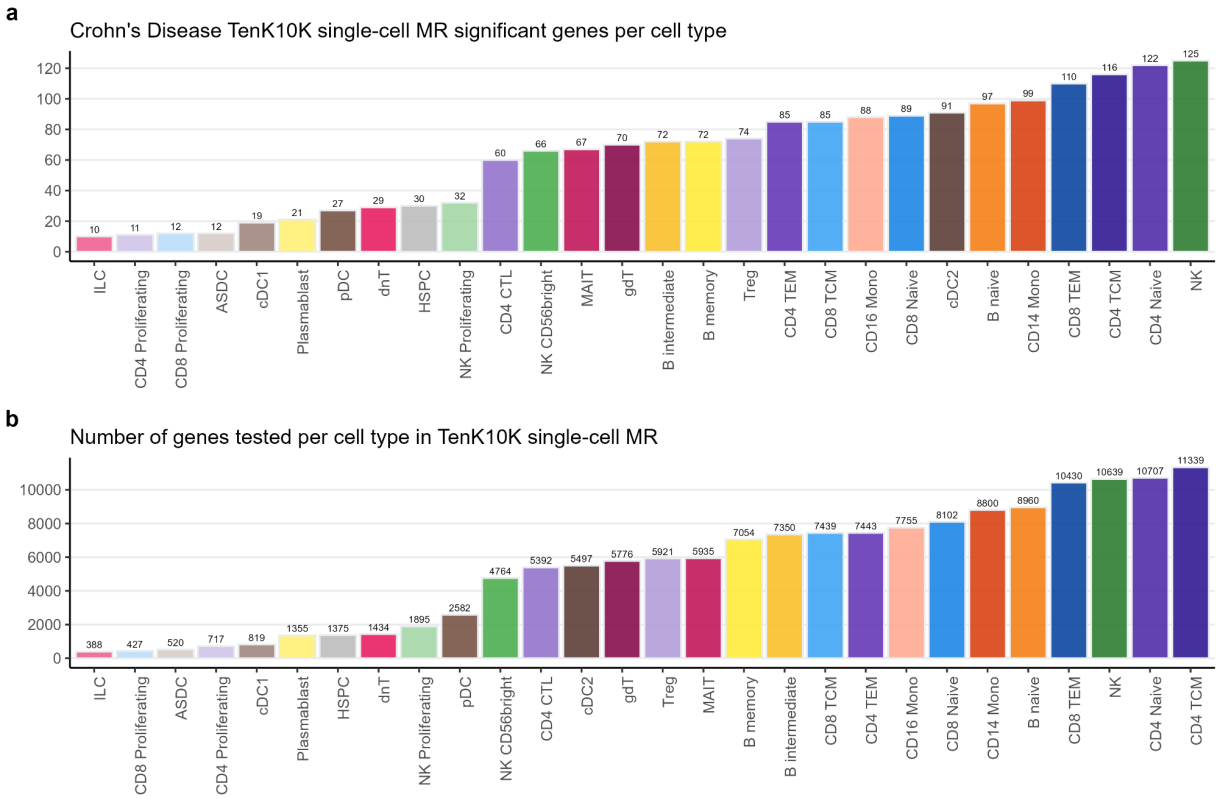
Supplementary Figure 10. Distribution of single-cell disease relevance score (scDRS) for Crohn's disease across two sampling strategies: 1) left panel: per-cell-type proportional sampling of 10% of cells analysed in the TenK10K phase 1 study, and 2) right panel: maximum of 10,000 cells per cell type (as used in the main analysis presented in the manuscript).



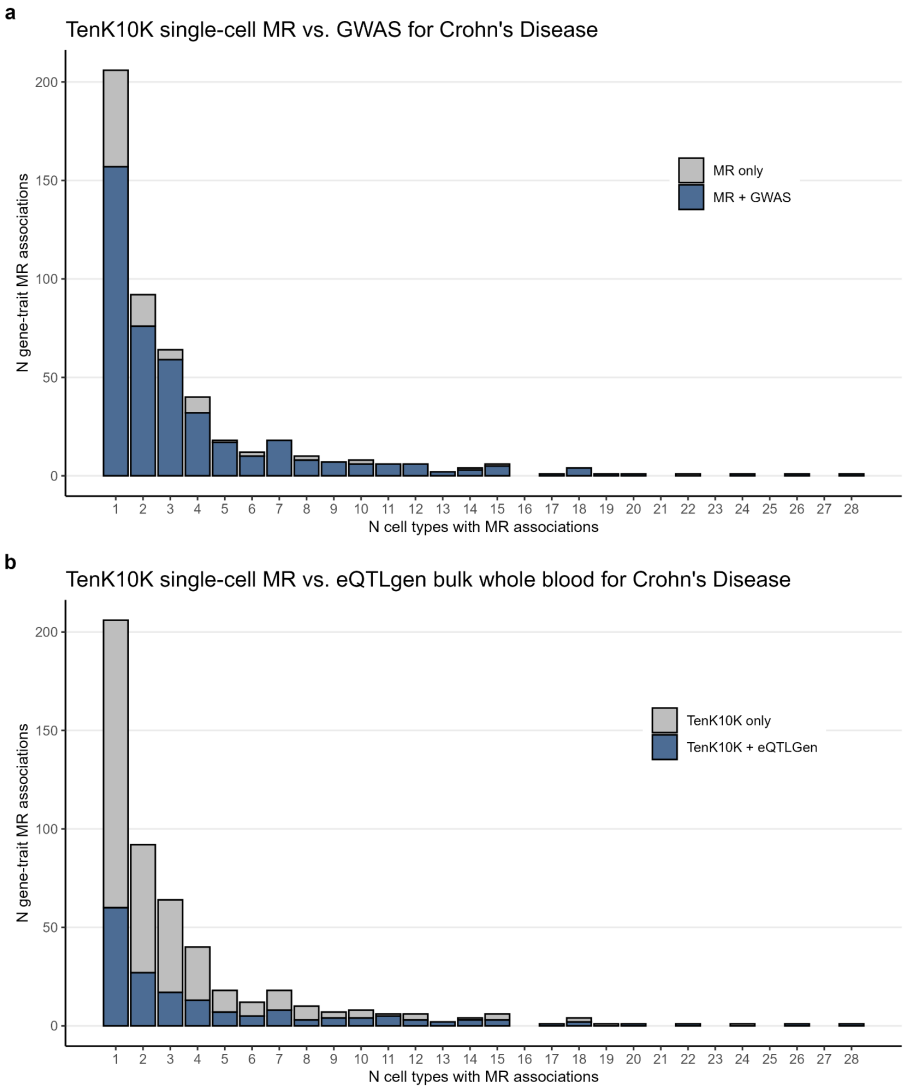
Supplementary Figure 11. Proportion of cells showing polygenic enrichment for Crohn's disease (cell-level Monte-Carlo *P* value) < 0.05 derived from single-cell disease relevance score (scDRS) across two sampling strategies: 1) left panel: per-cell-type proportional sampling of 10% of cells analysed in the TenK10K phase 1 study, and 2) right panel: maximum of 10,000 cells per cell type (as used in the main analysis presented in the manuscript).



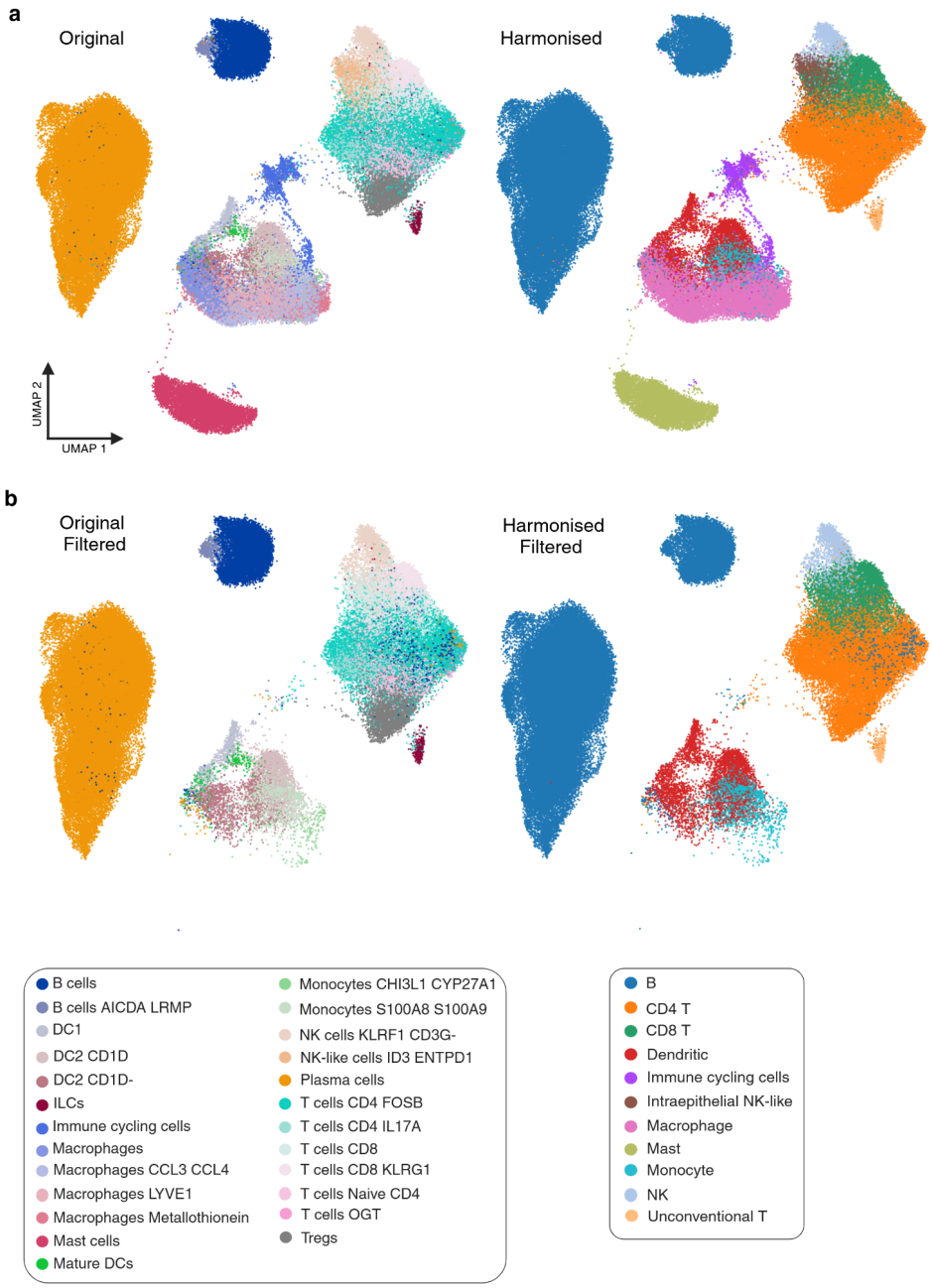
Supplementary Figure 12. Target-indication pairs in Open Targets Platform (version 25.06) with single-cell MR support.



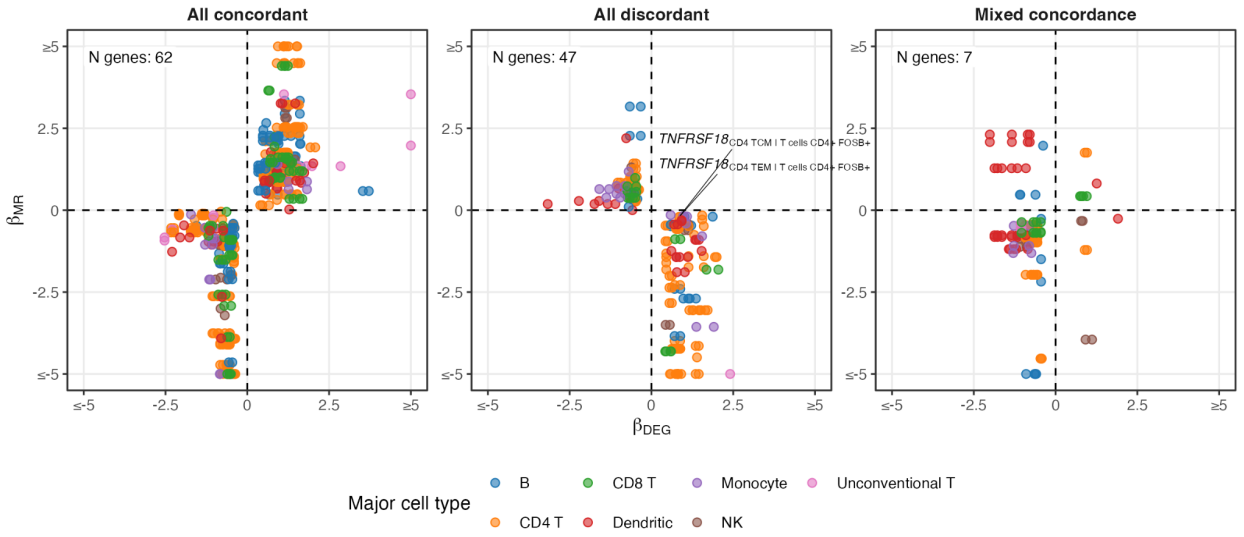
Supplementary Figure 13. Summary of cell type-resolved MR results for Crohn's disease **a)** Number of genes showing putative causal associations with Crohn's disease ($Q_{MR} < 0.05$; $P_{HEIDI} \geq 0.01$) across cell types. **b)** Total number of tested genes across cell types.



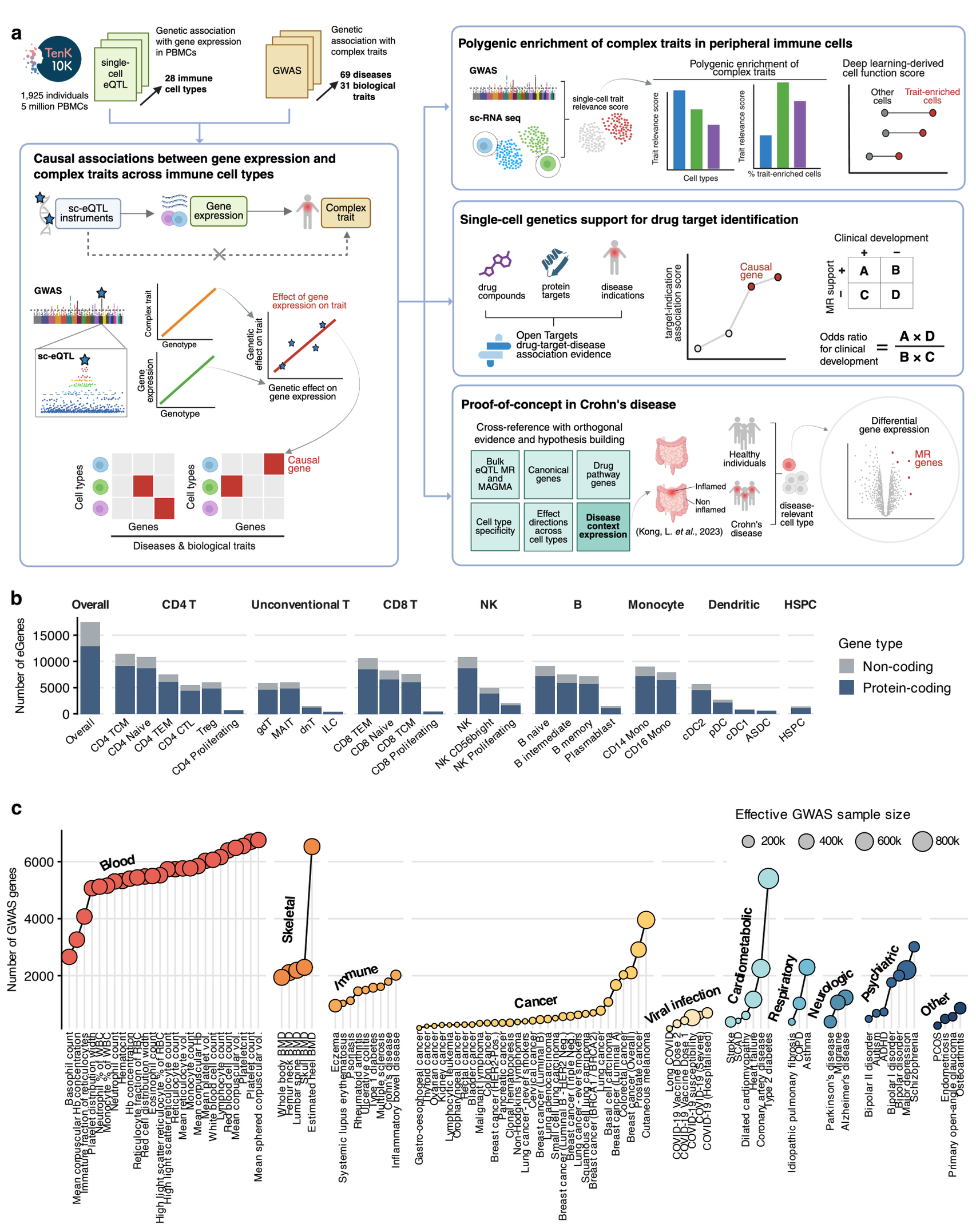
Supplementary Figure 14. Comparison of TenK10K single-cell MR with GWAS and eQTLgen bulk MR for Crohn's disease. Number of significant MR genes grouped by the number of cell types they are significant in. This is compared with the number of genes that are also significant in a) MAGMA, a GWAS-only based gene prioritisation method and b) bulk whole blood eQTLgen eQTL MR using the same Crohn's GWAS summary statistics.

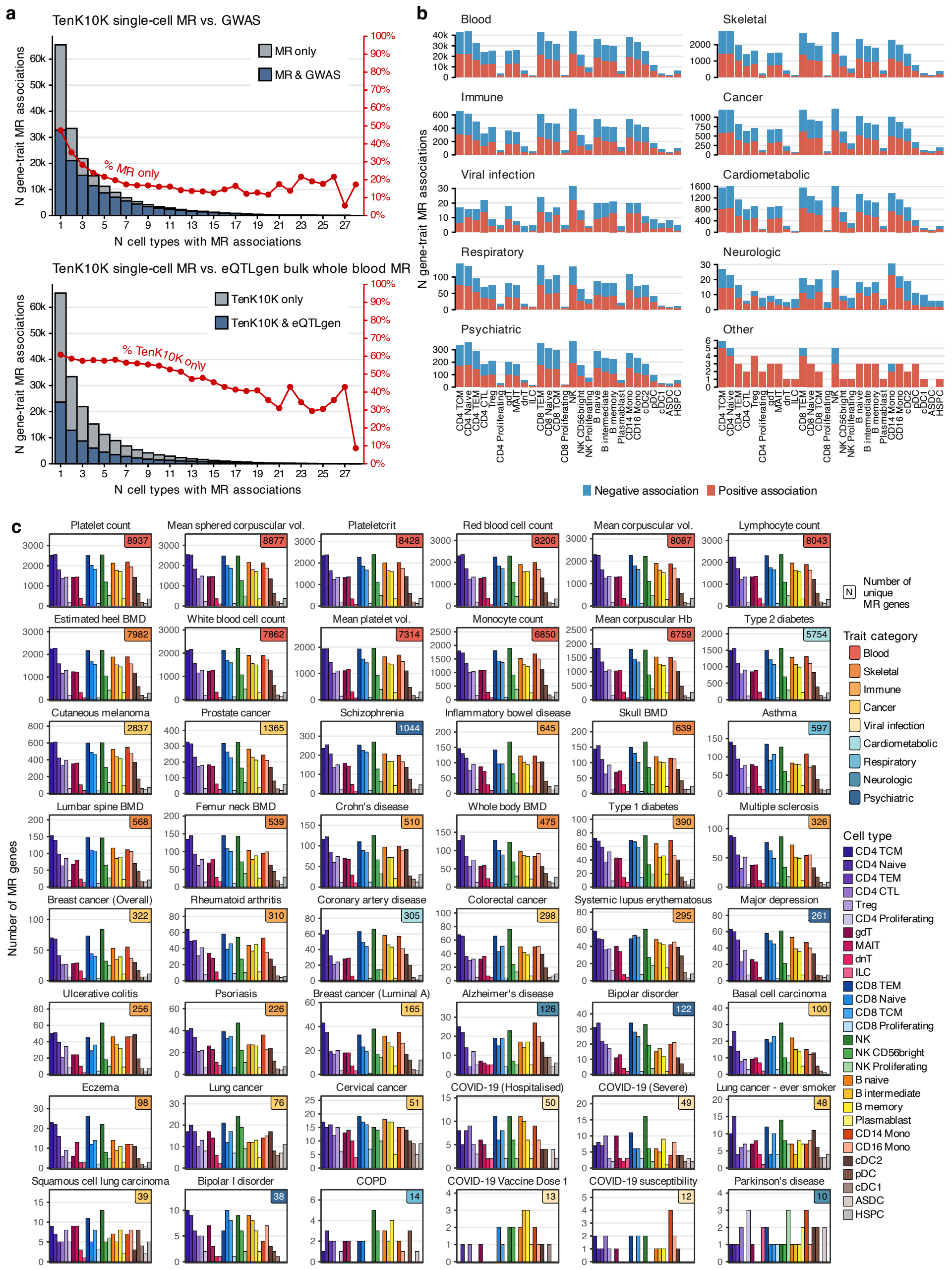


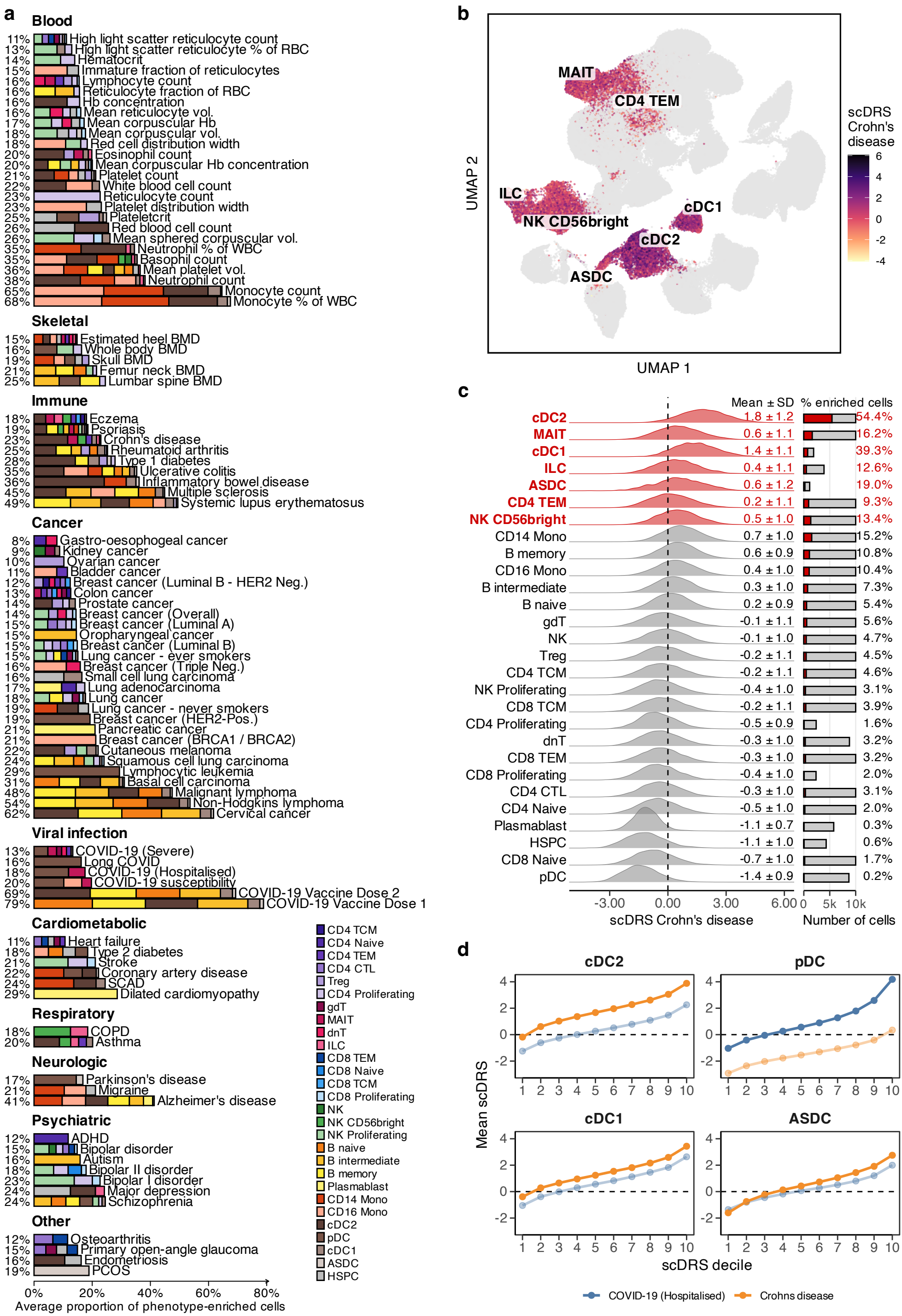
Supplementary Figure 15. Harmonised cell type annotation between MR and DEG**. a)** Left: The original annotation for the Crohn's colon immune cell dataset. Right: Harmonised annotation based on major cell types (**Supplementary Table 17**). **b)** Filtered version of (a) based on matched major cell types between MR and DEG. The cell types filtered out are macrophages, immune cycling cells, intraepithelial NK-like cells, and mast cells.

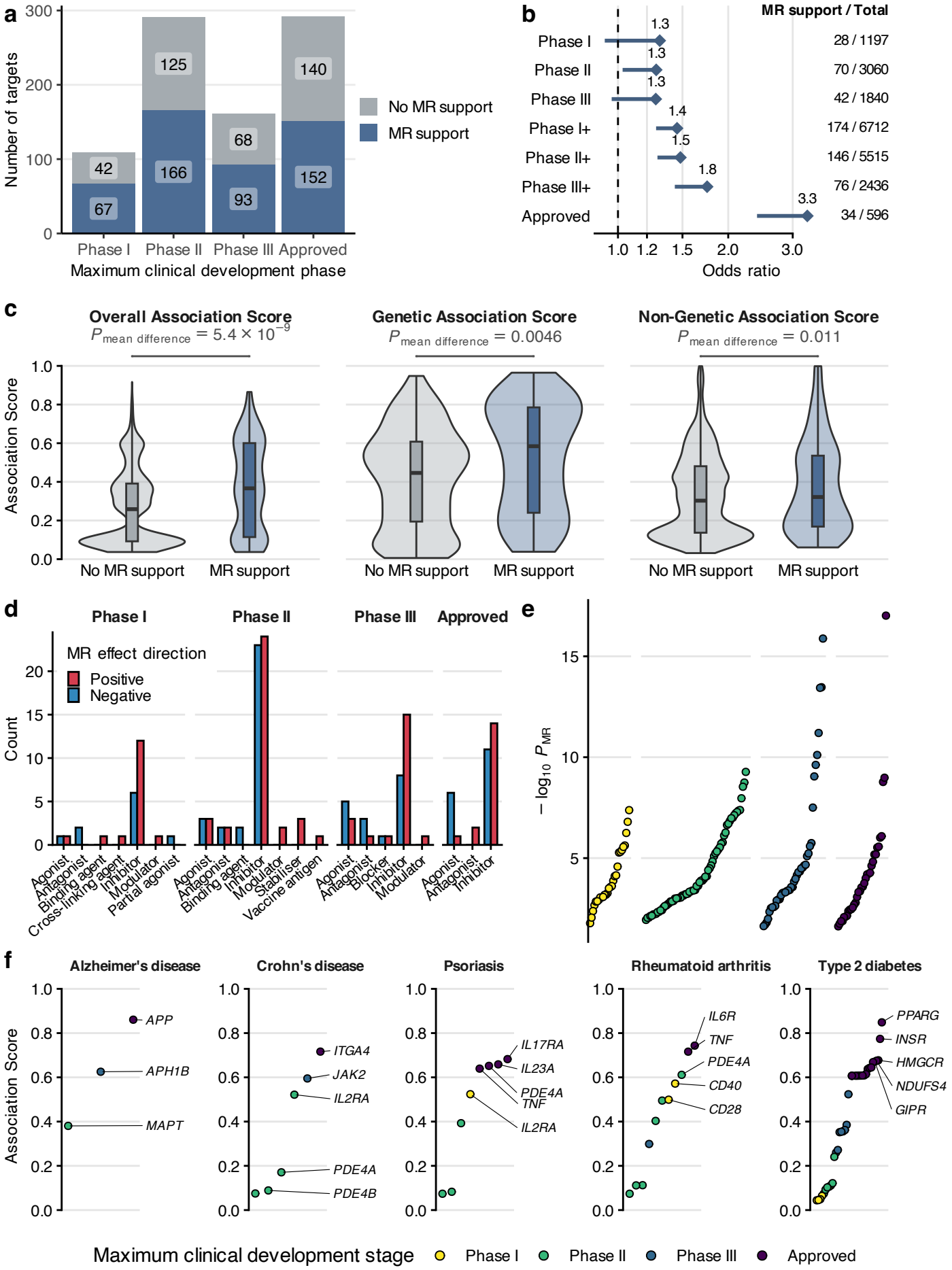


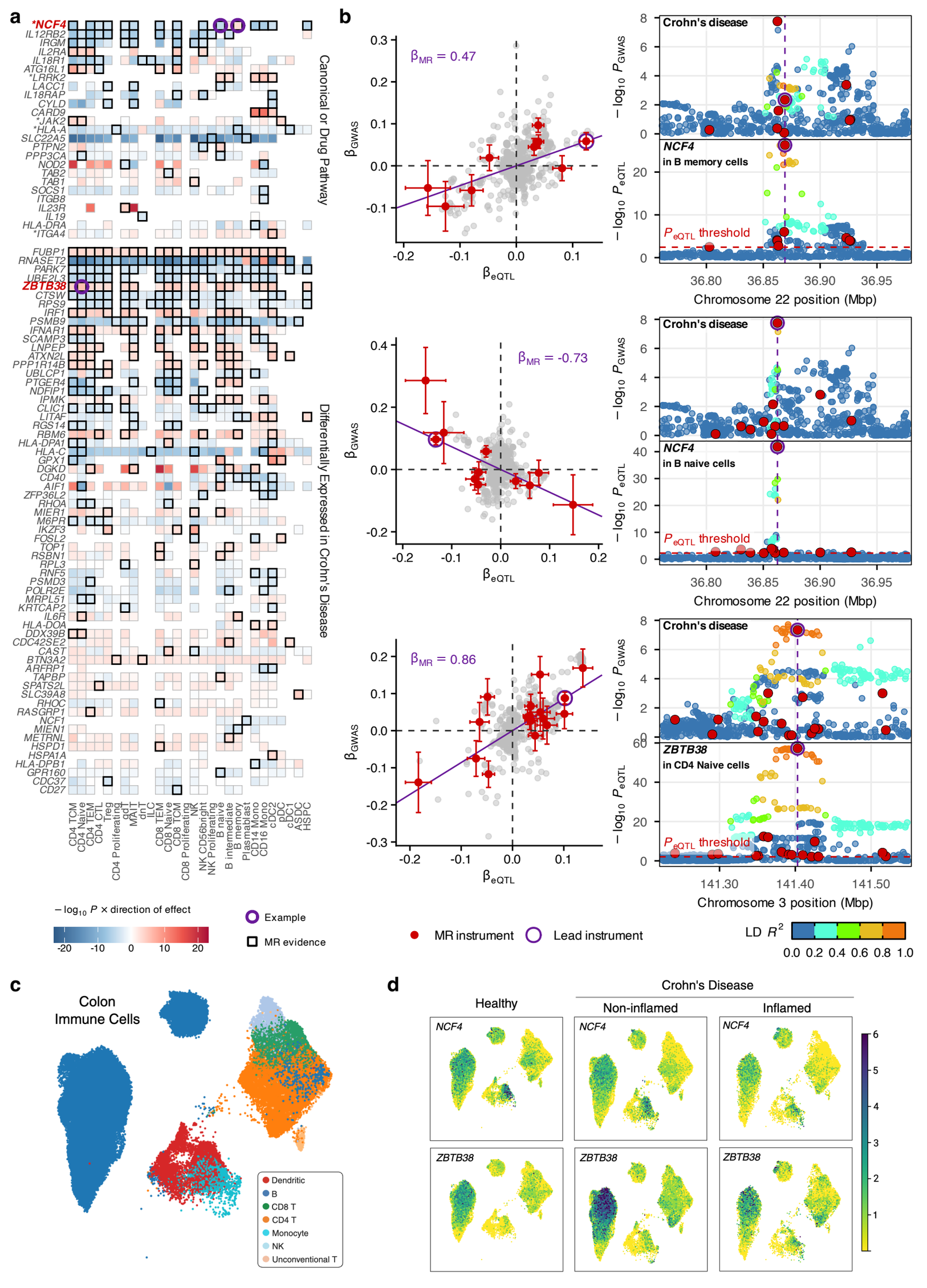
Supplementary Figure 16. Concordance of direction of effect in MR and DEG for genes with matched major cell type pairs. Each dot represents a gene-MR cell type-DEG cell type set, which belongs to a single matched major cell type between studies. *TNFRSF18* is highlighted as an example, labelled by 'MR cell type | DEG cell type', under the CD4 T major cell type group. 'All concordant' corresponds to cases where every gene-MR cell type-DEG cell type set has a matched direction of effect between MR and DEG. 'All discordant' corresponds to cases where every gene-MR cell type-DEG cell type set has an unmatched direction of effect between MR and DEG. 'Mixed concordance' corresponds to genes which have discordant directions between or within major cell types (see **Methods**, **Supplementary Table 14**).


a
TenK 10K
1,925 individuals
5 million PBMCs
single-cell eQTL
Genetic association with gene expression in PBMCs
28 immune cell types
GWAS
Genetic association with complex traits
69 diseases
31 biological traits
Causal associations between gene expression and complex traits across immune cell types
sc-eQTL instruments
Gene expression
Complex trait
GWAS
sc-eQTL
Complex trait
Genotype
Gene expression
Genotype
Effect of gene expression on trait
Genetic effect on trait
Genetic effect on gene expression
Cell types
Genes
Causal gene
Diseases & biological traits
Polygenic enrichment of complex traits in peripheral immune cells
GWAS
sc-RNA seq
single-cell trait relevance score
Polygenic enrichment of complex traits
Trait relevance score
Cell types
% trait-enriched cells
Deep learning-derived cell function score
Other cells
Trait-enriched cells
Single-cell genetics support for drug target identification
drug compounds
protein targets
disease indications
Open Targets drug-target-disease association evidence
target-indication association score
Causal gene
Clinical development
MR support
Odds ratio for clinical development
Proof-of-concept in Crohn's disease
Cross-reference with orthogonal evidence and hypothesis building
Bulk eQTL MR and MAGMA
Canonical genes
Drug pathway genes
Cell type specificity
Effect directions across cell types
Disease context expression
(Kong, L. et al., 2023)
Inflamed
Non inflamed
Crohn's disease
Healthy individuals
disease-relevant cell type
Differential gene expression
MR genes
b
Overall
CD4 T
Unconventional T
CD8 T
NK
B
Monocyte
Dendritic
HSPC
Number of eGenes
Gene type
Non-coding
Protein-coding
Overall
CD4 TCM
CD4 Naive
CD4 TEM
CD4 CTL
Treg
CD4 Proliferating
gdT
MAIT
dnT
ILC
CD8 TEM
CD8 Naive
CD8 TCM
CD8 Proliferating
NK
NK CD56bright
NK Proliferating
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC2
pDC
cDC1
ASDC
HSPC
c
Effective GWAS sample size
200k
400k
600k
800k
Number of GWAS genes
Blood
Skeletal
Immune
Cancer
Viral infection
Cardiometabolic
Respiratory
Neurologic
Psychiatric
Other


a
TenK10K single-cell MR vs. GWAS
MR only
MR & GWAS
% MR only
N gene-trait MR associations
N cell types with MR associations
TenK10K single-cell MR vs. eQTLgen bulk whole blood MR
TenK10K only
TenK10K & eQTLgen
% TenK10K only
b
Blood
Skeletal
Immune
Cancer
Viral infection
Cardiometabolic
Respiratory
Neurologic
Psychiatric
Other
Negative association
Positive association
c
Number of MR genes
Number of unique MR genes
Trait category
Blood
Skeletal
Immune
Cancer
Viral infection
Cardiometabolic
Respiratory
Neurologic
Psychiatric
Cell type
CD4 TCM
CD4 Naive
CD4 TEM
CD4 CTL
Treg
CD4 Proliferating
gdT
MAIT
dnT
ILC
CD8 TEM
CD8 Naive
CD8 TCM
CD8 Proliferating
NK
NK CD56bright
NK Proliferating
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC2
pDC
cDC1
ASDC
HSPC
Platelet count 8937
Mean sphered corpuscular vol. 8877
Plateletcrit 8428
Red blood cell count 8206
Mean corpuscular vol. 8087
Lymphocyte count 8043
Estimated heel BMD 7982
White blood cell count 7862
Mean platelet vol. 7314
Monocyte count 6850
Mean corpuscular Hb 6759
Type 2 diabetes 5754
Cutaneous melanoma 2837
Prostate cancer 1365
Schizophrenia 1044
Inflammatory bowel disease 645
Skull BMD 639
Asthma 597
Lumbar spine BMD 568
Femur neck BMD 539
Crohn's disease 510
Whole body BMD 475
Type 1 diabetes 390
Multiple sclerosis 326
Breast cancer (Overall) 322
Rheumatoid arthritis 310
Coronary artery disease 305
Colorectal cancer 298
Systemic lupus erythematosus 295
Major depression 261
Ulcerative colitis 256
Psoriasis 226
Breast cancer (Luminal A) 165
Alzheimer's disease 126
Bipolar disorder 122
Basal cell carcinoma 100
Eczema 98
Lung cancer 76
Cervical cancer 51
COVID-19 (Hospitalised) 50
COVID-19 (Severe) 49
Lung cancer - ever smoker 48
Squamous cell lung carcinoma 39
Bipolar I disorder 38
COPD 14
COVID-19 Vaccine Dose 1 13
COVID-19 susceptibility 12
Parkinson's disease 10


a
Blood
11% High light scatter reticulocyte count
13% High light scatter reticulocyte % of RBC
14% Hematocrit
15% Immature fraction of reticulocytes
16% Lymphocyte count
16% Reticulocyte fraction of RBC
16% Hb concentration
16% Mean reticulocyte vol.
17% Mean corpuscular Hb
18% Mean corpuscular vol.
18% Red cell distribution width
20% Eosinophil count
20% Mean corpuscular Hb concentration
21% Platelet count
22% White blood cell count
23% Reticulocyte count
23% Platelet distribution width
25% Plateletcrit
26% Red blood cell count
26% Mean sphered corpuscular vol.
35% Neutrophil % of WBC
35% Basophil count
36% Mean platelet vol.
38% Neutrophil count
65% Monocyte count
68% Monocyte % of WBC
Skeletal
15% Estimated heel BMD
16% Whole body BMD
19% Skull BMD
21% Femur neck BMD
25% Lumbar spine BMD
Immune
18% Eczema
19% Psoriasis
23% Crohn's disease
25% Rheumatoid arthritis
28% Type 1 diabetes
35% Ulcerative colitis
36% Inflammatory bowel disease
45% Multiple sclerosis
49% Systemic lupus erythematosus
Cancer
8% Gastro-oesophogeal cancer
9% Kidney cancer
10% Ovarian cancer
11% Bladder cancer
12% Breast cancer (Luminal B - HER2 Neg.)
13% Colon cancer
14% Prostate cancer
14% Breast cancer (Overall)
15% Breast cancer (Luminal A)
15% Oropharyngeal cancer
15% Breast cancer (Luminal B)
15% Lung cancer - ever smokers
16% Breast cancer (Triple Neg.)
16% Small cell lung carcinoma
17% Lung adenocarcinoma
18% Lung cancer
19% Lung cancer - never smokers
19% Breast cancer (HER2-Pos.)
21% Pancreatic cancer
21% Breast cancer (BRCA1 / BRCA2)
22% Cutaneous melanoma
24% Squamous cell lung carcinoma
29% Lymphocytic leukemia
31% Basal cell carcinoma
48% Malignant lymphoma
54% Non-Hodgkins lymphoma
62% Cervical cancer
Viral infection
13% COVID-19 (Severe)
16% Long COVID
18% COVID-19 (Hospitalised)
20% COVID-19 susceptibility
69% COVID-19 Vaccine Dose 2
79% COVID-19 Vaccine Dose 1
Cardiometabolic
11% Heart failure
18% Type 2 diabetes
21% Stroke
22% Coronary artery disease
24% SCAD
29% Dilated cardiomyopathy
Respiratory
18% COPD
20% Asthma
Neurologic
17% Parkinson's disease
21% Migraine
41% Alzheimer's disease
Psychiatric
12% ADHD
15% Bipolar disorder
16% Autism
18% Bipolar II disorder
23% Bipolar I disorder
24% Major depression
24% Schizophrenia
Other
12% Osteoarthritis
15% Primary open-angle glaucoma
16% Endometriosis
19% PCOS
0% 20% 40% 60% 80%
Average proportion of phenotype-enriched cells
CD4 TCM
CD4 Naive
CD4 TEM
CD4 CTL
Treg
CD4 Proliferating
gdT
MAIT
dnT
ILC
CD8 TEM
CD8 Naive
CD8 TCM
CD8 Proliferating
NK
NK CD56bright
NK Proliferating
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC2
pDC
cDC1
ASDC
HSPC
b
MAIT
CD4 TEM
ILC
NK CD56bright
cDC1
cDC2
ASDC
UMAP 2
UMAP 1
scDRS Crohn's disease
6
4
2
0
-2
-4
c
Mean ± SD
% enriched cells
cDC2 1.8 ± 1.2 54.4%
MAIT 0.6 ± 1.1 16.2%
cDC1 1.4 ± 1.1 39.3%
ILC 0.4 ± 1.1 12.6%
ASDC 0.6 ± 1.2 19.0%
CD4 TEM 0.2 ± 1.1 9.3%
NK CD56bright 0.5 ± 1.0 13.4%
CD14 Mono 0.7 ± 1.0 15.2%
B memory 0.6 ± 0.9 10.8%
CD16 Mono 0.4 ± 1.0 10.4%
B intermediate 0.3 ± 1.0 7.3%
B naive 0.2 ± 0.9 5.4%
gdT -0.1 ± 1.1 5.6%
NK -0.1 ± 1.0 4.7%
Treg -0.2 ± 1.1 4.5%
CD4 TCM -0.2 ± 1.1 4.6%
NK Proliferating -0.4 ± 1.0 3.1%
CD8 TCM -0.2 ± 1.1 3.9%
CD4 Proliferating -0.5 ± 0.9 1.6%
dnT -0.3 ± 1.0 3.2%
CD8 TEM -0.3 ± 1.0 3.2%
CD8 Proliferating -0.4 ± 1.0 2.0%
CD4 CTL -0.3 ± 1.0 3.1%
CD4 Naive -0.5 ± 1.0 2.0%
Plasmablast -1.1 ± 0.7 0.3%
HSPC -1.1 ± 1.0 0.6%
CD8 Naive -0.7 ± 1.0 1.7%
pDC -1.4 ± 0.9 0.2%
-3.00 0.00 3.00 6.00
scDRS Crohn's disease
0 5k 10k
Number of cells
d
cDC2
pDC
cDC1
ASDC
Mean scDRS
4
2
0
-2
1 2 3 4 5 6 7 8 9 10
scDRS decile
COVID-19 (Hospitalised)
Crohns disease


a
Number of targets
300
200
100
0
42
67
125
166
68
93
140
152
No MR support
MR support
Phase I
Phase II
Phase III
Approved
Maximum clinical development phase
b
MR support / Total
Phase I
Phase II
Phase III
Phase I+
Phase II+
Phase III+
Approved
1.3
1.3
1.3
1.4
1.5
1.8
3.3
28 / 1197
70 / 3060
42 / 1840
174 / 6712
146 / 5515
76 / 2436
34 / 596
1.0
1.2
1.5
2.0
3.0
Odds ratio
c
Overall Association Score
$P_{\text{mean difference}} = 5.4 \times 10^{-9}$
Genetic Association Score
$P_{\text{mean difference}} = 0.0046$
Non-Genetic Association Score
$P_{\text{mean difference}} = 0.011$
Association Score
No MR support
MR support
d
Phase I
Phase II
Phase III
Approved
MR effect direction
Positive
Negative
Count
Agonist
Antagonist
Binding agent
Cross-linking agent
Inhibitor
Modulator
Partial agonist
Stabiliser
Vaccine antigen
Blocker
e
$-\log_{10} P_{MR}$
f
Alzheimer's disease
Crohn's disease
Psoriasis
Rheumatoid arthritis
Type 2 diabetes
APP
APH1B
MAPT
ITGA4
JAK2
IL2RA
PDE4A
PDE4B
IL17RA
IL23A
PDE4A
TNF
IL2RA
IL6R
TNF
PDE4A
CD40
CD28
PPARG
INSR
HMGCR
NDUFS4
GIPR
Maximum clinical development stage
Phase I
Phase II
Phase III
Approved


a
*NCF4
IL12RB2
IRGM
IL2RA
IL18R1
ATG16L1
*LRRK2
LACC1
IL18RAP
CYLD
CARD9
*JAK2
*HLA-A
SLC22A5
PTPN2
PPP3CA
NOD2
TAB2
TAB1
SOCS1
ITGB8
IL23R
IL19
HLA-DRA
*ITGA4
Canonical or Drug Pathway
FUBP1
RNASET2
PARK7
UBE2L3
ZBTB38
CTSW
RPS9
IRF1
PSMB9
IFNAR1
SCAMP3
LNPEP
ATXN2L
PPP1R14B
UBLCP1
PTGER4
NDFIP1
IPMK
CLIC1
LITAF
RGS14
RBM6
HLA-DPA1
HLA-C
GPX1
DGKD
CD40
AIF1
ZFP36L2
RHOA
MIER1
M6PR
IKZF3
FOSL2
TOP1
RSBN1
RPL3
RNF5
PSMD3
POLR2E
MRPL51
KRTCAP2
IL6R
HLA-DOA
DDX39B
CDC42SE2
CAST
BTN3A2
ARFRP1
TAPBP
SPATS2L
SLC39A8
RHOC
RASGRP1
NCF1
MIEN1
METRNL
HSPD1
HSPA1A
HLA-DPB1
GPR160
CDC37
CD27
Differentially Expressed in Crohn's Disease
CD4 TCM
CD4 Naive
CD4 TEM
CD4 CTL
Treg
CD4 Proliferating
gdT
MAIT
dnT
ILC
CD8 TEM
CD8 Naive
CD8 TCM
CD8 Proliferating
NK
NK CD56bright
NK Proliferating
B naive
B intermediate
B memory
Plasmablast
CD14 Mono
CD16 Mono
cDC2
pDC
cDC1
ASDC
HSPC
−log10 P × direction of effect
-20 -10 0 10 20
Example
MR evidence
b
βMR = 0.47
βGWAS
βeQTL
Crohn's disease
−log10 PGWAS
NCF4
in B memory cells
−log10 PeQTL
PeQTL threshold
Chromosome 22 position (Mbp)
βMR = -0.73
NCF4
in B naive cells
βMR = 0.86
ZBTB38
in CD4 Naive cells
Chromosome 3 position (Mbp)
MR instrument
Lead instrument
LD R2
0.0 0.2 0.4 0.6 0.8 1.0
c
Colon
Immune Cells
Dendritic
B
CD8 T
CD4 T
Monocyte
NK
Unconventional T
d
Healthy
Crohn's Disease
Non-inflamed
Inflamed
NCF4
ZBTB38