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 674,764 causal associations across 16,078 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 17,444 eGenes (including 12,873 protein-coding genes) across 28 immune cell types, with a median of 5,947 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 15,101,807 gene–cell type–trait combinations, of which 732,822, (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, 674,764 causal associations were identified from combinations of 16,078 unique genes (12,076 protein-coding), 31 traits, and 28 cell types. The median number of unique genes with causal associations was 4,883 per cell type, 6,077 per trait, and 788 per cell type-trait pair (**Supplementary Figure 2, Supplementary Table 3**). Across the 28 immune cell types, the number of significant MR gene-trait associations ranged from 2,088 for ILC cells to 50,692 for NK cells (median = 28,954). 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 = 5.0%, 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 188,634 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 188,634 unique gene-trait associations identified by single-cell MR, 65,355 (35%) were not identified by using MAGMA with GWAS data only, with a median of 38 (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 114,631 (61%) gene-trait associations were not identified (median = 178 / 65% 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.

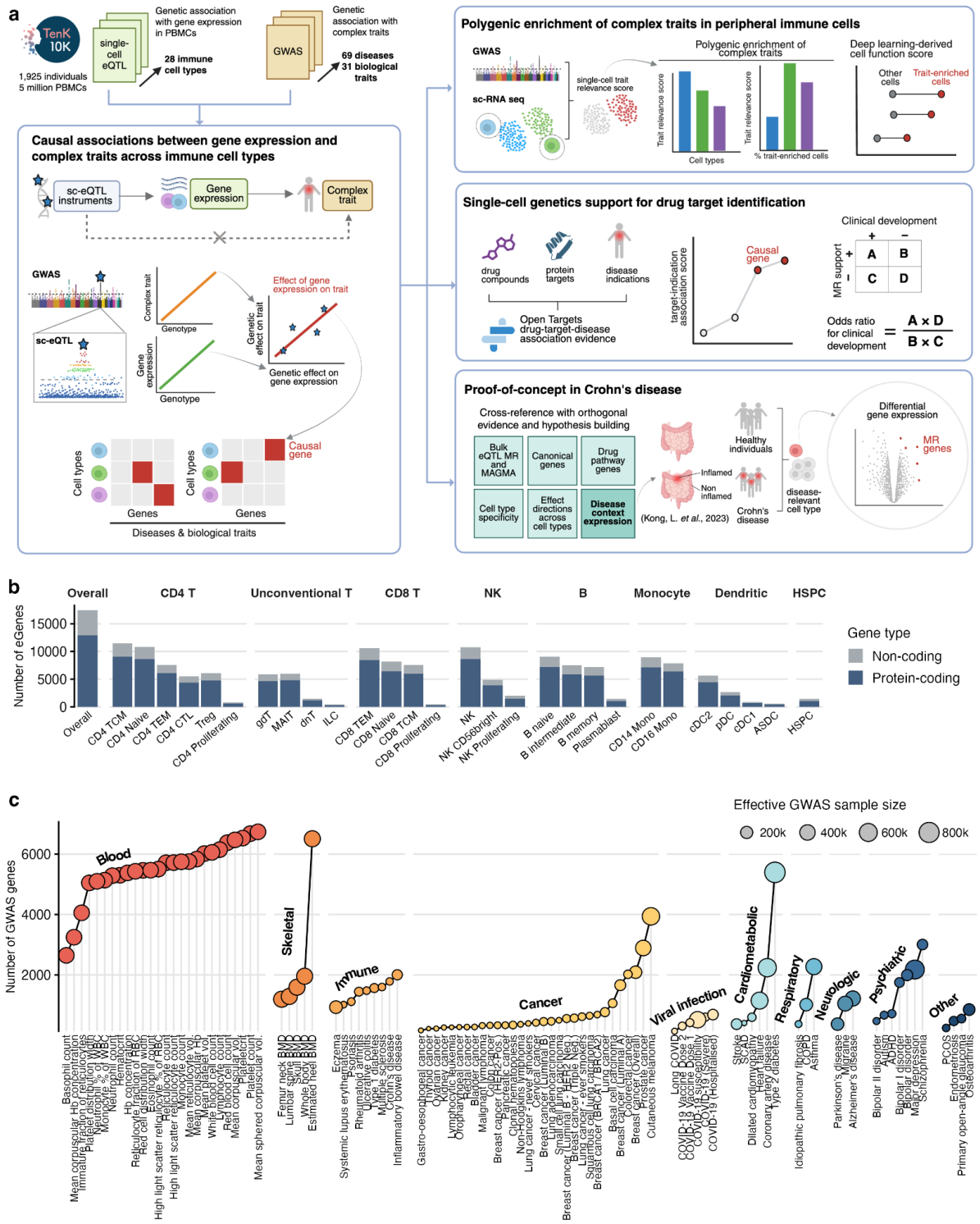


Figure 1 | Study design and causal inference using single-cell eQTLs.
a) Overview of study design. We integrated single-cell eQTL data from 5 million peripheral blood mononuclear cells (PBMCs) (covering 28 immune cell types) profiled in the TenK10K Phase 1 cohort (1,925 donors) with GWAS summary statistics 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 5,947 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).

172 In addition to identifying causal genes, MR also helps in resolving the causal effect
173 direction on trait outcomes, providing critical information to distinguish between targeting
174 upregulation and downregulation of gene expression in drug design. Across all significant MR
175 associations in our study, the direction of effect was evenly split between positive (*n*=365,985)
176 and negative (*n*=366,837), indicating a balanced contribution of activating and repressing
177 regulatory effects on trait outcomes (**Figure 2b**). This symmetry was consistent across both
178 phenotype categories and cell types, implicating that disease-relevant gene networks involve a
179 balance of activation and inhibition.

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

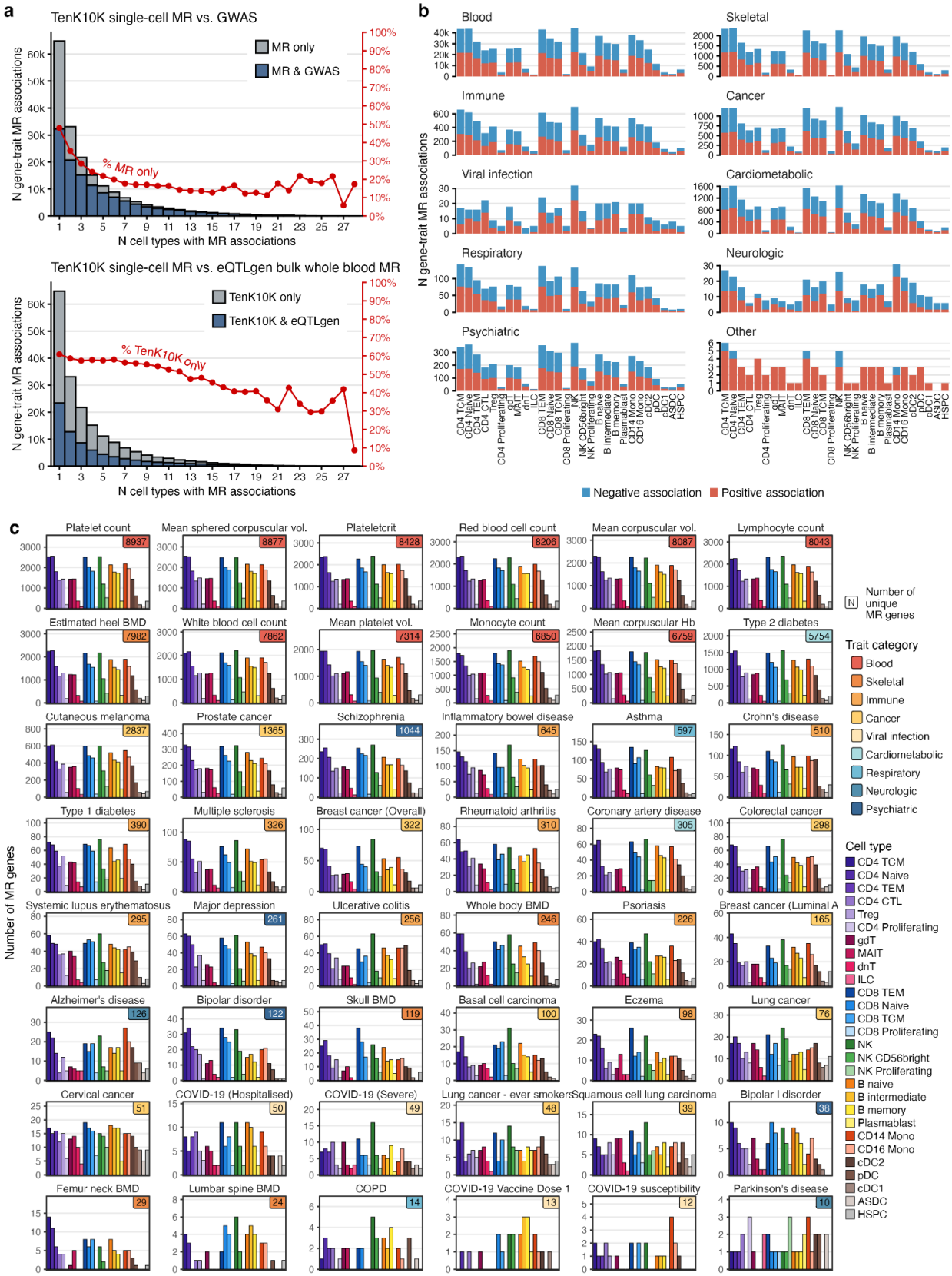


Figure 2 | Cell type-specific causal gene-trait associations across complex traits.
a) Distribution of gene-trait associations with MR evidence by the number of cell types in which the association is detected. Left y-axis: total number of gene–trait pairs (bars). Colours distinguish associations identified 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 (39 traits), cDC2 (37 traits), and CD4 proliferating T cells (31 traits) showing the strongest polygenic enrichment (**Supplementary Figure 7**). In total, 95 of 100 traits were enriched in at least one cell type (mean = 4.3, max = 10). Immune-mediated diseases showed the highest number of enriched cell types, as expected; for example, psoriasis 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 = 18.7%) (**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 from enriched cell types 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.

241 Using Crohn's disease as a case study, we identified 12,199 out of 56,920 sampled cells
242 (21%) as enriched for polygenic signals across eight cell types (**Figure 3b**). The most enriched
243 cell types, ranked by Z_{scDRS} statistics, were cDC2, MAIT, cDC1, NK CD56^{bright}, ASDC, CD14
244 Monocytes, and CD4 TEM ILC cells. Enrichment proportions ranged from 9.2% in CD4+
245 TEM to 54.3% in cDC2 (**Figure 3c**). Dendritic cells, particularly cDC2 and cDC1, showed
246 the strongest enrichment. These results are consistent with prior reports of dendritic cell
247 infiltration and activation in Crohn's-affected intestinal tissue[18], including overexpression of
248 chemokine receptors involved in T cell recruitment. We also observed enrichment in ASDCs,
249 a less-characterised dendritic subset with unclear roles in autoimmunity[19]. In contrast,
250 plasmacytoid dendritic cells (pDCs) did not show enrichment for Crohn's disease but were the
251 only dendritic subtype enriched for COVID-19 hospitalisation (**Figure 3d)**, consistent with
252 their established role in antiviral immunity and type I interferon production[20].

253 The polygenic enrichment analysis also facilitates the identification of trait-related cell
254 functions through integration with scDeepID, an in-house deep learning method to quantify
255 cell function as cell states at single-cell resolution (see **Methods**). Using B cells in Systemic
256 Lupus Erythematosus (SLE) as an example, we leveraged scDeepID to compute 400
257 representative cell function scores in 539,820 B cells (including B naive, B intermediate, and
258 B memory cell types) identified in the TenK10K phase 1 study. We extracted cell function
259 scores for the 30,000 sampled B cells used in the scDRS analysis, and compared the average
260 cell function scores between 17,264 SLE-enriched cells (cell-level $P_{scDRS\text{-}SLE} < 0.05$) and
261 12,736 non-SLE-enriched cells using a one-sided *T*-test. We identified 202 cell functions with
262 a higher average score in the SLE-enriched cells compared to non-SLE-enriched cells at $Q <$
263 0.05 (**Supplementary Figure 8, Supplementary Table 6**). The cell functions with the highest
264 mean score difference based on *Q* value include cellular response to type II interferon, positive
265 regulation of inflammatory response, tumour necrosis factor-mediated signalling pathway,
266 biological process involved in interaction with host, and leukocyte-mediated cytotoxicity
267 (**Supplementary Figure 9a-b)**. The scores of these five cell functions also showed a trend of
268 linear increase across deciles of $scDRS_{SLE}$ (**Supplementary Figure 9c**), implicating
269 involvement of these pathways and their gene members in the pathogenesis of SLE.

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

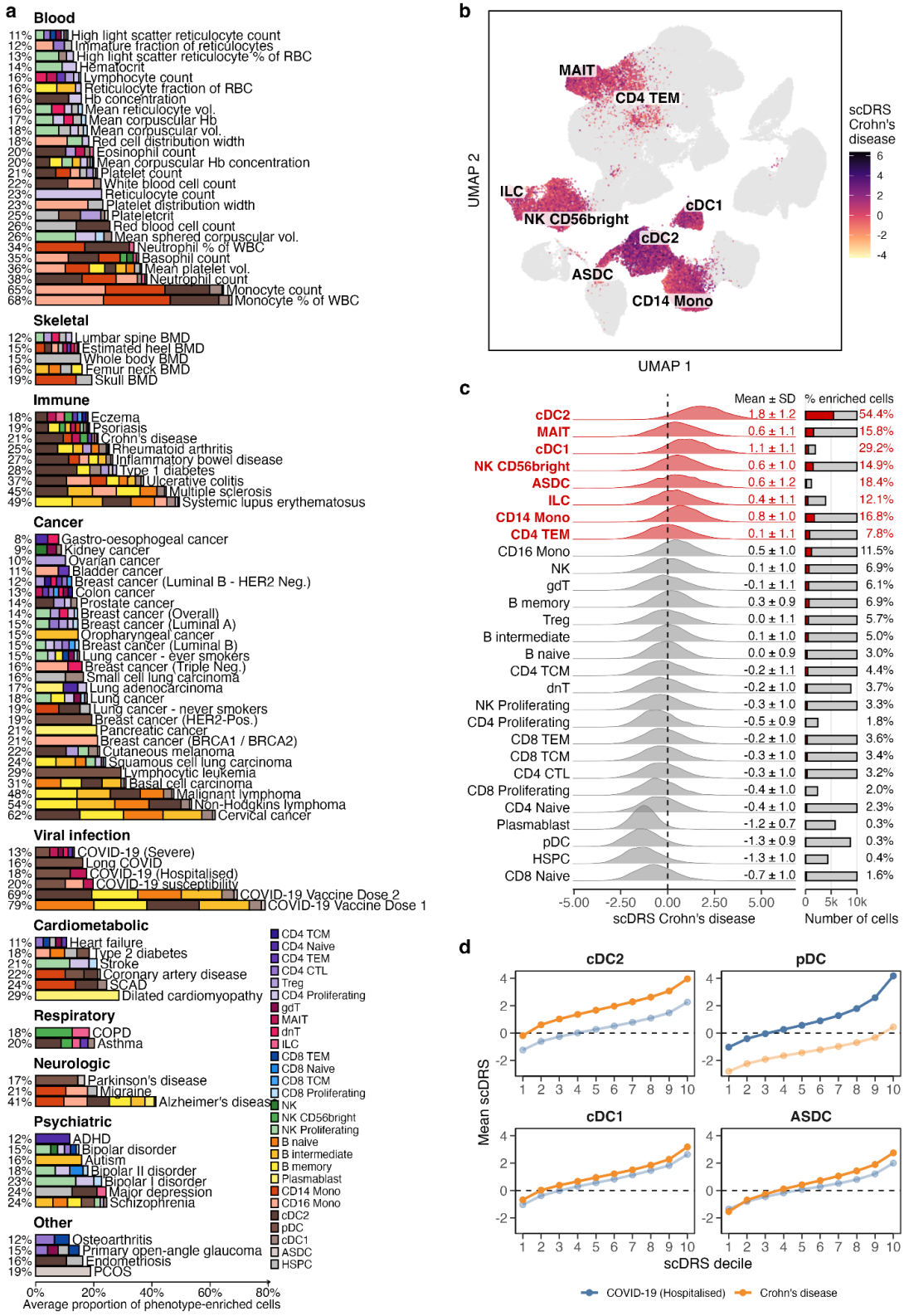


Figure 3 | Polygenic enrichment of complex traits across immune cell types. **a)** Average proportion of phenotype-enriched cells across 100 complex traits, stratified by immune cell types. Enrichment was calculated using the single-cell disease relevance score (scDRS); cells were considered enriched 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 $P_{scDRS} < 0.05$) 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 ($P_{scDRS} < 0.05$) 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 $P_{scDRS} < 0.05$). Cell types without evidence of enrichment ($P_{scDRS} \geq 0.05$) 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 pairs in both genetic evidence (mean = 0.52 vs. 0.42; $P = 7.7 \times 10^{-3}$) and non-genetic sources (mean = 0.38 vs. 0.33; $P = 6.3 \times 10^{-3}$), with a more pronounced effect in the genetic category.

322 Next, we assessed the concordance between the direction of MR effect and the
323 mechanism of action (MoA) of drugs at the most advanced clinical development stage for each
324 target–indication pair (**Figure 4d, Supplementary Table 10**). On average, MR-supported
325 pairs more often showed concordant MoA, for instance, risk-increasing (positive effect) MR
326 results aligning with inhibitors, or protective (negative effect) results aligning with agonists.
327 This pattern was especially pronounced for drugs in Phase III trials or with regulatory approval.
328 Moreover, MR-supported targets in later development stages were associated with stronger
329 evidence of causal effect (**Figure 4e, Supplementary Table 10**).

330 Finally, we examined selected examples from the 174 MR-supported target–indication
331 pairs (covering 131 genes and 25 diseases) in more detail (**Figure 4f, Supplementary Figure**
332 **12, Supplementary Table 10**). These included several known immune targets of approved
333 therapies: JAK2 and TNF for Crohn's disease, IL6R for rheumatoid arthritis, and IL23A and
334 IL17RA for psoriasis. Importantly, our analysis also identified high-confidence targets in
335 diseases not classically associated with immune dysfunction, including APP (amyloid
336 precursor protein) for Alzheimer's disease and GLP1R (glucagon-like peptide-1 receptor) for
337 type 2 diabetes.

338 APP encodes a precursor to Amyloid-β peptides, whose aggregation forms amyloid
339 plaques linked to synaptic dysfunction and cognitive decline in Alzheimer's disease[26]. GLP-
340 1R, a receptor primarily expressed in the pancreas and brain, modulates insulin secretion and
341 glucose metabolism through binding to GLP-1 proteins[27]. Initially developed for diabetes,
342 GLP-1R agonists (e.g., semaglutide) have recently shown efficacy in reducing cardiovascular
343 risk[28] and promoting weight loss[29] in individuals without diabetes. These examples highlight
344 how cell type-specific MR results can identify therapeutically actionable targets across
345 complex diseases including cases where the encoding genes are primarily expressed in other
346 tissues, potentially due to shared genetic regulation.

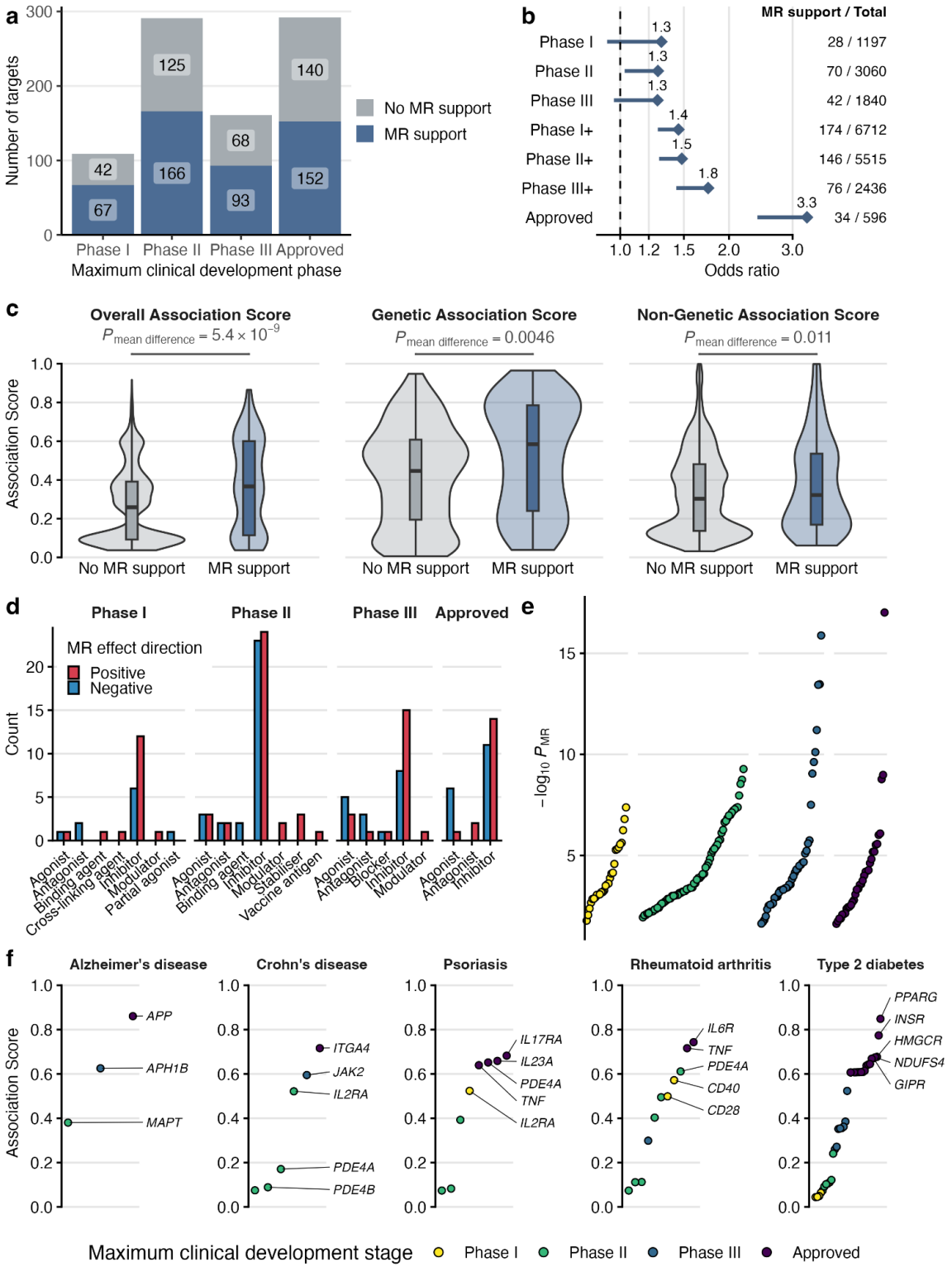


Figure 4 | Single-cell MR supports drug target validation and clinical development success. **a)** Number of protein-coding gene targets with MR support ($Q_{MR} < 0.05$; $P_{HEIDI} \geq 0.01$) across clinical development stages, as defined by target–indication pairs in the Open Targets Platform (OTP).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

355 side annotations indicate counts of MR-supported target–indication pairs over the total number of
356 clinically investigated pairs per phase. **c)** Distribution of OTP evidence scores for target–indication
357 pairs with and without MR support. Scores are shown for all evidence sources (overall), as well as
358 genetic and non-genetic sources separately. **d)** Number of MR-supported target–indication pairs,
359 stratified by predicted direction of MR effect (positive or negative) and mechanism of action (MoA) of
360 drugs with the highest OTP evidence score. Data are grouped by maximum clinical development stage.
361 **e)** Strength of MR evidence ($-\log_{10} P_{MR}$) for target–indication pairs across clinical development stages.
362 **f)** MR-supported targets for five selected diseases. Each bar represents a target–indication pair, ordered
363 by OTP overall association score and coloured by maximum clinical development phase. The top five
364 targets by OTP score are labelled.
365

366 Single-cell genetics reveals causal cell-type-specific mechanisms in 367 Crohn's disease

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

373 Using our single-cell MR framework, we identified 1,791 gene-cell type pairs and 510
374 unique genes (407 protein coding) with evidence of a causal relationship with Crohn's disease
375 across 28 immune cell types (**Supplementary Figure 13**). Of these, 88 were not detected using
376 MAGMA-based gene prioritisation from GWAS data alone, and 339 were not identified using
377 bulk whole blood-based MR (**Supplementary Figure 14**). A total of 209 genes (163 protein-
378 coding) exhibited cell-type-specific MR associations, defined as significant in only one cell
379 type. Notably, 146 (71%) of these were not detected in bulk-tissue whole blood MR,
380 highlighting the unique resolution provided by single-cell approaches (**Supplementary Table**
381 **11**). Of the 1,791 gene-cell type pairs with MR evidence, between 66 to 1,076 pairs were also
382 causally associated with other immune diseases (in descending order of shared signals:
383 inflammatory bowel disease, ulcerative colitis, psoriasis, type 1 diabetes, rheumatoid arthritis,
384 systemic lupus erythematosus, multiple sclerosis, and eczema), potentially revealing shared
385 causal molecular pathways. Of the 1,791 gene-cell type pairs with MR evidence, 232 (13%)
386 demonstrate discordant directions of effect between diseases, highlighting disease context-
387 dependent causal effects (**Supplementary Table 12**).

388 As a proof-of-principle, we recovered known therapeutic targets for Crohn's disease,
389 including *TNF*, *IL23R*, and canonical risk genes such as *ATG16L*, *NOD2*, and *IRGM*
390 (**Supplementary Table 13**, **Figure 5a**). Interestingly, we observed divergent directions of
391 causal effect for 61 genes (20%), where the same gene exhibited opposing MR directions in
392 different cell types. For example, *NCF4* (which encodes a component of an enzyme complex
393 in phagocytes), showed a positive effect in B memory cells but a negative effect in B naive
394 cells (**Figure 5b**), suggesting context-specific regulatory roles. The negative effect of *NCF4* in
395 B naive cells is also observed in eczema. Interestingly, while the positive effect of *NCF4* in B
396 memory cells is also observed in psoriasis, it has a negative effect in multiple sclerosis
397 (**Supplementary Table 12**).

398 To validate the biological relevance of our findings, we compared MR-identified genes
399 with differentially expressed genes from single-cell RNA-seq of colon biopsies from 46
400 patients with Crohn's disease and 25 healthy controls[33]. While some intestinal immune cells
401 are tissue-resident, many cell type lineages, such as B cells, T cells, dendritic cells, and NK
402 cells, are shared between peripheral and tissue compartments. We harmonised cells between
403 datasets by major cell types, excluding tissue-specific subsets such as macrophages
404 (**Supplementary Figure 15**), and examined the consistency of MR directionality with
405 differential expression.

406 We reasoned that genes with a positive MR effect (*i.e.*, risk-increasing) would be
407 upregulated in disease tissue, whereas those with a negative MR effect (protective) would be
408 downregulated. We identified a total of 116 genes with matched major cell types. Of these, 62
409 genes had concordant direction of effect between MR and DEG, 47 with discordant effects and
410 7 with both concordant and discordant effects, depending on concordance between and within
411 major cell types (**Supplementary Table 14, Supplementary Figure 16**). We highlight
412 *ZBTB38* as an interesting example with convergent evidence: it showed a positive MR effect
413 and was upregulated in B cells, CD4+ T cells, and CD8+ T cells from Crohn's disease colon
414 samples (**Figure 5c**). Conversely, *TNFRSF18* (a known immune modulator) showed a
415 protective MR effect but was also upregulated in CD4+ T cells from Crohn's disease colon,
416 suggesting potential compensatory mechanisms due to reverse causation (**Supplementary**
417 **Table 14, Supplementary Figure 16**).

418 Together, these results demonstrate how single-cell MR can identify both established
419 and uncharacterised causal genes with cell-type-specific effects, refine interpretation of
420 differential expression studies, and inform mechanistic hypotheses for complex disease
421 biology.

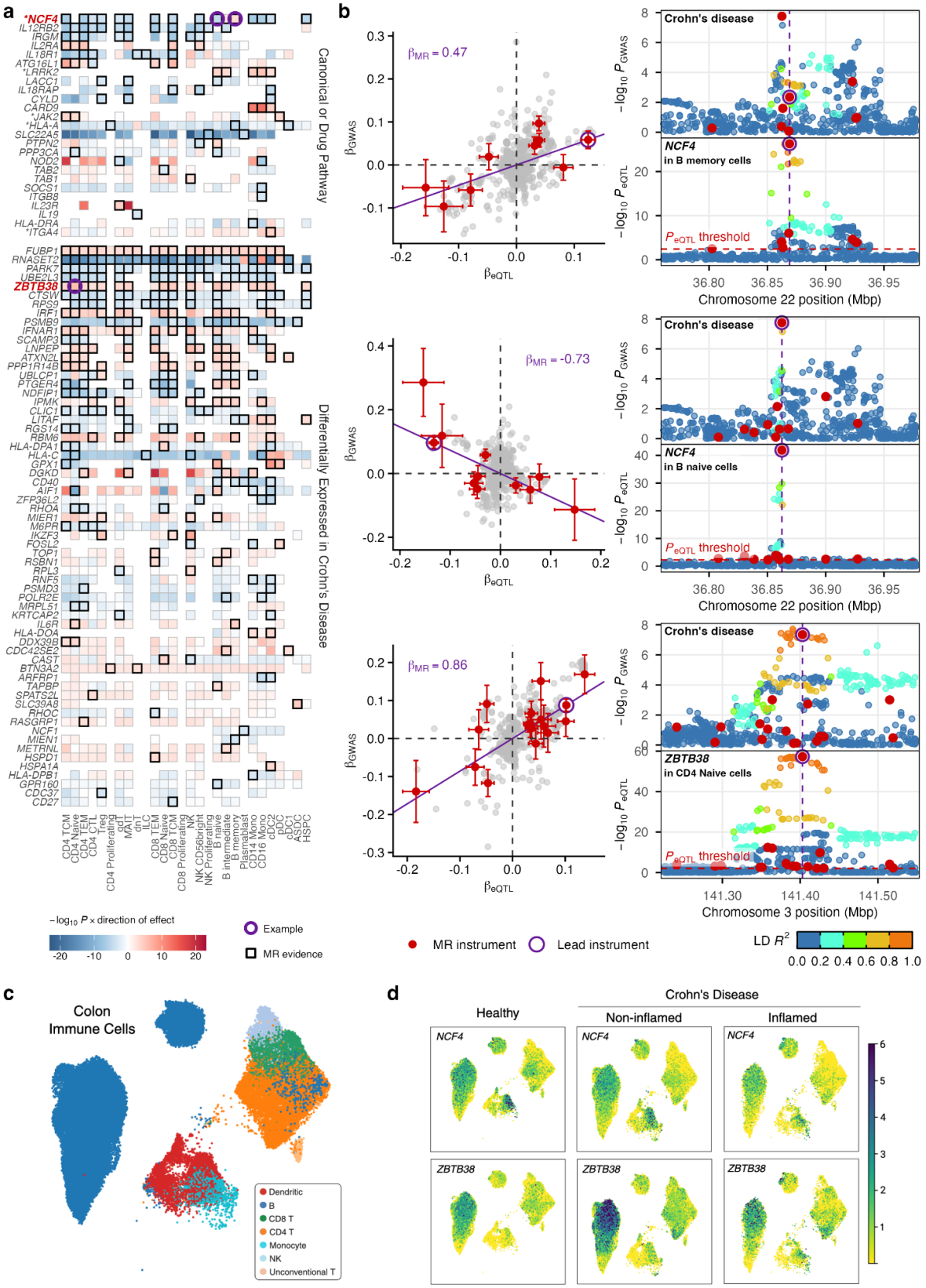


Figure 5 | Cell-type-specific causal and transcriptional signatures in Crohn's disease.
a) Selected MR-significant genes in Crohn's disease, shown as $-\log_{10} P_{MR}$ multiplied by direction of MR estimate of the lead instrument. Only protein coding genes are shown. A positive direction denotes increased disease risk with increased expression, and negative direction denotes reduced disease risk

427 for increased expression. Gene categories include canonical risk genes, approved drug targets and their
428 pathways (combined), and differentially expressed genes in matched major cell types of colon immune
429 cells of Crohn's disease patients (DEGs) with concordant direction of effect (See **Methods** and
430 **Supplementary Table 12**). Genes denoted with an asterisk are both canonical or drug target genes and
431 DEGs. DEGs were identified from colon biopsy scRNA-seq data (Kong et al., 2023)[33] by matching
432 genes and major immune cell types with MR results. Empty grids indicate the gene was not an eGene
433 in that cell type and was therefore not tested. Black outlines denote MR associations ($Q_{MR} < 0.05$; P_{HEIDI}
434 > 0.01). **b)** Causal effect plots showing GWAS effect sizes (*y*-axis) against cell type-specific sc-eQTL
435 SNP effect sizes (*x*-axis), paired with locus zoom plots for selected genes. Data points in both plots
436 represent common genetic variants in the *cis*-region of the gene that are available in both GWAS and
437 sc-eQTL datasets. Variants used as instruments for gene expression in the MR analysis are highlighted
438 in red, and lead instruments (lowest sc-eQTL *P*) are circled in purple. The slope in the effect plot
439 corresponds to the point estimate of the MR effect size (coefficient) of the lead instrument. Row 1 and
440 2, respectively: positive causal effect of *NCF4* in B memory cells and negative causal effect of *NCF4*
441 in B naive cells; Row 3: positive causal effect of *ZBTB38* in CD4+ naive cells. **c)** Colon immune cells
442 from Kong et al., 2023, colour-coded by major cell type annotations derived by mapping major cell
443 types with the TenK10K study (see **Methods, Supplementary Table 16**). Unmatched major cell types
444 are excluded (original annotations of the full datasets are presented in **Supplementary Figure 16**). **d)**
445 Expression of *NCF4* and *ZBTB38*, which both have causal effects on Crohn's disease in circulating
446 immune cells and differential expression in colon immune cells. Colon immune cells are projected on
447 the UMAP coordinates from panel c, stratified by disease and inflammation status.
448

449 ## Discussion

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

462 Approximately one-third (34%) of gene–trait associations identified through MR were
463 not recovered using conventional gene-based GWAS approaches (*e.g.*, MAGMA), and the
464 majority of these were specific to a single or small number of cell types. This underscores the
465 value of cell-resolved sc-eQTL data for uncovering regulatory effects that are masked in bulk
466 tissue analyses. Notably, we found that 61% of the gene-trait associations identified in our
467 study were undetectable in a comparable MR analysis using genetic instruments derived from
468 bulk-tissue eQTL mapping of whole blood sample from eQTLgen[17], despite having 16 times
469 larger sample size for eQTL discovery. These findings affirm the hypothesis that cell-specific
470 genetic regulations are undetectable in aggregated datasets, further reinforcing the utility of
471 single-cell resolved eQTL mapping.

While our analysis focused on PBMCs, we identified causal associations not only for immune-related diseases and blood traits, but also for systemic traits such as type 2 diabetes, schizophrenia, cancer, and bone mineral density. These results potentially suggest that peripheral immune cell dysregulation is directly involved in the pathogenesis of systemic diseases, or that they may share cell-level genetic regulation with other cell types in disease-relevant tissues. These two hypotheses can be observed in the polygenic enrichment patterns observed in the single-cell disease relevance score (scDRS) analysis, which shows that polygenic signatures of systemic disease were indeed observed across PBMC cells, but to a lesser extent compared to traits directly related to immune dysregulation and peripheral blood composition. Given that nearly half of cis-eQTLs are shared across tissues[20], these findings implicate that large-scale sc-eQTL maps may provide some insights into causal biology of systemic diseases in other tissues with similar expression of disease-relevant genes or with shared genetic regulation.

At the cell type level, we found striking enrichment of immune-mediated disease associations in dendritic cell (DC) populations. In Crohn's disease, for example, multiple DC subsets (cDC1, cDC2, ASDC) showed strong polygenic enrichment and MR-supported causal associations, consistent with prior evidence that DCs accumulate and become activated in inflamed intestinal tissue[9]. In contrast, COVID-19 genetic risk was specifically enriched in plasmacytoid DCs (pDCs), aligning with their established role in antiviral immunity and rapid interferon production[11]. Focusing on B cells in the context of SLE, we found SLE-enriched cells display significantly a higher cell function score for cellular response to type II interferon, a B cell-intrinsic mechanism implicated in SLE pathogenesis via autoimmune response due to dysregulated germinal centre[38]. These findings suggest that large-scale sc-eQTL data, when integrated with GWAS and functional annotation of relevant biological context, can resolve the cell type-specific contributions of immune populations to disease even in the absence of disease-context tissue sampling.

Translationally, our data show that causal gene–trait relationships derived from single-cell MR are predictive of clinical development success. By intersecting MR results with drug target and trial data from the Open Targets Platform, we found that target–indication pairs supported by single-cell MR were 3.3 times more likely to reach regulatory approval. These pairs also showed higher evidence scores across both genetic and non-genetic data sources and were more likely to display concordant direction of effect with the mechanism of action of approved or late-phase drugs. Our framework successfully recapitulated known therapeutic targets, including *JAK2* and *TNF* for Crohn's disease, *IL6R* for rheumatoid arthritis, *APP* for Alzheimer's disease, and *GLP1R* for type 2 diabetes, while nominating cell type-specific genes with strong genetic support for new indications. In addition, our data show more than half of targets tested in clinical trials (including 151 that are already approved) had single-cell MR support for at least one tested indication, implicating opportunities for drug repurposing.

To highlight the utility of cell-type-resolved MR for identifying both canonical and novel therapeutic targets, we further explored 510 genes associated with Crohn's disease as a case study. We identified 510 MR-significant genes, including those with well-established

513 roles such as *NCF4*, *ATG16L1*, and *IL23R*, as well as less characterised candidates such as
514 *ZBTB38*. The transcriptional repressor *ZBTB38* exhibited a positive causal effect in T and B
515 cell subsets and was independently upregulated in disease-affected colon tissue. *ZBTB38* has
516 been linked to Crohn's disease[39] and other inflammatory diseases[40,41], however its functional
517 role in T and B cells has not been explored. Our findings here offer a testable hypothesis that
518 *ZBTB38* may modulate cell type-specific programs relevant to disease pathogenesis, supported
519 by differential expression in disease. A second example, *NCF4,* is a known risk gene that
520 encodes a subcomponent of the NADPH complex, playing a critical role in oxidative stress and
521 antimicrobial function[42]. In Crohn's disease, *NCF4* exhibited a positive effect in B memory
522 cells, in contrast to a negative effect in other cell types including B naive cells. Interestingly,
523 independent eQTL data show that the top eSNP in B memory cells, which has a uniquely
524 positive effect, is a fine-mapped SNP with a negative eQTL effect in monocytes, macrophages,
525 neutrophils, but a positive splicing QTL (sQTL) effect in B-cell derived lymphoblastoid cell
526 lines[24,43]. Many eQTL effects are explained by splicing QTLs[44,45], and these results highlight
527 the need for long-read sequencing to resolve cell type specific sQTLs and explain novel causal
528 associations. In this case study, we form links between MR results within diseases as well as
529 independent, orthogonal sources of evidence such as differential gene expression[46].

530 We also identified MR-supported associations that may inform mechanisms of action
531 for existing therapies. For example, for Crohn's disease, we found a positive causal association
532 of *IL23R* in γδ T cells and a negative (protective) association of *IL12Rβ2* in multiple immune
533 cell types, while *IL12Rβ1* showed no significant effect. These observations are consistent with
534 clinical evidence that ustekinumab (targeting the shared *IL12*/*IL23* p40 subunit, blocking the
535 *IL23R*/*IL12Rβ1* receptor complex) acts predominantly through *IL23* blockade, and may help
536 explain the increased efficacy of *IL23*-specific therapies like guselkumab[47–49]. Our data suggest
537 that targeted modulation of *IL23R* and sparing of *IL12RB2* may optimise therapeutic efficacy
538 while reducing off-target effects.

539 Our study has several limitations. First, MR estimates are subject to bias from
540 overestimation in sc-eQTL discovery (winner's curse) and residual pleiotropy not fully
541 accounted for by the HEIDI test. Second, PBMCs may not capture causal regulation in disease-
542 relevant tissues or tissue-resident immune compartments. Third, the statistical power of our
543 study may be limited for rare cell types genes with low expression. While we downsampled
544 abundant cell types to reduce bias in the polygenic enrichment analysis, sensitivity analyses
545 suggest this had minimal impact. Furthermore, OpenTargets may not capture all active drug
546 development programs, and many apparently negative results may reflect the absence of
547 evidence rather than evidence of absence. Depending on the particular disease, signals may not
548 be detected if their key causal effects are mediated in non-peripheral immune cells, or tissue-
549 specific cells such as epithelial or stromal cells. Finally, functional validation of hypotheses is
550 essential to confirm the causal relevance of newly implicated genes and their therapeutic
551 potential.

552 In summary, we present a large-scale, cell type-resolved atlas of causal gene–trait
553 associations by integrating single-cell eQTL mapping with GWAS through Mendelian

554 randomisation. Our results highlight extensive and cell-specific polygenic contributions to
555 complex traits, identify new regulatory mechanisms in disease, and demonstrate that MR-
556 informed cell type-specific targets are predictive of drug development success. This resource
557 provides a powerful foundation for understanding the molecular underpinnings of complex
558 disease and for guiding precision therapeutic strategies.

559

560 Methods

561 Data sources

562 Single-cell eQTL

563 To perform single-cell MR analysis, we used genetic instruments derived from a large-
564 scale single-cell expression quantitative trait loci (sc-eQTL) association analysis from Phase 1
565 of the TenK10K project[15]. This dataset comprises matched whole-genome sequencing (WGS)
566 and single-cell RNA-sequencing (scRNA-seq) data from 1,925 individuals of European
567 ancestry from Tasmanian Ophthalmic Biobank (TOB) and the BioHEART study cohort,
568 covering over 5 million peripheral blood mononuclear cells (PBMCs) across 28 immune cell
569 types. The cell types were identified using a combined approach that uses scPred[50] and sc-
570 HPL[51] with Azimuth PBMC reference dataset[52].
571 The sc-eQTL association test was performed separately in each of the 28 cell types
572 using a Poisson mixed model for single-cell counts implemented in SAIGE-QTL[53] with the
573 following covariates: sex and age of the individuals, 7 genotype PCs, 5 cell type-specific
574 expression PCs, cohort status (TOB or BioHEART), percentage of counts from mitochondrial
575 genes and total counts. For each gene, common genetic variants (minor allele frequency >1%)
576 within *cis*-region of the gene, defined as 100 kbp window upstream and downstream site of the
577 tested eGene bodies based on ENCODE v44 hg38 coordinates, were tested for association with
578 expression level of the gene measured in raw counts. A total of 12,874,224 variants were
579 included in the analysis, which yielded 154,932 common variant sc-eQTL across 17,674 unique
580 eGenes across cell types at FDR 5%.

581 Genome-wide association study of complex traits

582 To perform MR, MAGMA and scDRS, analyses, we used publicly accessible genome-
583 wide association results of 100 complex diseases and biological traits across 10 phenotype
584 categories: Blood (26 traits)[54], Skeletal traits (5 traits)[55,56], Cancer (31 diseases)[57–66], Immune
585 (9 diseases)[67–73], Viral infection (6 diseases)[74–76], Cardiometabolic (6 diseases)[77–83], Neurologic
586 (3 diseases)[84–86], Psychiatric (7 diseases)[87–91], Respiratory (3 diseases)[92–94], and Other (4
587 diseases)[95–98]. For each trait, we obtained GWAS summary statistics from public repositories.
588 Summary statistics were harmonised to the TenK10K study by matching variants'
589 genomic coordinates and alleles, performing conversion to hg38 with liftOver and
590 harmonisation of minor and major alleles based on allele frequency in individual-level
591 genotype from the TenK10K study. To ensure consistency of variants' allele frequency and

592 linkage disequilibrium pattern with TenK10K sc-eQTL dataset, analysis was limited to GWAS
593 performed in European ancestries. If unreported in the GWAS summary statistics, values such
594 as the beta value as log(OR), standard error and p-values were derived from other reported
595 values. The effective GWAS sample size for binary disease traits presented in **Figure 1c** was
596 calculated as 4 / (1/number of cases + 1/number of controls). More details of each trait, GWAS
597 data source, reference, and number of samples are provided in **Supplementary Table 2**.
598

599 Mendelian randomisation analysis

600 To identify genes associated with disease and biomarker traits in PBMC cell types, we
601 performed a multi-instrument Mendelian randomisation (MR) analysis using the SMR
602 approach as implemented in the SMR v1.3.1 software package[13]. Genetic instruments for MR
603 were selected from common variants within the *cis*-region of the tested gene bodies passing
604 cell type-specific thresholds. The cell type-specific thresholds were derived from the maximum
605 *P*-value across sc-eQTL for eGenes identified at FDR <5% in the TenK10K phase I study[15].
606 Following the default parameter of the SMR software[14], variants that are not available in the
607 GWAS summary statistics of tested phenotype or those with allele frequency difference > 0.2
608 are excluded prior to the instrument selection steps. In the MR analysis, we pruned variants
609 that are highly correlated (linkage disequilibrium (LD) $r^2 > 0.9$) with the top sc-eQTL within
610 the region, and used set-based approximation accounting for residual LD to combine MR
611 estimates from the remaining variants, with *P*-value derived using Saddlepoint approximation.
612 MR analysis was performed by modelling the effect of increased gene expression level on
613 disease risk or trait levels so that a positive association implies a trait-increasing (or risk-
614 increasing) effect and a negative association implies a trait-decreasing (or protective) effect.

615 To adjust for multiple testing, MR *P*-values were adjusted for false discovery rate
616 (FDR) by calculating *Q*-values based on *P*-values distribution across cell type-gene
617 combinations per trait with the *qvalue* R package[99]. To minimise confounding by pleiotropic
618 effects on other genes, we further performed a test for heterogeneity in dependent instruments
619 (HEIDI)[13] for each MR test we performed. An MR estimate with a $Q < 0.05$ and a $P_{HEIDI} \geq$
620 0.01 were considered as evidence of a causal association.

621 The comparison with eQTLGen was performed using publicly available pre-formatted
622 SMR summary statistics from eQTLgen Consortium[17], downloaded from
623 https://eqtlgen.org/cis-eqtls.html. Consistent with the original publication, we used a threshold
624 of $P_{eQTL} < 2.02 \times 10^{-5}$ to select genetic instrument for MR. Multiple testing correction for
625 MRwas performed consistently with the TenK10K study.

626

627 Identification of GWAS genes

628 To identify GWAS genes for comparison with MR results, we performed a gene-level
629 aggregation of single-variant estimates from GWAS summary statistics using MAGMA[16]. The

aggregation for each gene was performed across variants within the transcription start and end sites of the gene (based on ENCODE v44 hg38 coordinates) using a SNP-wise model with linkage disequilibrium between variants modeled using reference data TenK10K phase I individual-level genotype. *P*-value for a gene-trait association was derived using mean chi-squared statistics by approximation of known sampling distribution. To account for multiple testing, we calculated Q-values across all analytical sets. Genes associated with a trait at $Q < 0.05$ were considered as putative GWAS genes for the trait.

Single-cell polygenic enrichment of complex traits

We utilized the single-cell disease relevance score (scDRS)[10] method to quantify polygenic enrichment of complex traits at both individual-cell and cell-type levels. scDRS integrates single-cell gene expression profiles with GWAS summary statistics to associate individual cells to the GWAS trait, by testing overexpression of trait-relevant genes derived from GWAS in a given cell relative to other genes with similar expression profile across all cells. To run scDRS for each trait, we first extracted the top 1,000 trait-associated genes based on *Z*-score derived from the MAGMA analysis. Next, we used scDRS to compute the aggregate expression of the associated genes in each cell, weighted by the GWAS MAGMA *Z*-score and inversely weighted by gene-specific technical covariates in the single-cell data, using the mean-variance model. To compute cell-level *P*-value for overexpression of trait-relevant genes, we followed scDRS method to first generate 1,000 sets of cell-specific raw control scores computed using Monte Carlo (MC) samples of matched control gene sets (matching gene set size, mean expression and expression variance of the selected genes). The *P*-value was then derived by comparing normalised trait scores against normalised control scores based on the empirical distribution of the pooled normalised control scores across all control gene sets and all cells.

Single-cell RNA sequencing data to perform scDRS analysis were extracted from ~5 million PBMCs that passed through the processing and quality control pipeline in the TenK10K phase I study[15]. The cell-level scoring in scDRS was performed using sex, age, number of genes, and cohort (TOB and BioHEART) as covariates. To manage computational efficiency and account for variations in cell type proportions, we randomly sampled a maximum of 10,000 cells for each of the 28 immune cell types in the TenK10K phase 1 study, resulting in 229,239 total sampled cells. To test the robustness of this sampling strategy, we repeated this procedure using Crohn's disease as an example with an alternative sampling strategy by taking 10% of cells from each cell type.

Quantification of cell function score in B cells

We computed cell function scores in B cells as cell states using scDeepID, a flexible multi-task transformer that leverages different biological pathway databases as guides to train the model. First, we trained a scDeepID model with 539,820 B cells (including B naive, B intermediate, and B memory cells) from TenK10K phase 1 single-cell RNA sequencing data.

The training used expression features of 2,351 highly variable genes selected with Scanpy function highly_variable_genes (min_mean=0.0123, max_mean=3, min_disp=0.5) with human Biological Process Gene Ontology (GOBP) database as a guide, filtered to the top 400 most informative GOBP pathways based on the presence of pathway genes in the highly variable gene list. We then extracted the attention-based latent space for UMAP representation, and extracted the cell function scores followed by transformation of each cell function score to approximate a normal distribution with a zero mean and one unit variance.

As a proof-of-concept to explore the function of cells that are enriched for specific phenotypes, we intersected scDeepID cell function score with scDRS score for systemic lupus erythematosus ($scDRS_{SLE}$) amongst 30,000 subsampled B cells used in the scDRS analysis. We first grouped these cells into SLE-enriched cells (cell-level $P_{scDRS\text{-}SLE}$) and non-SLE-enriched cells, and compared mean cell function scores between the two groups using a one-sided *T*-test for each pathway under analysis followed by FDR correction.

Estimating the support of single-cell genetics for drug target validation

To evaluate the utility of single-cell genetics for drug target identification, we cross-referenced our cell-type resolved MR results with a target-evidence association score and drug development activity dataset sourced from Open Targets Platform (OTP) [24] version 25.06. Analysis were limited to a set of 12,873 protein-coding genes tested in the MR analysis and 57 diseases with at least one matching indication available in the Open Targets Platform database, resulting in a total of 733,761 possible pairs of target (gene) and indication (disease) – the target-indication universe in this study. We used Ensembl gene ID to match tested genes to OTP targets, and used Experimental Factor Ontology (EFO) ID to match tested phenotypes to OTP indications.

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

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

For every category, we computed the 2x2 intersection between target-indication inside or outside the category and target-indication with or without MR support, resulting in four possible combinations: in the category & with MR support, in the category & without MR support, outside the category & with MR support, and outside the category & without MR support. Using this 2x2 contingency table, we then computed the odds ratio for being in the category, comparing between target-indication pairs with and without MR support. *P*-value and 95% lower confidence interval were calculated using a one-sided Fisher's exact test implemented in *fisher.test()* R function. For this analysis, MR support was defined as a significant MR association ($Q_{MR} < 0.05$ and $P_{HEIDI} \geq 0.01$) in at least one cell type under analysis.

Second, for every matched target-indication pair under analysis, we extracted the overall OTP evidence score which quantifies aggregated evidence from data sources of genetic association, somatic mutation, known drug, affected pathway, literature, RNA expression, and animal model. We took the maximum of direct (evidence referring directly to indication ID) and indirect association score (evidence referring to indication ID or any of its descendant IDs in EFO). The association score takes a value between 0 and 1, where a higher score indicates more evidence supporting the target-indication association relative to a theoretical maximum. We computed the mean score of target-indication with and without MR support, and tested the difference using a two-sample Welch's *T*-test implemented in *t.test()* R function. We then extracted the evidence score stratified by data score, computed the genetic evidence score (unweighted mean of scores from genetic association sources) and non-genetic evidence (unweighted mean of scores from non-genetic association sources), and repeated the test procedures for the two evidence sources.

Third, we assessed the concordance between MR-predicted effect direction and drug mechanism across clinical development phases. The MR-predicted effect direction for a target-indication pair was derived by taking the majority of effect directions across cell types with significant MR associations. To remove ambiguity, target-indication with equal number of positive and negative effects across cell types were excluded from the analysis. The drug mechanism effects were extracted from ChEMBL drug data in OTP, and matched with ChEMBL ID of drug with maximum clinical development phase for a given target-indication pair.

Crohn's disease case study

We extracted all MR associations for Crohn's disease which passed the discovery threshold of FDR-adjusted $Q < 5\%$ and $P_{HEIDI} > 0.01$.. For each unique MR gene identified, we constructed a Boolean annotation (TRUE / FALSE) based on membership in: 1) the canonical gene list, 2) the drug target gene list 3) the drug target pathways, 4) cell type-specific MR association (whether the MR association is only identified in one cell type), 5) discordance of MR effect direction in at least two cell types, and 6) differentially expressed in Crohn's disease colon tissue where the major cell types are matched (**Supplementary Table 13**).

740 Canonical genes from Crohn's disease were downloaded from a systematic review of
741 genes associated with Crohn's disease[100] The full list (n = 1,172) was filtered to belong to
742 Categories "Experimental evidence of variant", "Other evidences of genetic alterations",
743 "Treatment Response" , "Biologically related but no evidence of mutation", with a Document
744 Score >= 8 and Abstract Number >= 10, resulting in a list of 49 genes (**Supplementary Table**
745 **15**). The drug target genes were extracted from review papers[101,102], and the drug pathway
746 genes (IL23/12, ITGA4, JAK/STAT, TNF) were extracted from GSEA platform[103,104]
747 following the approach described previously[71] (**Supplementary Table 16**). The list of
748 differentially expressed genes was derived from an external dataset of single-cell RNA
749 sequencing in intestinal tissues from 46 Crohn's disease patients and 25 healthy controls,
750 comparing both inflamed and non-inflamed disease tissue with healthy tissue.[33] To derive this
751 annotation, we extracted differentially expressed genes in the colon (pre-computed Discrete
752 FDR < 0.05) across immune cell types. We harmonised the original cell type annotation with
753 the TenK10K cell type annotation used in this study by matching major cell types
754 (**Supplementary Table 17**). We then identified MR and DEG genes in matched major cell
755 types, and checked for concordance of direction of effect by comparing the beta sign between
756 MR and the Discrete model of DEG (**Supplementary Table 14**). This generated multiple
757 combinations of gene-major cell type pairs, based on unique combinations of original MR cell
758 types and DEG cell types. 'All concordant' corresponds to cases where every gene-MR cell
759 type-DEG cell type set has a matched direction of effect between MR and DEG. 'All
760 discordant' corresponds to cases where every gene-MR cell type-DEG cell type set has an
761 unmatched direction of effect between MR and DEG. 'Mixed concordance' corresponds to
762 genes which have discordant directions between or within major cell types.
763 To evaluate expression of the genes of interest in colon tissue (**Figure 5d**), we extracted
764 sc-RNAseq data from colon tissue of Crohn's disease patients[33], downloaded from
765 CELLxGENE[105]. We present the gene expression as normalised counts on the original Uniform
766 Manifold Approximation and Projection (UMAP) embedding using Scanpy, stratified by
767 'Type' into "Healthy", "Non-inflamed Crohn's disease", and "Inflamed Crohn's disease" and
768 filtered to immune cell types with matched major cell types annotation in the TenK10K. The
769 original cell annotation and distribution of cells before and after filtering is presented in
770 **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 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.

Acknowledgements

The authors acknowledge the use of computational resources and support from the Australian National Computing Infrastructure and the Data Science Platform of the Garvan Institute of Medical Research. The authors acknowledge contributions from Prof. Peter Croucher's Lab and Dr. John Kemp's Lab for access to GWAS summary statistics of estimated heel bone mineral density.

Ethics Statement

The Human Research Ethics Committee of St Vincent's Hospital gave ethical approval for this work. The National Statement on Ethical Conduct in Human Research of the National Health and Medical Research Council gave ethical approval for this work.

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

References

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

867 20. Swiecki, M. & Colonna, M. The multifaceted biology of plasmacytoid dendritic cells. *Nat. Rev.*
868 *Immunol.* **15**, 471–485 (2015).

869 21. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic
870 evidence on clinical success. *Nature* **629**, 624–629 (2024).

871 22. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications.
872 *Nature Genetics* **47**, 856–860 (2015).

873 23. Dann, E. *et al.* Estimating the impact of single-cell RNA sequencing of human tissues on drug
874 target validation. *medRxiv* 2024.04.04.24305313 (2024) doi:10.1101/2024.04.04.24305313.

875 24. Buniello, A. *et al.* Open Targets Platform: facilitating therapeutic hypotheses building in drug
876 discovery. *Nucleic Acids Res* **53**, D1467–D1475 (2025).

877 25. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple
878 bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024).

879 26. Hampel, H. *et al.* The amyloid-β pathway in Alzheimer's disease. *Mol. Psychiatry* **26**, 5481–
880 5503 (2021).

881 27. Zheng, Z. *et al.* Glucagon-like peptide-1 receptor: mechanisms and advances in therapy. *Signal*
882 *Transduct. Target. Ther.* **9**, 234 (2024).

883 28. Lincoff, A. M. *et al.* Semaglutide and cardiovascular outcomes in obesity without diabetes. *N.*
884 *Engl. J. Med.* **389**, 2221–2232 (2023).

885 29. Ryan, D. H. *et al.* Long-term weight loss effects of semaglutide in obesity without diabetes in the
886 SELECT trial. *Nat. Med.* **30**, 2049–2057 (2024).

887 30. Roda, G. *et al.* Crohn's disease. *Nat. Rev. Dis. Primers* **6**, 22 (2020).

888 31. Torres, J., Mehandru, S., Colombel, J.-F. & Peyrin-Biroulet, L. Crohn's disease. *Lancet* **389**,
889 1741–1755 (2017).

890 32. de Lange, K. M. & Barrett, J. C. Understanding inflammatory bowel disease via
891 immunogenetics. *J. Autoimmun.* **64**, 91–100 (2015).

892 33. Kong, L. *et al.* The landscape of immune dysregulation in Crohn's disease revealed through
893 single-cell transcriptomic profiling in the ileum and colon. *Immunity* **56**, 444–458.e5 (2023).

894 34. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic
895 architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.*
896 **19**, 110–124 (2018).

897 35. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common disease is more
898 complex than implied by the core gene omnigenic model. *Cell* **173**, 1573–1580 (2018).

899 36. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of
900 quantitative traits. *Nat. Rev. Genet.* **25**, 639–657 (2024).

901 37. Stankey, C. T. *et al.* A disease-associated gene desert directs macrophage inflammation through
902 ETS2. *Nature* **630**, 447–456 (2024).

903 38. Jackson, S. W. *et al.* B cell IFN-γ receptor signaling promotes autoimmune germinal centers via
904 cell-intrinsic induction of BCL-6. *J. Exp. Med.* **213**, 733–750 (2016).

905 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).

44. Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F. & Guigó, R. Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat. Commun.* **12**, 727 (2021).

45. Yamaguchi, K. *et al.* Splicing QTL analysis focusing on coding sequences reveals mechanisms for disease susceptibility loci. *Nat. Commun.* **13**, 4659 (2022).

46. Munafò, M. R., Higgins, J. P. T. & Smith, G. D. Triangulating evidence through the inclusion of genetically informed designs. *Cold Spring Harb. Perspect. Med.* **11**, a040659 (2021).

47. Luo, J. *et al.* Structural basis for the dual recognition of IL-12 and IL-23 by ustekinumab. *J. Mol. Biol.* **402**, 797–812 (2010).

48. Ma, C., Panaccione, R., Khanna, R., Feagan, B. G. & Jairath, V. IL12/23 or selective IL23 inhibition for the management of moderate-to-severe Crohn's disease? *Best Pract. Res. Clin. Gastroenterol.* **38-39**, 101604 (2019).

49. Verstockt, B. *et al.* IL-12 and IL-23 pathway inhibition in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **20**, 433–446 (2023).

50. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).

51. Michielsen, L., Reinders, M. J. T. & Mahfouz, A. Hierarchical progressive learning of cell identities in single-cell data. *Nat. Commun.* **12**, 2799 (2021).

52. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

53. Zhou, W. *et al.* Efficient and accurate mixed model association tool for single-cell eQTL analysis. *Genetic and Genomic Medicine* (2024).

54. Vuckovic, D. *et al.* The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020).

55. Chai, R. C. *et al.* Multiscale analysis and functional validation of the cellular and genetic determinants of skeletal disease. *Genetics* (2024).

56. Miao, J. *et al.* Valid inference for machine learning-assisted genome-wide association studies. *Nat. Genet.* **56**, 2361–2369 (2024).

57. Rashkin, S. R. *et al.* Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat. Commun.* **11**, 4423 (2020).

946 58. Fernandez-Rozadilla, C. *et al.* Deciphering colorectal cancer genetics through multi-omic
947 analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat.*
948 *Genet.* **55**, 89–99 (2023).

949 59. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci
950 and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**, 1126–
951 1132 (2017).

952 60. Adolphe, C. *et al.* Genetic and functional interaction network analysis reveals global enrichment
953 of regulatory T cell genes influencing basal cell carcinoma susceptibility. *Genome Med.* **13**, 19
954 (2021).

955 61. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population.
956 *Nature* **613**, 508–518 (2023).

957 62. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million
958 Veteran Program. *Science* **385**, eadj1182 (2024).

959 63. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility
960 loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).

961 64. Wang, A. *et al.* Characterizing prostate cancer risk through multi-ancestry genome-wide
962 discovery of 187 novel risk variants. *Nat. Genet.* **55**, 2065–2074 (2023).

963 65. Kar, S. P. *et al.* Genome-wide analyses of 200,453 individuals yield new insights into the causes
964 and consequences of clonal hematopoiesis. *Nat. Genet.* **54**, 1155–1166 (2022).

965 66. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat.*
966 *Genet.* **53**, 1415–1424 (2021).

967 67. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics.
968 *Nature* **594**, 398–402 (2021).

969 68. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and
970 adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**,
971 1457–1464 (2015).

972 69. Ishigaki, K. *et al.* Multi-ancestry genome-wide association analyses identify novel genetic
973 mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).

974 70. Stuart, P. E. *et al.* Transethnic analysis of psoriasis susceptibility in South Asians and Europeans
975 enhances fine-mapping in the MHC and genomewide. *HGG Adv.* **3**, 100069 (2022).

976 71. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple
977 integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).

978 72. International Multiple Sclerosis Genetics Consortium (IMSGC) *et al.* Analysis of immune-
979 related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–
980 1360 (2013).

981 73. Budu-Aggrey, A. *et al.* European and multi-ancestry genome-wide association meta-analysis of
982 atopic dermatitis highlights importance of systemic immune regulation. *Nat. Commun.* **14**, 6172
983 (2023).

984 74. COVID-19 Host Genetics Initiative. A second update on mapping the human genetic architecture
985 of COVID-19. *Nature* **621**, E7–E26 (2023).

986 75. Lammi, V. *et al.* Genome-wide association study of long COVID. *Nat. Genet.* **57**, 1402–1417

(2025).

76. Bian, S. *et al.* Genetic determinants of IgG antibody response to COVID-19 vaccination. *Am. J. Hum. Genet.* **111**, 181–199 (2024).

77. Suzuki, K. *et al.* Multi-ancestry genome-wide study in >2.5 million individuals reveals heterogeneity in mechanistic pathways of type 2 diabetes and complications. *medRxiv* (2023) doi:10.1101/2023.03.31.23287839.

78. Zhao, C., Ma, T., Cheng, X., Zhang, G. & Bai, Y. Genome-wide association study of cardiometabolic multimorbidity in the UK Biobank. *Clin. Genet.* **106**, 72–81 (2024).

79. Adlam, D. *et al.* Genome-wide association meta-analysis of spontaneous coronary artery dissection identifies risk variants and genes related to artery integrity and tissue-mediated coagulation. *Nat. Genet.* **55**, 964–972 (2023).

80. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).

81. Zheng, S. L. *et al.* Genome-wide association analysis provides insights into the molecular etiology of dilated cardiomyopathy. *Nat. Genet.* **56**, 2646–2658 (2024).

82. Henry, A. *et al.* Genome-wide association study meta-analysis provides insights into the etiology of heart failure and its subtypes. *Nat. Genet.* **57**, 815–828 (2025).

83. Surakka, I. *et al.* Multi-ancestry meta-analysis identifies 2 novel loci associated with ischemic stroke and reveals heterogeneity of effects between sexes and ancestries. *bioRxiv* (2022) doi:10.1101/2022.02.28.22271647.

84. Kim, J. J. *et al.* Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. *Nat. Genet.* **56**, 27–36 (2024).

85. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).

86. Hautakangas, H. *et al.* Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nat. Genet.* **54**, 152–160 (2022).

87. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208 (2023).

88. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

89. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

90. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).

91. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

92. Tsuo, K. *et al.* Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. *Cell Genom.* **2**, 100212 (2022).

93. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom.* **2**, 100192 (2022).

1027 94. Allen, R. J. *et al.* Genome-wide association study of susceptibility to idiopathic pulmonary
1028 fibrosis. *Am. J. Respir. Crit. Care Med.* **201**, 564–574 (2020).

1029 95. McDonald, M.-L. N. *et al.* Novel genetic loci associated with osteoarthritis in multi-ancestry
1030 analyses in the Million Veteran Program and UK Biobank. *Nat. Genet.* **54**, 1816–1826 (2022).

1031 96. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests
1032 shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**, e1007813 (2018).

1033 97. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain and
1034 inflammatory conditions. *Nat. Genet.* **55**, 423–436 (2023).

1035 98. Lo Faro, V. *et al.* Novel ancestry-specific primary open-angle glaucoma loci and shared biology
1036 with vascular mechanisms and cell proliferation. *Cell Rep. Med.* **5**, 101430 (2024).

1037 99. John D. Storey with contributions from Andrew J. Bass, A. A. D. R. *Qvalue*. (Bioconductor,
1038 2017). doi:10.18129/B9.BIOC.QVALUE.

1039 100. Garza-Hernandez, D. *et al.* A systematic review and functional bioinformatics analysis of genes
1040 associated with Crohn's disease identify more than 120 related genes. *BMC Genomics* **23**, 302
1041 (2022).

1042 101. Feng, Z. *et al.* Breaking through the therapeutic ceiling of inflammatory bowel disease: Dual-
1043 targeted therapies. *Biomed. Pharmacother.* **158**, 114174 (2023).

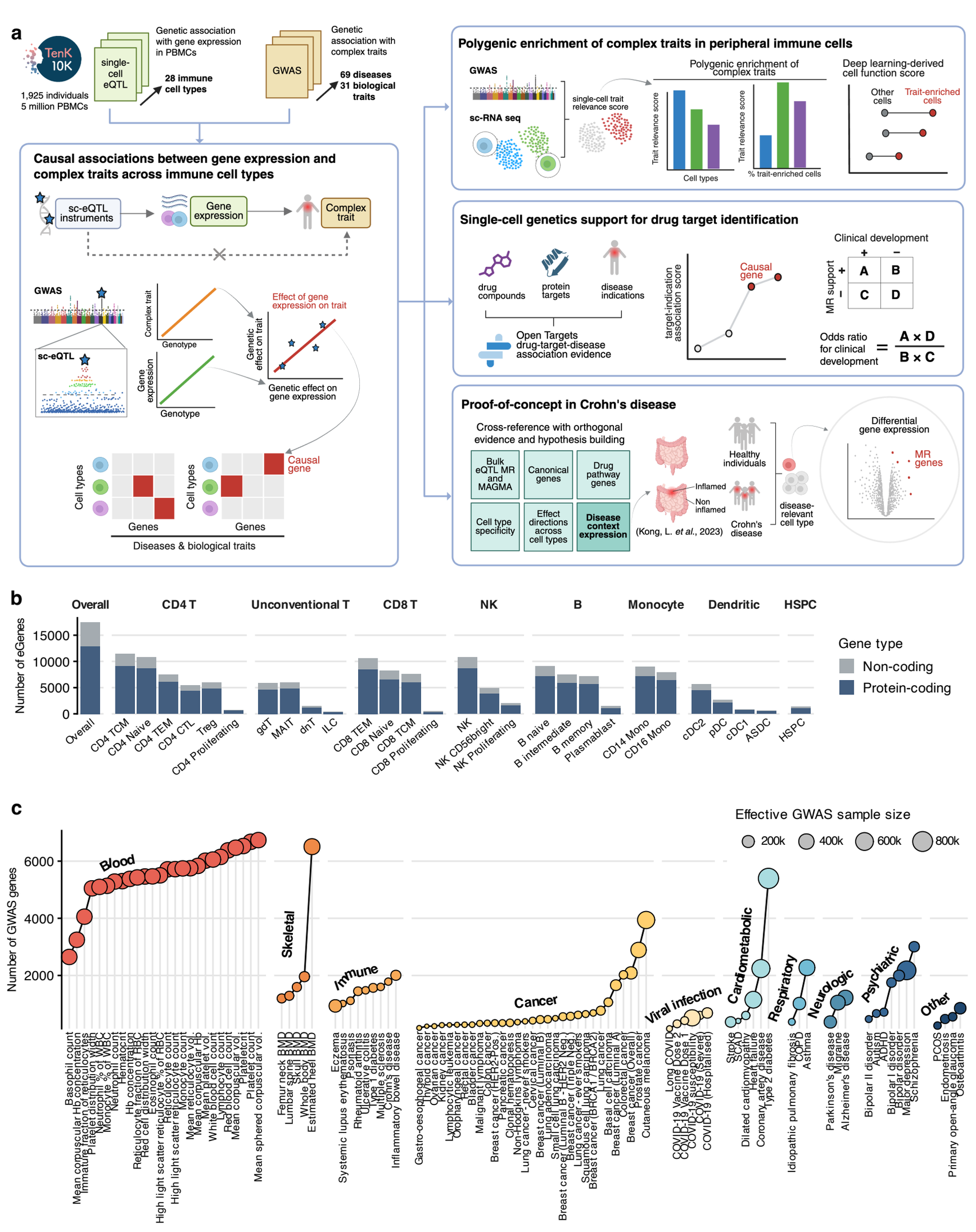
1044 102. Tyebally, R., Xue, A. & Powell, J. E. The potential clinical impact of cell type-specific genetic
1045 regulation: Crohn's disease. *Clin. Transl. Med.* **13**, e1474 (2023).

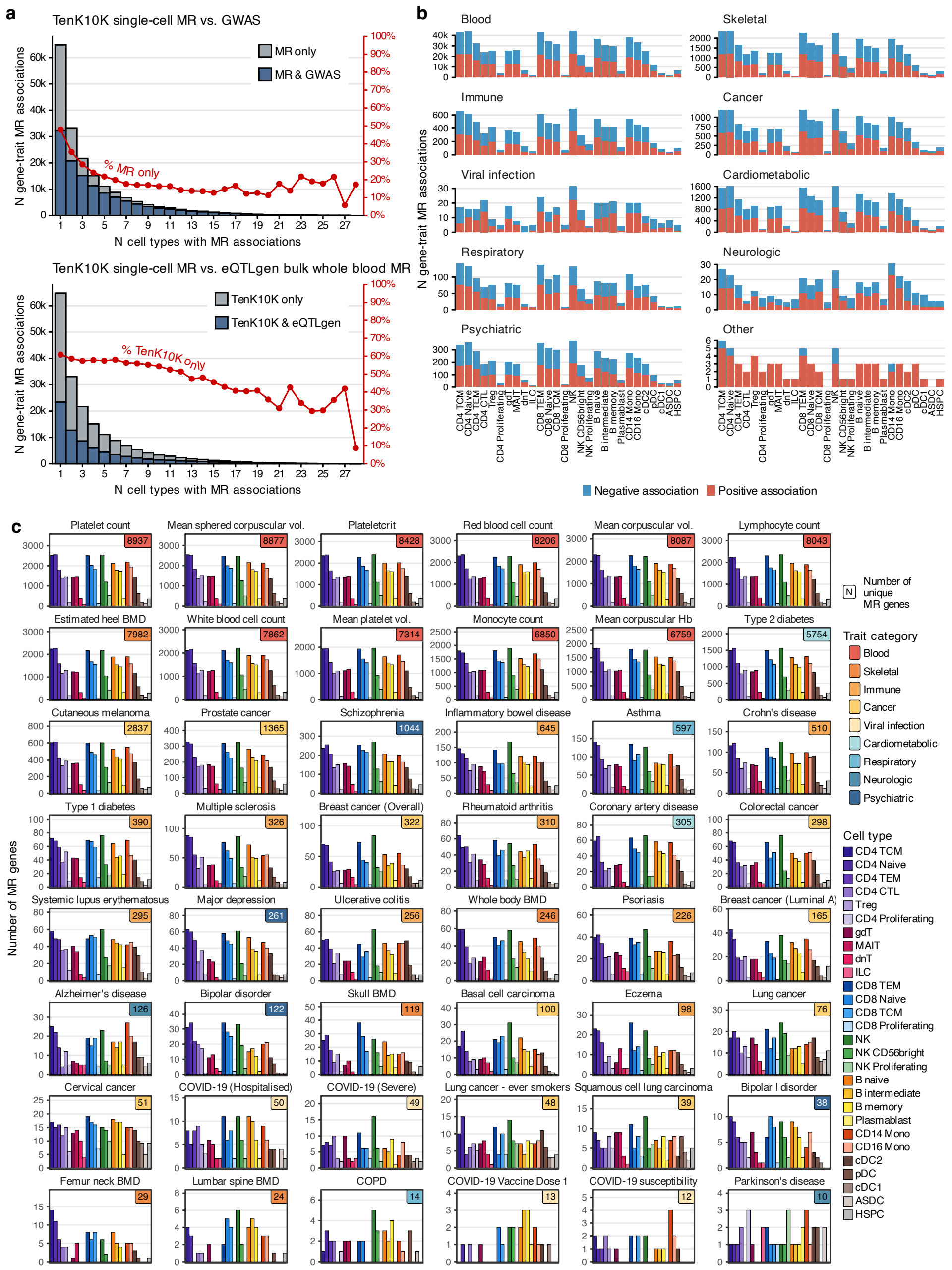
1046 103. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1047 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550
1048 (2005).

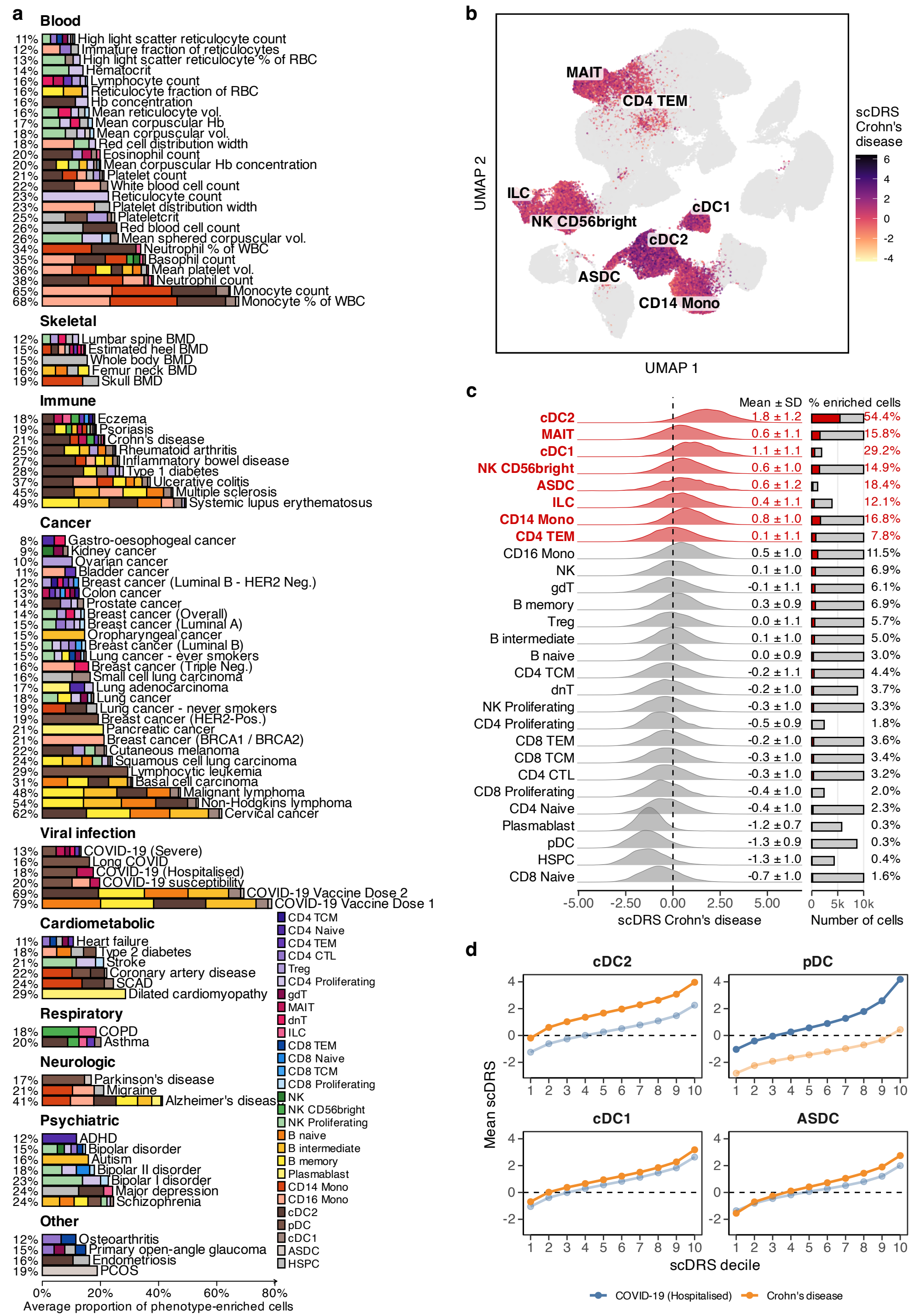
1049 104. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are
1050 coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

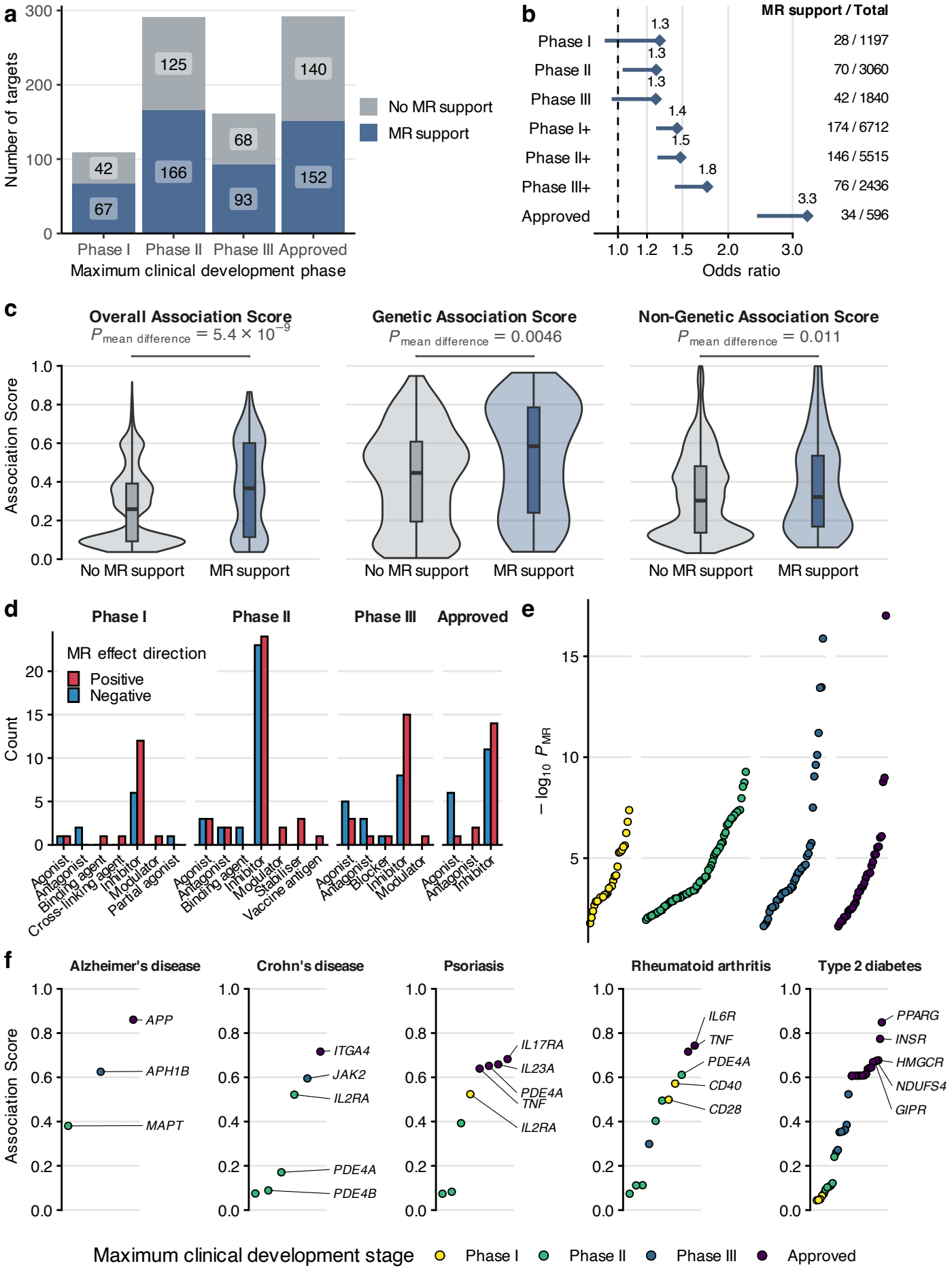
1051 105. CZI Single-Cell Biology Program *et al.* CZ CELL×GENE Discover: A single-cell data platform
1052 for scalable exploration, analysis and modeling of aggregated data. *bioRxiv* (2023)
1053 doi:10.1101/2023.10.30.563174.

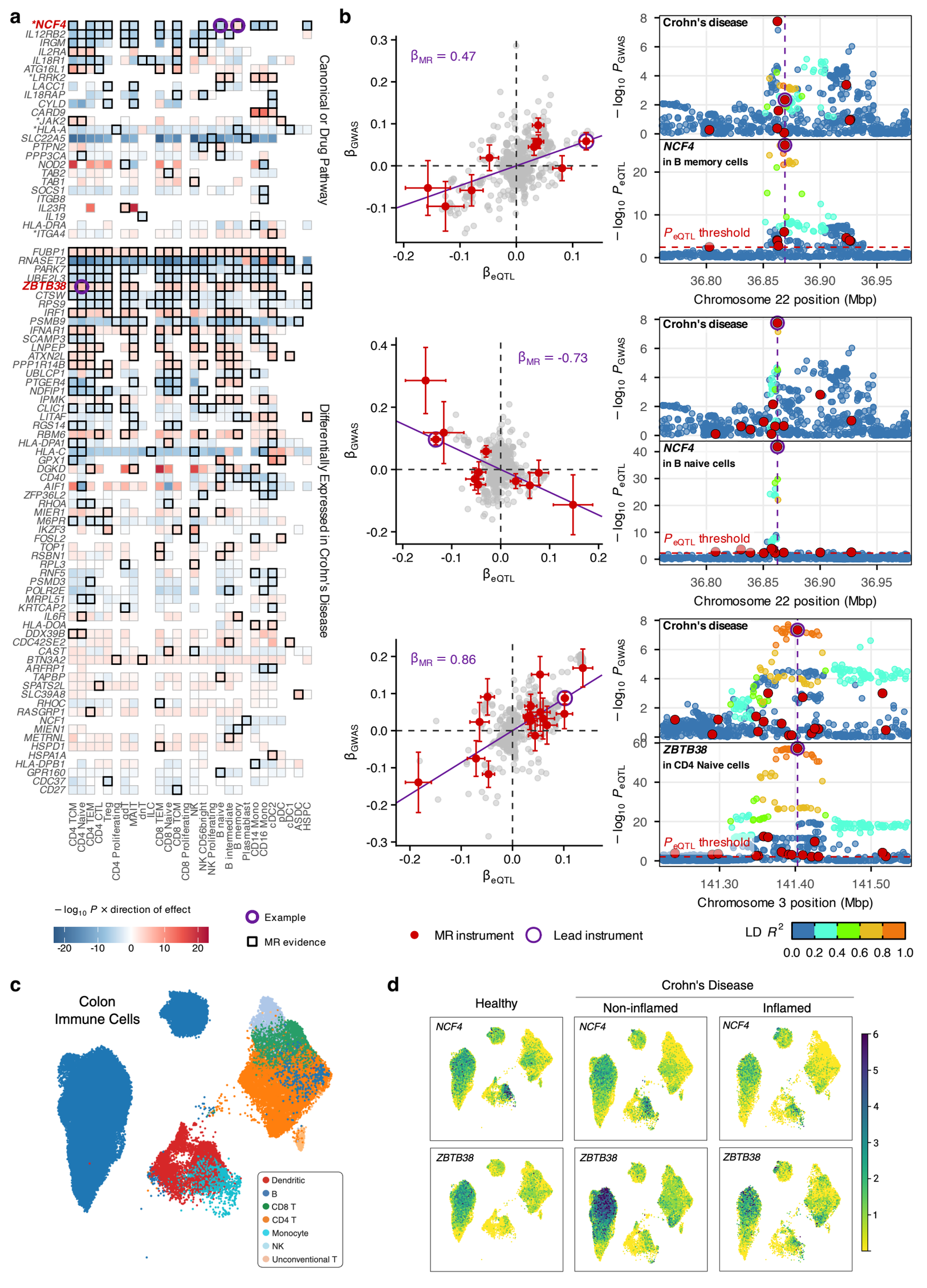
1054


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
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 relevance score
% 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
+
−
A
B
C
D
Odds ratio for clinical development = A × D / B × C
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
15000
10000
5000
0
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
Gene type
Non-coding
Protein-coding
c
Effective GWAS sample size
200k
400k
600k
800k
Number of GWAS genes
6000
4000
2000
Blood
Skeletal
Immune
Cancer
Viral infection
Cardiometabolic
Respiratory
Neurologic
Psychiatric
Other
Basophil count
Mean corpuscular Hb concentration
Immature fraction of reticulocytes
Platelet distribution width
Neutrophil % of WBC
Monocyte % of WBC
Neutrophil count
Hematocrit
Hb concentration
Reticulocyte fraction of RBC
Red cell distribution width
Eosinophil count
High light scatter reticulocyte % of RBC
High light scatter reticulocyte count
Monocyte count
Mean reticulocyte vol.
Mean platelet vol.
Mean corpuscular Hb
White blood cell count
Lymphocyte count
Red blood cell count
Mean corpuscular vol.
Plateletcrit
Platelet count
Mean sphered corpuscular vol.
Femur neck BMD
Lumbar spine BMD
Skull BMD
Whole body BMD
Estimated heel BMD
Eczema
Systemic lupus erythematosus
Psoriasis
Rheumatoid arthritis
Ulcerative colitis
Type 1 diabetes
Multiple sclerosis
Crohn's disease
Inflammatory bowel disease
Gastro-oesophageal cancer
Thyroid cancer
Ovarian cancer
Kidney cancer
Lymphocytic leukemia
Oropharyngeal cancer
Hepatic cancer
Bladder cancer
Malignant lymphoma
Breast cancer (HER2 Pos.)
Pancreatic cancer
Clonal hematopoiesis
Non-Hodgkins lymphoma
Lung cancer - never smokers
Cervical cancer
Breast cancer (Luminal B)
Lung adenocarcinoma
Small cell lung carcinoma
Breast cancer (Triple Neg.)
Lung cancer - ever smokers
Squamous cell lung carcinoma
Breast cancer (BRCA1/BRCA2)
Lung cancer
Basal cell carcinoma
Breast cancer
Colorectal cancer
Breast cancer (Overall)
Prostate cancer
Cutaneous melanoma
COVID-19 Vaccine Dose 2
COVID-19 Vaccine Dose 1
COVID-19 susceptibility
COVID-19 (Severe)
COVID-19 (Hospitalised)
Stroke
SCAD
Dilated cardiomyopathy
Heart failure
Coronary artery disease
Type 2 diabetes
Idiopathic pulmonary fibrosis
COPD
Asthma
Parkinson's disease
Migraine
Alzheimer's disease
Bipolar II disorder
Autism
ADHD
Bipolar I disorder
Bipolar disorder
Major depression
Schizophrenia
PCOS
Endometriosis
Primary open-angle glaucoma
Osteoarthritis


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
Asthma 597
Crohn's disease 510
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
Whole body BMD 246
Psoriasis 226
Breast cancer (Luminal A) 165
Alzheimer's disease 126
Bipolar disorder 122
Skull BMD 119
Basal cell carcinoma 100
Eczema 98
Lung cancer 76
Cervical cancer 51
COVID-19 (Hospitalised) 50
COVID-19 (Severe) 49
Lung cancer - ever smokers 48
Squamous cell lung carcinoma 39
Bipolar I disorder 38
Femur neck BMD 29
Lumbar spine BMD 24
COPD 14
COVID-19 Vaccine Dose 1 13
COVID-19 susceptibility 12
Parkinson's disease 10


a
Blood
11% High light scatter reticulocyte count
12% Immature fraction of reticulocytes
13% High light scatter reticulocyte % of RBC
14% Hematocrit
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.
34% Neutrophil % of WBC
35% Basophil count
36% Mean platelet vol.
38% Neutrophil count
65% Monocyte count
68% Monocyte % of WBC
Skeletal
12% Lumbar spine BMD
15% Estimated heel BMD
15% Whole body BMD
16% Femur neck BMD
19% Skull BMD
Immune
18% Eczema
19% Psoriasis
21% Crohn's disease
25% Rheumatoid arthritis
27% Inflammatory bowel disease
28% Type 1 diabetes
37% Ulcerative colitis
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
CD14 Mono
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 15.8%
cDC1 1.1 ± 1.1 29.2%
NK CD56bright 0.6 ± 1.0 14.9%
ASDC 0.6 ± 1.2 18.4%
ILC 0.4 ± 1.1 12.1%
CD14 Mono 0.8 ± 1.0 16.8%
CD4 TEM 0.1 ± 1.1 7.8%
CD16 Mono 0.5 ± 1.0 11.5%
NK 0.1 ± 1.0 6.9%
gdT -0.1 ± 1.1 6.1%
B memory 0.3 ± 0.9 6.9%
Treg 0.0 ± 1.1 5.7%
B intermediate 0.1 ± 1.0 5.0%
B naive 0.0 ± 0.9 3.0%
CD4 TCM -0.2 ± 1.1 4.4%
dnT -0.2 ± 1.0 3.7%
NK Proliferating -0.3 ± 1.0 3.3%
CD4 Proliferating -0.5 ± 0.9 1.8%
CD8 TEM -0.2 ± 1.0 3.6%
CD8 TCM -0.3 ± 1.0 3.4%
CD4 CTL -0.3 ± 1.0 3.2%
CD8 Proliferating -0.4 ± 1.0 2.0%
CD4 Naive -0.4 ± 1.0 2.3%
Plasmablast -1.2 ± 0.7 0.3%
pDC -1.3 ± 0.9 0.3%
HSPC -1.3 ± 1.0 0.4%
CD8 Naive -0.7 ± 1.0 1.6%
-5.00 -2.50 0.00 2.50 5.00
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)
Crohn's 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