nature genetics

Article

https://doi.org/10.1038/s41588-026-02681-0

High-resolution promoter interaction analysis implicates genes involved in activation of type 3 innate lymphoid cells in immune disease risk

Received: 28 October 2022

Accepted: 17 June 2026

Published online: 4 August 2026

Check for updates

Valeriya Malysheva[1,2,3,4,5,6,42] ✉, Helen Ray-Jones[1,2,3,7,42], Nora Lakes[8,9,10,42], Rachel A. Brown[11,12], Tareian A. Cazares[8,13,31], Owen Clay[10,14,32], David E. Ohayon[10], Pavel Artemov[1,2,33], Joseph A. Wayman[13], Zi F. Yang[13], Monica Della Rosa[1,2,34], Carmen Petitjean[1,2,35,36], Clarissa Booth[15], Joseph I. J. Ellaway[1,2,37], Jenna R. Barnes[9,10], Andrew W. Dangel[11,12], Ankita Saini[11,12], William R. Orchard[1,16,38], Xiaoting Chen[10], Sreeja Parameswaran[10], Frances Burden[17,18,39], Mattia Frontini[17,18,19], Takashi Nagano[20,40,41], Peter Fraser[20,21], Stefan Schoenfelder[20], Matthew T. Weirauch[8,9,10,22,23,24,25], Leah C. Kottyan[8,9,10,22,25], David F. Smith[22,26,27], Nick Powell[28], Jill M. Weimer[15], Eugene M. Oltz[11,12], Chris Wallace[29,30,43], Emily R. Miraldi[8,9,13,22,24,43], Stephen N. Waggoner[8,9,10,14,22,43] ✉ & Mikhail Spivakov[1,2] ✉

Innate lymphoid cells (ILCs) are rare tissue-resident lymphocytes that functionally mirror cells of CD4$^+$ T helper lineage but lack antigen receptors. Type 3 ILCs (ILC3s) are enriched at barrier sites, regulating inflammation and promoting tissue integrity. Here we profile the promoter-anchored chromosomal contacts of primary human ILC3s using low-input, high-resolution targeted chromosome conformation capture and compare them with those in CD4$^+$ T cells. We use these data to link Crohn's disease genome-wide association study variants with target genes, implicating both known and unanticipated candidates, including *CLN3*, a causal gene for Batten disease. We show that *Cln3* overexpression in a mouse ILC3-like cell line alters stimulation-induced transcriptional programs and cytokine secretion. Extending our approach to five additional immune genome-wide association study traits reveals enrichment for regulators of ILC3 activation. Our work develops methods, maps long-range gene regulation in ILC3s, and prioritizes immune disease risk genes with roles in this clinically relevant immune cell type.

Innate lymphoid cells (ILCs) are tissue-resident lymphocytes that mirror cells of CD4$^+$ T helper lineage but lack antigen receptors, responding instead to environmental cues through cytokine production and immunomodulatory activity[1,2]. Three major ILC groups have been defined based on cytokine profiles and developmental regulators[2,3]. Among these, type 3 ILCs (ILC3s) are enriched at barrier sites, including the gut, where they regulate inflammation and epithelial integrity[2–4]. ILC3-derived cytokines such as IL-17 and IL-22 promote mucosal homeostasis, and their overexpression has been linked to the development or exacerbation of Crohn's disease[5–8].

Immune disorders, including Crohn's disease, are known to have a genetic component, with genome-wide association studies (GWAS) identifying hundreds of disease susceptibility variants associated with these conditions[9]. Given the importance of ILC3s in immune control, it is likely that some of these variants affect ILC3 function. However, as most GWAS variants are noncoding and these studies are, by design,

A full list of affiliations appears at the end of the paper. ✉e-mail: valeriya.malysheva@vib.be; stephen.waggoner@cchmc.org; mikhail.spivakov@lms.mrc.ac.uk

cell-type agnostic, identification of causal genes and cell types implicated by GWAS variants is challenging. High-resolution chromosome conformation capture (Hi-C) techniques such as promoter capture Hi-C (PCHi-C) enable sensitive detection of promoter-anchored interactions and prioritization of target genes of noncoding GWAS variants that reside in distal regulatory regions such as enhancers[10,11]. However, it has been difficult to apply these techniques to rare cell types such as ILC3s.

Here we use a low-input PCHi-C protocol to map the three-dimensional regulatory architecture of primary human ILC3s, alongside CD4$^+$ T helper cells as an abundant adaptive immune population. We link promoters to distal regulatory elements based on these data with a combination of statistical interaction detection and a PCHi-C-adapted activity-by-contact (ABC) model. To integrate the enhancer–promoter contact maps with GWAS data, we develop multiCOGS, a Bayesian gene prioritization framework incorporating summary statistics imputation and multivariate fine-mapping. This strategy identifies both known and previously unrecognised candidate genes including *CLN3*, a Batten disease-causing gene, which we validate functionally in a mouse ILC3-like cell line. We extend ILC3 PCHi-C-based multiCOGS analysis to five additional immune diseases and show that the prioritized genes are enriched for regulators of ILC3 activation. Together, our results reveal *cis*-regulatory circuitries in ILC3s and prioritize GWAS effector genes implicating this rare, clinically relevant cell type in Crohn's disease and other immune disorders.

Results

A compendium of promoter-anchored chromosomal contacts in human ILC3s

To profile promoter-anchored chromosomal contacts in ILC3s, we used our low-input *DpnII*-based PCHi-C protocol[12,13] in ILC3s isolated from human tonsils (Fig. 1a). Significant promoter contacts were detected with CHiCAGO[14] at single-fragment resolution, as well as after pooling 'other end' fragments into ~5-kb bins, while leaving the baited promoter-containing *DpnII* fragment unbinned. Using this approach, we detected 31,003 contacts between promoters and promoter-interacting regions (PIRs) at single-fragment resolution and 58,632 contacts in 5-kb bins (Fig. 1b and Supplementary Table 1; Data S1 and S2 at OSF[15]). Binning resulted in detection of longer-range contacts, as we reported previously in other cell types[16] (Fig. 1c,d). For further analyses, we merged PIR sets detected at both resolutions (Methods).

A joint UMAP embedding of ILC3 promoter interaction profiles with those detected in 17 abundant blood cell types using *HindIII*-based PCHi-C[17] segregated ILC3s with other lymphoid cell types, consistent with patterns of promoter interactions reflecting lineage history (Extended Data Fig. 1a; Data S3 at OSF[15]; Methods). In total, more than half of PIRs (52.2%, 16,087 of 30,830) contacted at least one promoter in both ILC3s and another blood cell type (Data S3 at OSF[15]). Notably, the increased resolution afforded by using four-base cutter *DpnII* rather than six-base cutter *HindIII* enabled capture of alternative transcription start sites for 6,789 genes on separate baited fragments. Many of these genes displayed distinct interaction landscapes across isoforms (Extended Data Fig. 1b–d), as exemplified by the interaction patterns of three *INPP4B* alternative transcription start sites (Fig. 1e).

We next used public epigenetic data in human tonsil-derived ILC3s[18] to annotate PIR chromatin states. As expected, ILC3 PIRs were enriched for markers of accessible and/or active enhancers (assay for transposase-accessible chromatin (ATAC) peaks and histone H3 lysine 27 acetylation (H3K27ac)) and active transcription (histone H3 lysine 4 trimethylation (H3K4me3)), with 18,231 PIRs bearing combinations of these marks ('active PIRs'; Fig. 1f). Nearly two-thirds of active PIRs (65.9%, 12,021 of 18,231) overlapped with active regulatory regions annotated in the Ensembl Regulatory Build (ERB) ('promoter', 'promoter flanking regions', 'open chromatin'; Methods)[19], and 47.5% (8,658 of 18,231) overlapped with ERB-annotated CTCF binding sites (Fig. 1g). However, only 3% (636 of 18,231) of these PIRs overlapped with ERB-annotated enhancers[20], and nearly 20% (3,425 of 18,231) did not overlap with any ERB annotations (Fig. 1g).

Finally, to probe the relationship between enhancer–promoter connectivity and gene expression, we integrated these data with publicly available single-cell RNA sequencing (RNA-seq) profiles of human mucosal tissue ILC3s[21]. In agreement with studies in other cell types[17], we observed significant positive associations between numbers of active PIRs and gene expression, at both fragment and 5-kb resolution (Fig. 1h).

Enhancer–promoter contacts identified by ABCC complement significant promoter interactions

To further increase the sensitivity of detection of functional enhancer–promoter chromosomal interactions, we adapted the ABC approach[22] to PCHi-C data. In contrast to CHiCAGO, which detects statistically significant interactions relative to a distance-dependent background, ABC, which was originally developed for Hi-C, considers any observed contact frequency between a chromatin region and a promoter to be functionally meaningful. In addition, whereas CHiCAGO scores are independent of enhancer activity levels at PIRs, ABC scores incorporate both contact frequency and enhancer activity[22].

In our ABC adaptation, termed activity-by-captured-contact (ABCC), the contact frequencies are estimated from PCHi-C data that are imputed based on CHiCAGO-generated statistical models (Fig. 2a, Extended Data Fig. 2a,b and Methods). We validated ABCC using public enhancer perturbation data[22] and confirmed that this approach could accurately identify functional enhancer–promoter interactions in a lineage-specific manner and provide regulatory information complementary to CHiCAGO-detected contacts (Extended Data Fig. 2 and Supplementary Note).

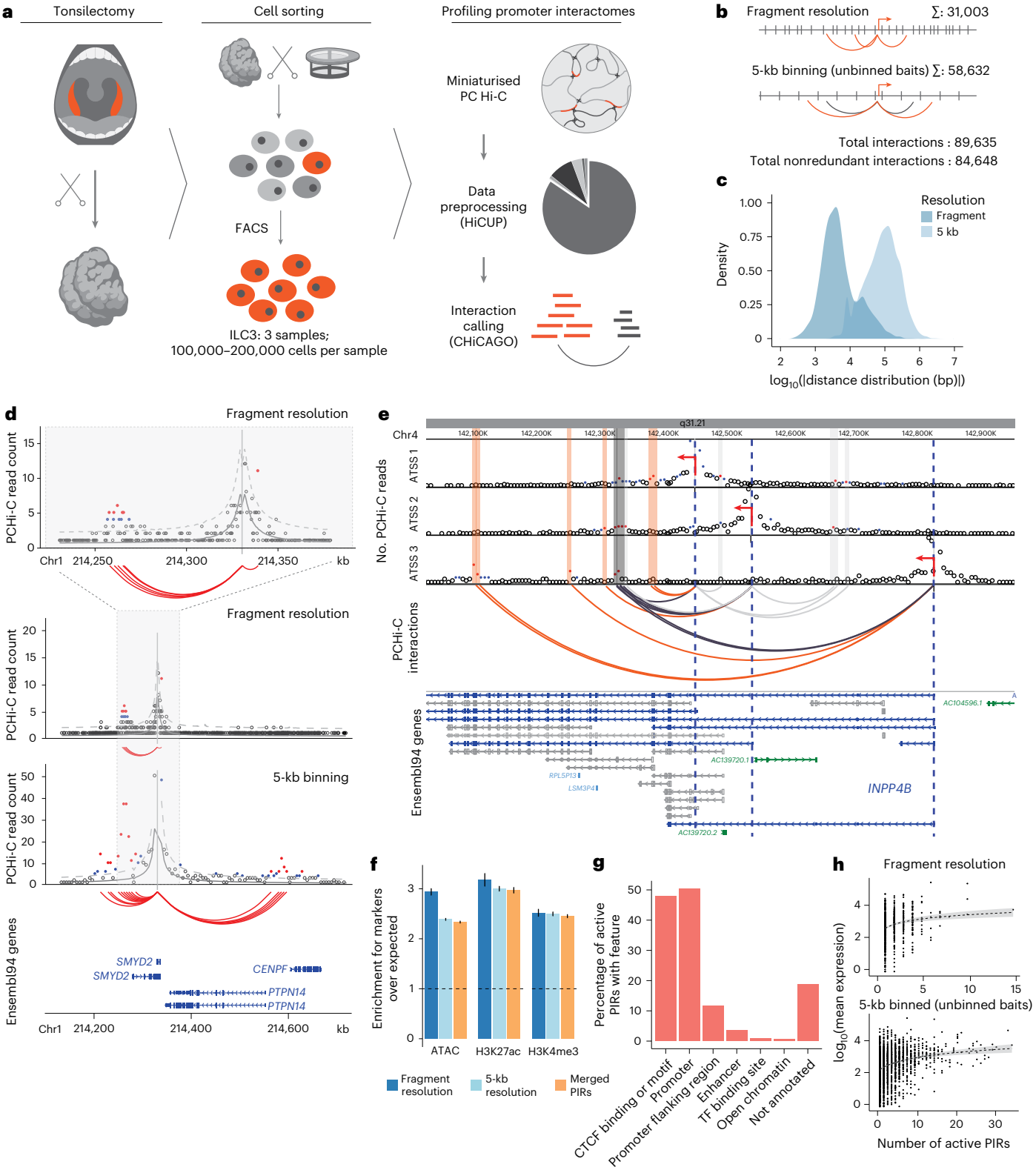
We applied ABCC to ILC3 PCHi-C data, producing 18,877 putative enhancer–promoter pairs across 17,690 genes (Data S4 at OSF[15]). Similar to the findings for CHiCAGO-defined PIRs, we observed positive

Fig. 1 | Compendium of promoter interactions in ILC3s. a, Outline of the study. **b**, Chromosomal interaction binning strategy. The analysis is done in two modes: fragment resolution (no binning) and 5-kb binning. In the 5-kb binning mode, the baited (captured) fragment containing a gene promoter is left unbinned to enable high-resolution linkage between the promoter and distal enhancers. Interactions uniquely detected in one mode only are shown as red arcs, and those detected in both modes are shown as gray arcs. The number of significant interactions is given for each mode individually and merged across both modes (Methods). **c**, Comparison of promoter–PIR distance distributions for PIRs detected at fragment and 5-kb resolution. **d**, Example of chromosomal interactions for the *SMYD2* gene at fragment and 5-kb resolution. The inset shows a zoomed-in view of the promoter interactions detected at fragment resolution. **e**, Example of multiple degrees of contact sharing between alternative promoters for the *INPP4B* gene. Captured alternative promoters are indicated by red arrows and blue dashed lines. The transcripts driven by these promoters (based on Ensembl 94) are shown in blue, and other *INPP4B* transcripts are shown in gray. Transcripts for processed pseudogenes are shown in light blue and long intergenic noncoding RNAs in green. PIRs are categorized as fully shared between alternative promoters (dark gray arcs), partially shared (light gray arcs) or distinct (red arcs). **f**, Enrichment of PIRs for markers of active enhancers and promoters (H3K27ac and H3K4me3) and accessible chromatin (ATAC) in ILC3s. Bars represent fold enrichment versus distance-matched background, with error bars denoting 95% confidence intervals from CHiCAGO permutation tests. **g**, Annotation of PIRs bearing active histone marks and/or ATAC-seq peaks in ILC3s ('active PIRs') according to the ERB. **h**, Associations between gene expression and numbers of regulatory elements identified in CHiCAGO PIRs at fragment and 5-kb resolution (leaving the baited fragments unbinned). The dashed line reflects a generalized additive model fit (smooth term estimate e.d.f. = 1 and 1.96, respectively; $P < 2 \times 10^{-16}$) and the gray ribbon its confidence interval. FACS, fluorescence-activated cell sorting.

https://doi.org/10.1038/s41588-026-02681-0

associations between numbers of ABCC-linked enhancers per gene and gene expression (Fig. 2b). However, ABCC interactions generally spanned shorter distances than CHiCAGO-detected pairs (median distance ~69 kb versus ~108 kb, respectively; Wilcoxon rank-sum test $P < 2.2 \times 10^{-16}$; Fig. 2c), and the two sets of contacts showed limited overlap (8.4%; Data S5 at OSF[15]). As expected from their definition, ABCC-linked enhancers were strongly enriched for features of active or accessible chromatin (Fig. 2d). They also showed enrichment for CTCF binding sites comparable to that of CHiCAGO-defined PIRs (Fig. 2d).

For downstream analyses, we combined ABCC- and CHiCAGO-defined promoter contacts (referred to collectively as PIRs hereafter). Representative examples of the regulatory landscapes detected by both approaches are shown in Fig. 2e,f.



https://doi.org/10.1038/s41588-026-02681-0

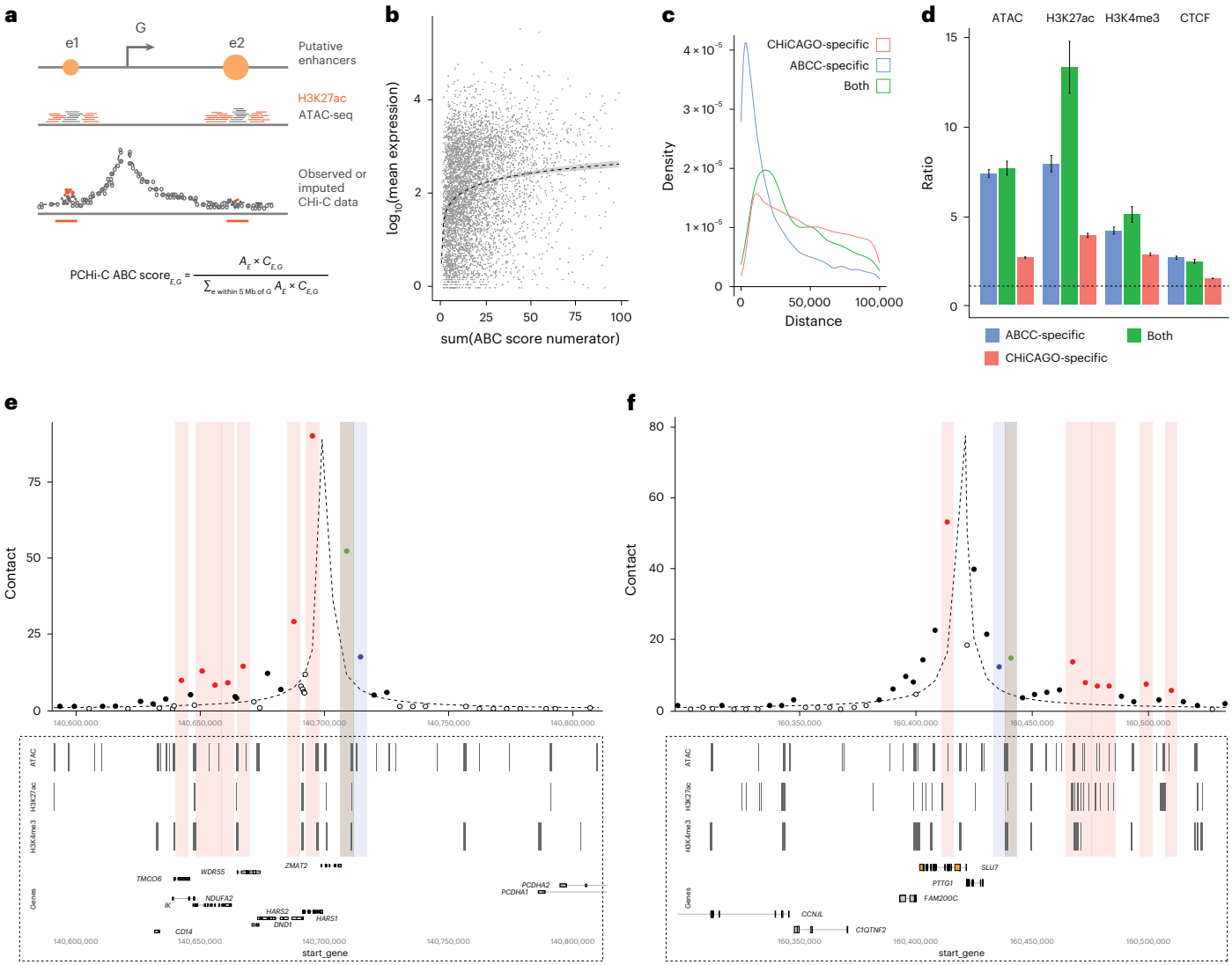


Fig. 2 | Combining ABCC and CHiCAGO to link distal elements with target genes. a, Schematic depicting adaptation of the ABC model for use with PCHi-C data, termed ABCC. **b**, Associations between gene expression and ABC numerator score summed across all predicted enhancers per gene. The dashed line shows a generalized additive model fit (smooth term estimate e.d.f. = 2.485, $P < 2 \times 10^{-16}$) and the gray ribbon its confidence interval. **c**, Interaction distance comparison across CHiCAGO-specific, ABCC-specific and shared interactions. **d**, Enrichment for markers of active and accessible chromatin within CHiCAGO-specific, ABCC-specific and shared PIRs. Bars represent fold enrichment versus distance-matched background, with error bars denoting 95% confidence intervals from CHiCAGO permutation tests. **e**,**f**, Representative examples of CHiCAGO- and ABCC-detected contacts for *ZMAT2* (**e**) and *SLU7* (**f**) promoters. The dashed line shows expected counts estimated using the CHiCAGO distance function. PIRs detected with CHiCAGO at 5-kb resolution are shown as red dots and shading, those detected with ABCC as blue dots and shading, and those detected by both approaches as green points and shading. Black-filled dots represent imputed counts considered by ABCC, corresponding to the maximum value between observed and expected counts. Unfilled dots represent observed counts falling below expected values.

Promoter interactomes in ILC3s and CD4$^+$ T cells reveal shared and differential regulatory circuits

ILC3s have developmental similarities[23,24] and share core 'immune modules' with CD4$^+$ T helper cells[24–26]; this prompted us to use this abundant cell type for comparative analysis and identification of ILC3-specific regulatory circuits. To this end, we generated and processed high-resolution PCHi-C data for CD4$^+$ T cells using the same protocol, identifying 31,252 and 87,348 interactions at single-fragment and 5-kb resolution, respectively (Data S6 and S7 at OSF[15]). In addition, we detected 30,258 enhancer–gene pairs in CD4$^+$ T cells with ABCC across 16,956 genes (Data S8 and S9 at OSF[15]), 30% of which were shared with ABCC pairs identified in ILC3s.

Differential analysis of promoter interactions between ILC3s and CD4$^+$ T cells with Chicdiff[27] revealed a total of 19,038 cell-type-specific contacts across 3,664 genes (1,818 at fragment resolution and 17,220 at 5-kb resolution; weighted adjusted $P < 0.05$; Fig. 3a and Extended Data Fig. 3). As expected, we also detected a significant association between differential interactions and differential expression (chi-squared test $P = 9.95 \times 10^{-7}$) (Fig. 3b). Pathway enrichment analysis revealed immune regulatory programs associated with ILC3-specific and CD4$^+$ T cell-specific contacts (Fig. 3c, Supplementary Tables 2–5 and Supplementary Note). Overall, this comparative analysis highlighted both shared and distinct *cis*-regulatory wiring of ILC3s and CD4$^+$ T cells, reflecting their specialized roles in innate versus adaptive immune responses and coordinated regulation of immune activation pathways.

PIRs in ILC3s and CD4$^+$ T cells are enriched for immune disease-associated variants

We used the RELI algorithm[28] to detect enrichment of active PIRs for variants associated with 495 healthy and pathological GWAS traits listed in the GWAS Catalog[29] (Fig. 4a and Methods).

https://doi.org/10.1038/s41588-026-02681-0

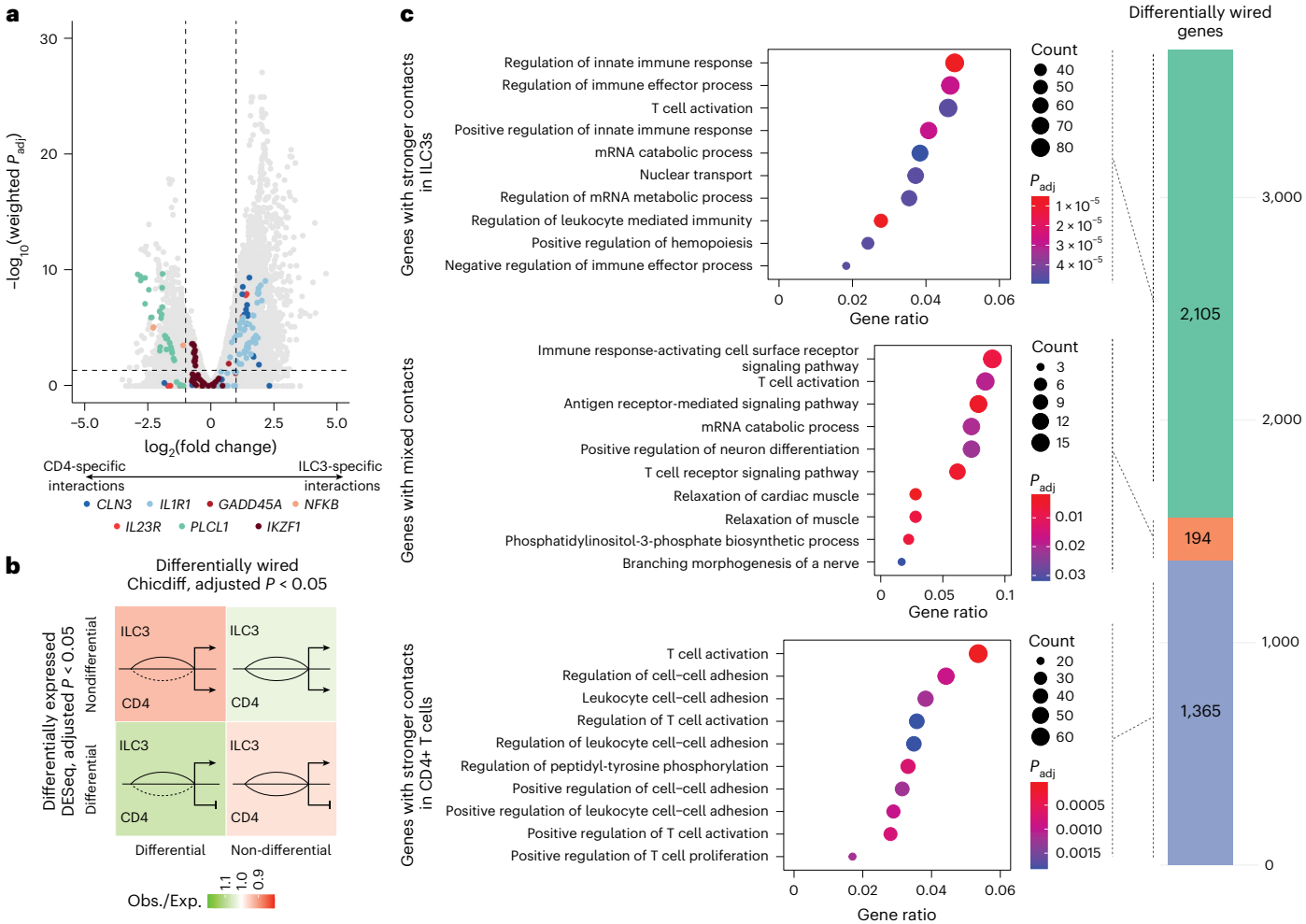


Fig. 3 | Differential enhancer–promoter interactions between ILC3s and CD4$^+$ T cells. a, Volcano plot of differential interactions between ILC3s and CD4$^+$ T cells detected by Chicdiff[27], highlighting those of selected immune-related genes (*CLN3*, *IL1R1*, *GADD45A*, *NFKB*, *IL23R*, *PLCL1*, *IKZF1*). Statistical significance was evaluated using two-sided Wald test within a negative binomial generalized linear model framework, with *P* values adjusted for multiple testing across all interactions using independent hypothesis weighting[84]. **b**, Relationship between differential expression (DESeq2, adjusted $P < 0.05$) and differential wiring of promoter contacts (Chicdiff, adjusted $P < 0.05$). Statistical significance was determined using Pearson's chi-squared test with Yates' continuity correction ($\chi^2 = 23.938$, d.f. = 1, exact $P = 9.95 \times 10^{-7}$). **c**, Gene ontology (GO) enrichment analysis of genes with stronger contacts in ILC3s (top), CD4$^+$ T cells (bottom) or a mixture of contacts that were stronger in either cell type (middle), showing biological processes related to immune cell activation, adhesion and differentiation. Pathway enrichment was assessed using one-sided hypergeometric test, with *P* values adjusted globally for multiple comparisons using the Benjamini–Hochberg false discovery rate procedure (adjusted $P < 0.05$ was considered to indicate significance; full results are listed in Supplementary Tables 2–5). Bubble size reflects the number of genes; color indicates adjusted *P* values. The bar plot shows the overlap between differentially wired genes (as evaluated by Chicdiff) in ILC3s and CD4$^+$ T cells. P_{adj}, adjusted *P* value.

In ILC3s, RELI revealed significant enrichment for variants associated with 21 of 495 traits (Bonferroni-adjusted $P < 0.05$; Fig. 4b, Extended Data Fig. 4a and Supplementary Table 6). Autoimmune diseases (EFO:0005140) affecting a broad array of tissues and organs where ILC3s are known to reside were overrepresented among these traits (one-tailed hypergeometric test, $P = 1.08 \times 10^{-5}$). These tissues and organs included the gut (Crohn's disease, celiac disease, ulcerative colitis, primary sclerosing cholangitis), airways (asthma, hay fever) and central nervous system (multiple sclerosis). We also noted enrichment for several peripheral blood cell traits, including platelet width, lymphocyte count and corpuscular volume (Supplementary Table 6).

In CD4$^+$ T helper cells, RELI detected significant enrichment for variants associated with 22 of 495 traits (Bonferroni-adjusted $P < 0.05$) (Extended Data Fig. 4b and Supplementary Table 7). There was a strong correlation between the traits enriched in the two cell types ($R^2 = 0.85$, correlation test $P = 5.2 \times 10^{-4}$) (Fig. 4b and Extended Data Fig. 4c), consistent with their overlapping *cis*-regulatory circuits. However, several traits displayed cell type specificity; these included allergic sensitization, mouth ulcers and IgG glycosylation in ILC3s, and primary biliary cirrhosis, rheumatoid arthritis and systemic lupus erythematosus in CD4$^+$ T cells (Fig. 4b, Extended Data Fig. 4c and Supplementary Table 7).

multiCOGS links imputed, fine-mapped Crohn's disease GWAS signals to candidate effector genes through promoter contacts in ILC3s and CD4$^+$ T cells

RELI analysis identified significant enrichment of Crohn's disease risk variants within active PIRs in both ILC3s and CD4$^+$ T helper cells (~2.3-fold; Bonferroni-adjusted $P = 6.96 \times 10^{-6}$ and $P = 1.19 \times 10^{-7}$, respectively), which we confirmed using stratified linkage disequilibrium (LD) score regression (Fig. 4c). Given the importance of both cell types in the pathogenesis of Crohn's disease[30–34], we next sought to prioritize effector genes linked to Crohn's disease risk variants in these cells using our PCHi-C data. To this end, we developed multiCOGS, a Bayesian prioritization algorithm that scores the association of each gene with a trait based on the location of fine-mapped GWAS

https://doi.org/10.1038/s41588-026-02681-0

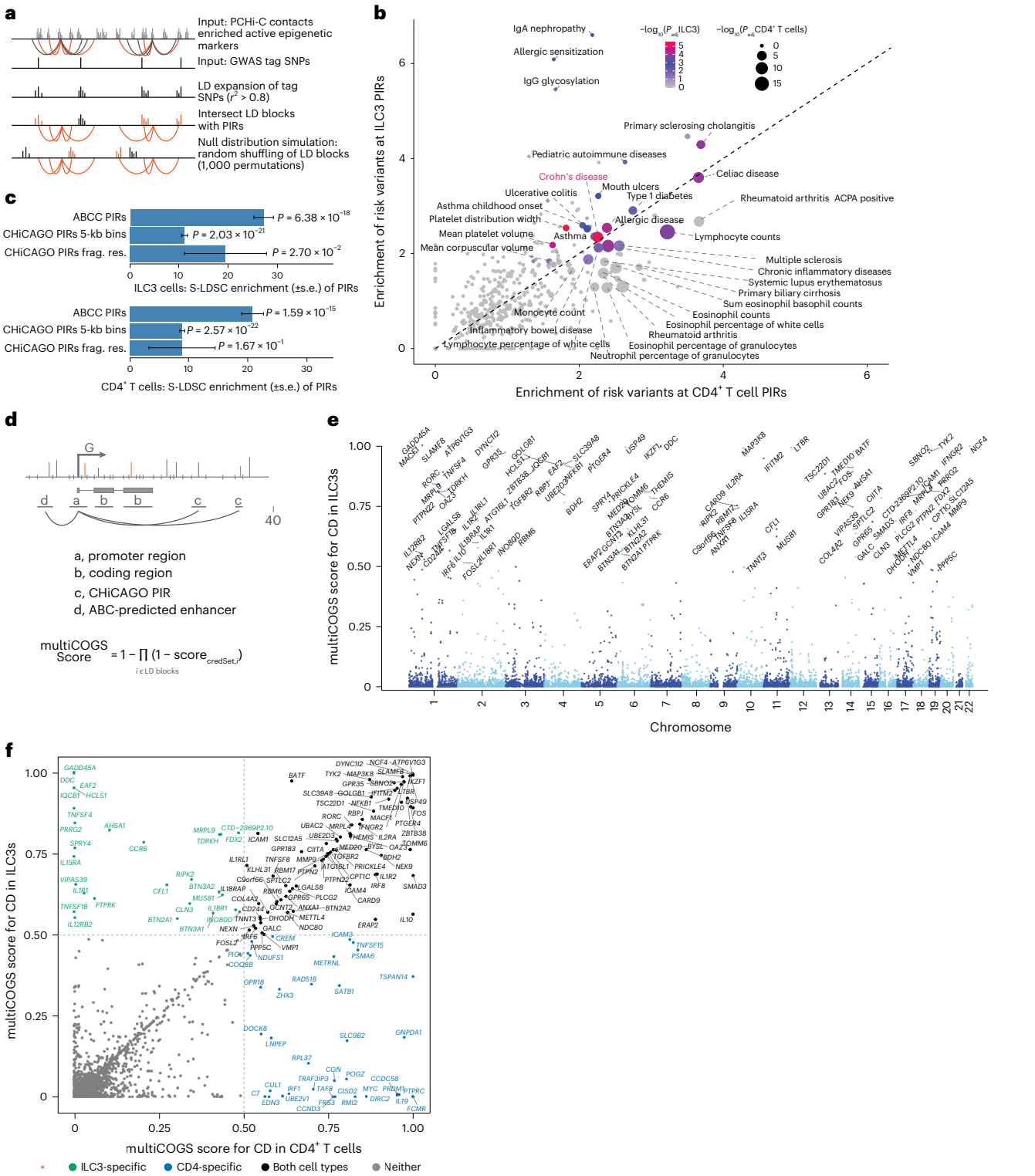


Fig. 4 | Statistical integration of PCHi-C results for ILC3s and CD4$^+$ T cells with GWAS enables gene prioritization for Crohn's disease. a, Schematic of the RELI algorithm used to estimate enrichment of genetic risk loci within PIRs. **b**, Enrichment of risk variants in ILC3 versus CD4$^+$ T cell 'active' PIRs across 495 diseases and traits, determined using the RELI method[28]. *P* values were adjusted for multiple testing using the Bonferroni procedure. Traits with adjusted *P* values < 0.05 in ILC3s and/or CD4$^+$ T cells are labeled. **c**, Stratified LD score regression analysis (using LDSC[85]) for enrichment of Crohn's disease risk heritability at PIRs of ILC3s and CD4$^+$ T cells. Bars show LDSC-estimated enrichment coefficients for each PIR category, and error bars show the standard errors reported by the model. Statistical significance was determined using one-sided *z*-test relative to the baseline model; exact *P* values are reported next to each bar. **d**, Schematic of the multiCOGS algorithm. **e**, Manhattan plot of multiCOGS gene prioritization scores for Crohn's disease risk based on integration of GWAS with promoter interactions in ILC3s. Genes with multiCOGS scores greater than 0.5 are labeled. **f**, Distribution of multiCOGS gene scores for Crohn's disease computed based on PCHi-C data in ILC3s and CD4$^+$ T cells. Genes are colored depending on whether they exceeded a multiCOGS score cutoff of 0.5 in ILC3s only (green), CD4$^+$ T cells only (blue) or both cell types (black). The rest of the genes are shown as gray dots. Frag. res., fragment resolution.

https://doi.org/10.1038/s41588-026-02681-0

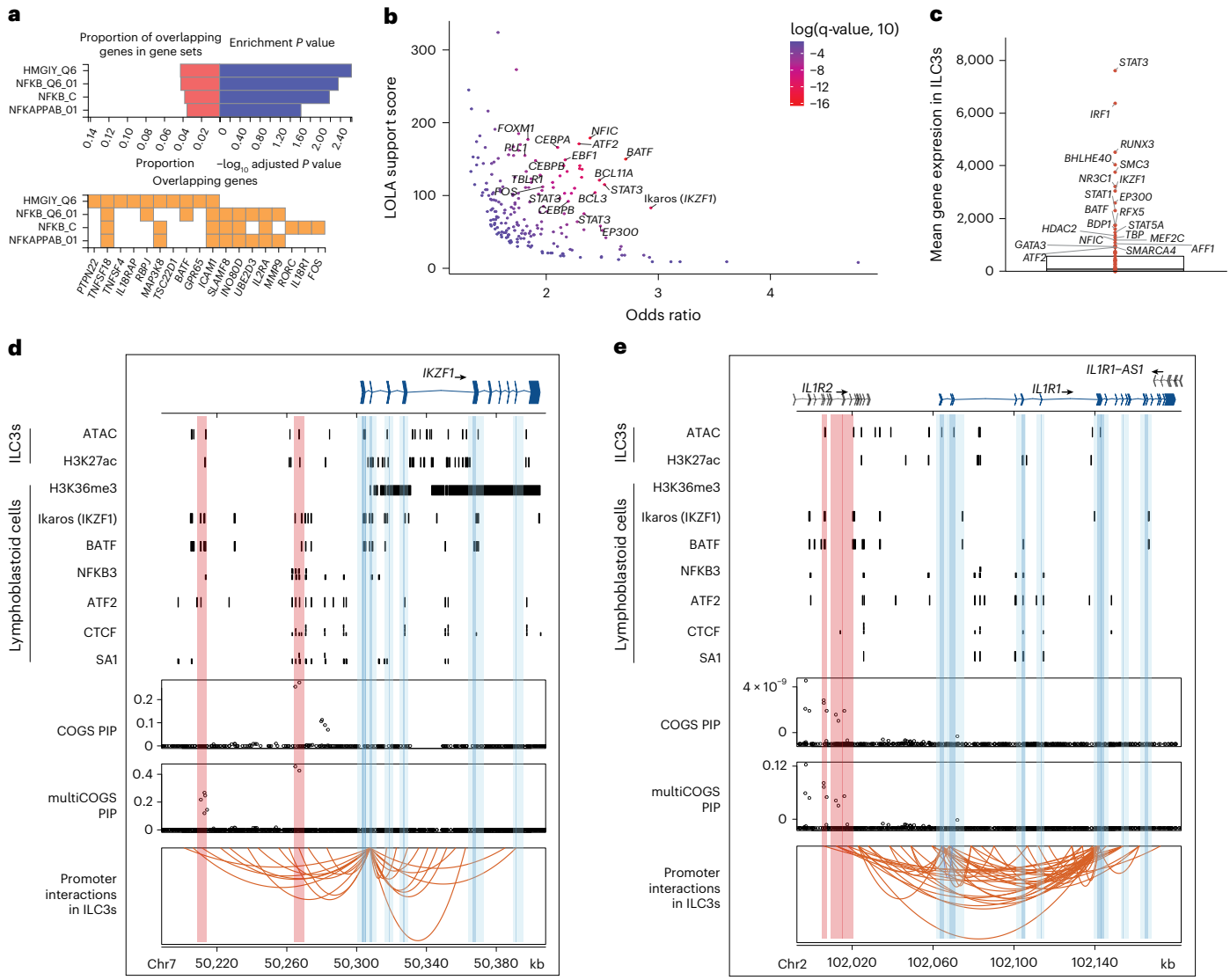


Fig. 5 | Characterization of genes associated with Crohn's disease risk prioritized by multiCOGS in ILC3s and their putative TF regulators. **a**, Significant sets of multiCOGS-prioritized genes predicted to bind specific TFs in their promoter regions, according to the MSigDB TF targets database, detected using the GENE2FUNC pipeline in FUMA (hypergeometric overrepresentation test)[86]. TF sets are labeled (rows), with the proportions of all multiCOGS genes per set and associated *P* values (adjusted for multiple testing via the Benjamini–Hochberg method) shown in the top panel and gene names in the bottom panel. **b**, Enrichment analysis for TF binding sites at active PIRs for genes prioritized by multiCOGS versus active PIRs of all genes submitted to multiCOGS analysis. **c**, Expression of TFs enriched at the PIRs of prioritized genes. Outliers were removed for clarity. **d**,**e**, Examples of genes prioritized by multiCOGS for Crohn's disease (*IKZF1* (**d**) and *IL1R1* (**e**)), showing patterns of TF binding in lymphoblastoid cell lines and posterior inclusion probability (PIP) profiles generated by the Wakefield fine-mapping procedure[87] assuming a single causal variant per LD block used in classic COGS ('COGS PIP') and the SuSiE[38] multivariate fine-mapping algorithm used in multiCOGS ('multiCOGS PIP'). Vertical dark blue and light blue bands, respectively, highlight annotated gene promoters and promoter-proximal regions (±5 restriction fragments) considered in (multi)COGS analysis in addition to PIRs. Vertical red bands highlight PIRs harboring Crohn's disease risk-associated SNPs with high PIPs. Orange arcs correspond to significant interactions (CHiCAGO score > 5) at 5-kb resolution for *IKZF1* and *IL1R1*, respectively.

signals within gene coding regions, gene promoters and PIRs (Fig. 4d and Methods). multiCOGS extends our previously published COGS[17,35] framework by incorporating summary-statistics-based imputation[36] and multivariate fine-mapping approaches such as SuSiE[37–39] to increase the accuracy and coverage of gene prioritization.

We applied multiCOGS to a Crohn's disease GWAS meta-analysis[40] using CHiCAGO- and ABCC-defined PIRs in ILC3s and CD4$^+$ T cells. At the established COGS score cutoff of 0.5 (ref. 17), multiCOGS prioritized 109 genes in ILC3s (Fig. 4e) and 118 genes in CD4$^+$ T cells (Extended Data Fig. 5a and Supplementary Tables 8 and 9), mostly through linking of noncoding, likely causal variants to gene promoters via CHiCAGO-defined contacts (Extended Data Fig. 5b). Notably, ABCC contributed to only ~11% of prioritized genes in both cell types (Extended Data Fig. 5c). The prioritized candidates had known functions in cytokine signaling (*IL10*, *IL1RL1*, *IL2RA*, *IFNGR2*), autophagy (*ATG16L1*, *GPR65*), antimicrobial responses (*PTPN2*, *IRF8*)[41,42] and transcriptional regulation (*FOS*, *TSC22D1*, *RBPJ*) and were enriched in immune pathways that have been implicated in inflammatory bowel disease (IBD)[43–45], including the IL6–JAK–STAT3, TNF–NF-κB and IL-23/T helper 17 cell pathways (Extended Data Fig. 6, Supplementary Table 10 and Supplementary Note). Comparison with standard COGS showed that multiCOGS substantially expanded and refined the prioritized genes (Extended Data Fig. 7, Supplementary Table 11 and Supplementary Note). Notably, at several loci, multiCOGS prioritized multiple

candidates through distinct credible sets. For example, at chr7p, both COGS and multiCOGS prioritized *IKZF1* (encoding the Ikaros transcription factor (TF)[46]) based on CHiCAGO-defined contacts. However, multiCOGS also prioritized DOPA decarboxylase gene *DDC*, which encodes a potential regulator of immune cell infiltration[47], based on a separate credible set overlapping ABCC-linked enhancers (Extended Data Fig. 7d).

We searched for prior evidence of association of multiCOGS-prioritized genes with Crohn's disease (or, more broadly, IBD) by querying them in OpenTargets, curated gene-to-disease databases and functional studies[48–52]. According to these sources, more than half of the genes prioritized in ILC3s (61 of 109) and CD4$^+$ T cells (67 of 118) had not previously been implicated in IBD (Supplementary Table 11). In particular, 20 of the 29 genes selectively prioritized in ILC3s and not CD4$^+$ T cells (Fig. 4f) had not previously been linked to IBD in the studied datasets. This set included genes such as *GADD45A* and *PTPRK* with known roles in immunity and T cell development[53,54] and some with unexpected functions, including *DDC* as mentioned above, as well as *CLN3*, which encodes a lysosomal protein whose mutations cause the neurodegenerative Batten disease[55].

Finally, we predicted potential transcriptional regulators of the prioritized loci in ILC3s using two complementary approaches. Gene-centric analysis identified HMGA1 and NF-κB as candidate regulators of genes prioritized in Crohn's disease (Fig. 5a and Supplementary Table 12), whereas region-centric analysis identified enrichment of the predicted binding sites for 97 TFs within their PIRs (Fig. 5b and Supplementary Table 12). These TFs included established regulators of inflammatory response, such as Ikaros (encoded by *IKZF1*)[46], BATF[56] and NFKB3 (also known as RELA)[57–59], all of which are highly expressed in ILC3s (Fig. 5c). For example, the promoter of *IKZF1* contacted two active PIRs harboring fine-mapped Crohn's disease variants in ILC3s, with predicted occupancy of Ikaros itself, BATF, NFKB3, ATF2, CTCF and SA1 inferred from lymphoblastoid cell line data (Fig. 5d). Similarly, the promoter of *IL1R1*, which encodes an ILC3-activating receptor[60], contacted active PIRs harboring Crohn's disease-associated variants and showing evidence of Ikaros, BATF and CTCF binding based on lymphoblastoid cell line data (Fig. 5e).

Overall, multiCOGS prioritization implicated inflammatory signaling genes and their upstream TF regulators as candidate mediators of Crohn's disease susceptibility in ILC3s.

CLN3 contributes to ILC3 inflammatory capacity

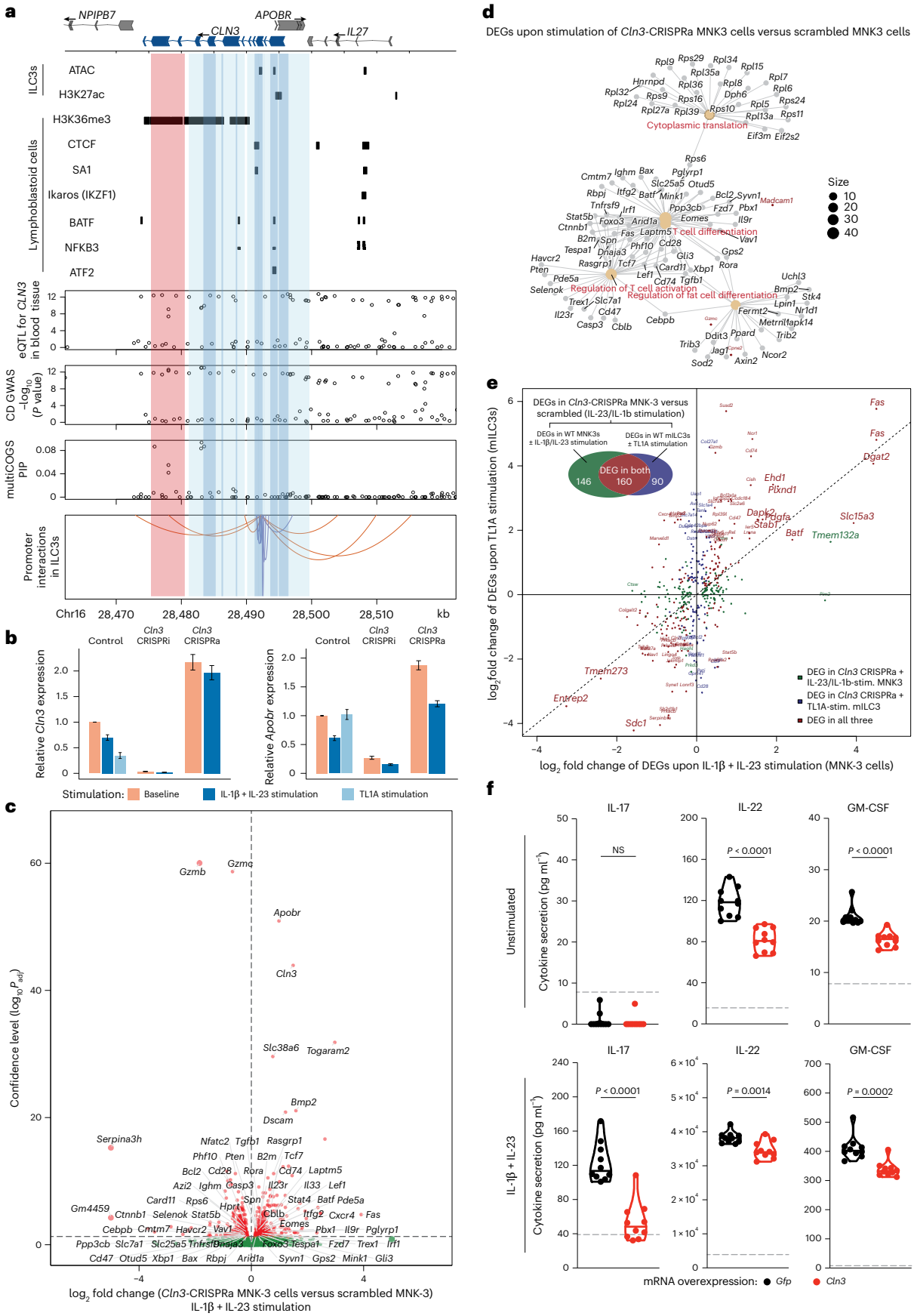
We next focused on *CLN3*, a Batten disease gene, which was selectively prioritized for Crohn's disease risk in ILC3s but not CD4$^+$ T cells and not previously linked to any immune-mediated diseases according to our database analysis. Examination of the SuSiE fine-mapped Crohn's disease GWAS locus underlying the prioritization of *CLN3* revealed a credible set of variants overlapping with two regions considered by multiCOGS, which were colocalized with expression quantitative trait loci for *CLN3* and nearby genes including *APOBR* (Fig. 6a and Supplementary Note).

To further investigate the role of the *CLN3* locus in ILC3s, we used mouse MNK-3 cells as a tractable model for ILC3 activation and effector function[61]. We found that *Cln3* expression was downregulated upon MNK-3 cell stimulation with IL-23 and IL-1β, cytokines that are essential for ILC3 effector function[62,63] (Fig. 6b, left). Consistent with this observation, analysis of published RNA-seq data from primary mouse ILC3s stimulated with TL1A[64] also showed reduced *Cln3* expression (Fig. 6b, left). Notably, adjacent gene *Apobr* was similarly downregulated under IL-23/IL-1β stimulation (Fig. 6b, right), in line with evidence of coordinated regulation of these genes in humans based on expression quantitative trait loci[65–68]. By contrast, TL1A stimulation did not affect *Apobr* expression (Fig. 6b, right).

To interrogate the transcriptional consequences of stimulation-induced *Cln3* repression, we used CRISPR activation (CRISPRa) based on dCas9–VP64 and MS2-p65–HSF1 to prevent *Cln3* downregulation in MNK-3 cells upon stimulation. CRISPRa targeted to the *Cln3* promoter produced an approximately threefold increase in *Cln3* expression in stimulated MNK-3 cells and an ~2.1-fold increase in the baseline condition (Fig. 6b, left). Notably, *Apobr* expression was also increased in both basal and stimulated conditions (Fig. 6b, right), potentially reflecting local effects of CRISPRa targeting but consistent with the coordinated regulation observed at this locus (Fig. 6b). Bulk RNA-seq analysis revealed widespread transcriptional changes following *Cln3*-CRISPRa, with 519 differentially expressed genes in unstimulated cells and 722 in stimulated cells relative to scrambled guide RNA (gRNA) controls (DESeq2 adjusted $P < 0.05$) (Fig. 6c, Extended Data Fig. 8a and Supplementary Table 13; Data S10 at OSF[15]). These genes were enriched for pathways involved in lymphocyte differentiation, activation and proliferation and included *Il23r*, *Cd74* and *Fas* (upregulated) and inflammatory serine protease genes *Gzmb* and *Gzmc* (downregulated) (Fig. 6d and Extended Data Fig. 8b). More than half the genes differentially expressed in stimulated *Cln3*-CRISPRa cells overlapped with genes altered by IL-23/IL-1β or TL1A stimulation in *Cln3*-unperturbed cells[64] (Fig. 6e), suggesting that sustained *Cln3* expression counteracts canonical activation-associated transcriptional programs. By contrast, CRISPR interference (CRISPRi; dCas9–KRAB)-mediated knockdown of *Cln3* resulted in few transcriptional changes beyond *Cln3* and *Apobr* themselves (Extended Data Fig. 8c,d and Supplementary Table 13; Data S11 at OSF[15]). The observed changes included upregulation of *Nos2*, which encodes a

Fig. 6 | Evidence of the role of CLN3 in ILC3 inflammatory function. **a**, Interaction profile of the human *CLN3* promoter alongside the tracks of TF binding, blood expression quantitative trait loci[65], Crohn's disease GWAS[40] – $\log_{10} P$ values and SuSiE PIP values. Dark blue and light blue bands, respectively, highlight locations of annotated *CLN3* promoters and promoter-proximal regions (±5 restriction fragments) considered by multiCOGS in addition to PIRs. Red band highlights the ILC3-specific PIR containing Crohn's disease-associated SNPs with high PIPs. Orange and purple arcs, respectively, depict significant interactions (CHiCAGO score > 5) in ILC3s at 5-kb and single-fragment resolution. Multiple vertical lines per TF reflect separate datasets. **b**, Up- and downregulation of *Cln3* and *Apobr* upon TL1A stimulation in mouse primary ILC3s (RNA-seq data from ref. 64) and upon IL-23/IL-1β stimulation in *Cln3*-targeted CRISPRi and CRISPRa MNK-3 cells (RNA-seq data from this study). Bars and error bars, respectively, indicate fold-change estimates and standard errors derived from DESeq2 (ref. 88) negative binomial regression models based on $n = 3$ replicates per condition. **c**, Differential expression of genes in IL-23/IL-1β-stimulated *Cln3*-CRISPRa MNK-3 cells relative to scrambled gRNA controls. Red dots show differentially expressed genes (stimulated *Cln3*-CRISPRa differentially expressed genes, DESeq2 adjusted $P < 0.05$), with other genes shown as green dots. **d**, Network-style representation of GO term enrichment analysis of stimulated *Cln3*-CRISPRa differentially expressed genes. **e**, Changes in expression of stimulated *Cln3*-CRISPRa differentially expressed genes (dots) upon either IL-23/IL-1β stimulation of unperturbed MNK-3 cells or TL1A stimulation of primary mouse ILC3s (data from ref. 64). Genes with $\log_2$-fold-changes greater than 1.5 in either condition are labeled in small letters and those with such changes in both conditions in large letters. **f**, Evidence that *Cln3* overexpression decreases inflammatory cytokine secretion. MNK-3 cells were electroporated with *GFP* mRNA (black) or *Cln3–Myc* mRNA (red) and then cultured either unstimulated (top row) or stimulated with IL-1β and IL-23 (bottom row) for 24 h. Cytokine concentrations (IL-17, IL-22, GM-CSF) in culture supernatants were quantified by enzyme-linked immunosorbent assay. Each point represents an individual biological replicate ($n = 10$ per condition). Data shown are from one representative experiment of three independent experiments performed. The dotted line indicates the lower limit of quantification for each assay. Statistical significance was assessed using unpaired two-sided Welch's *t*-test. DEGs, differentially expressed genes; NS, not significant; WT, wild type.

https://doi.org/10.1038/s41588-026-02681-0



https://doi.org/10.1038/s41588-026-02681-0

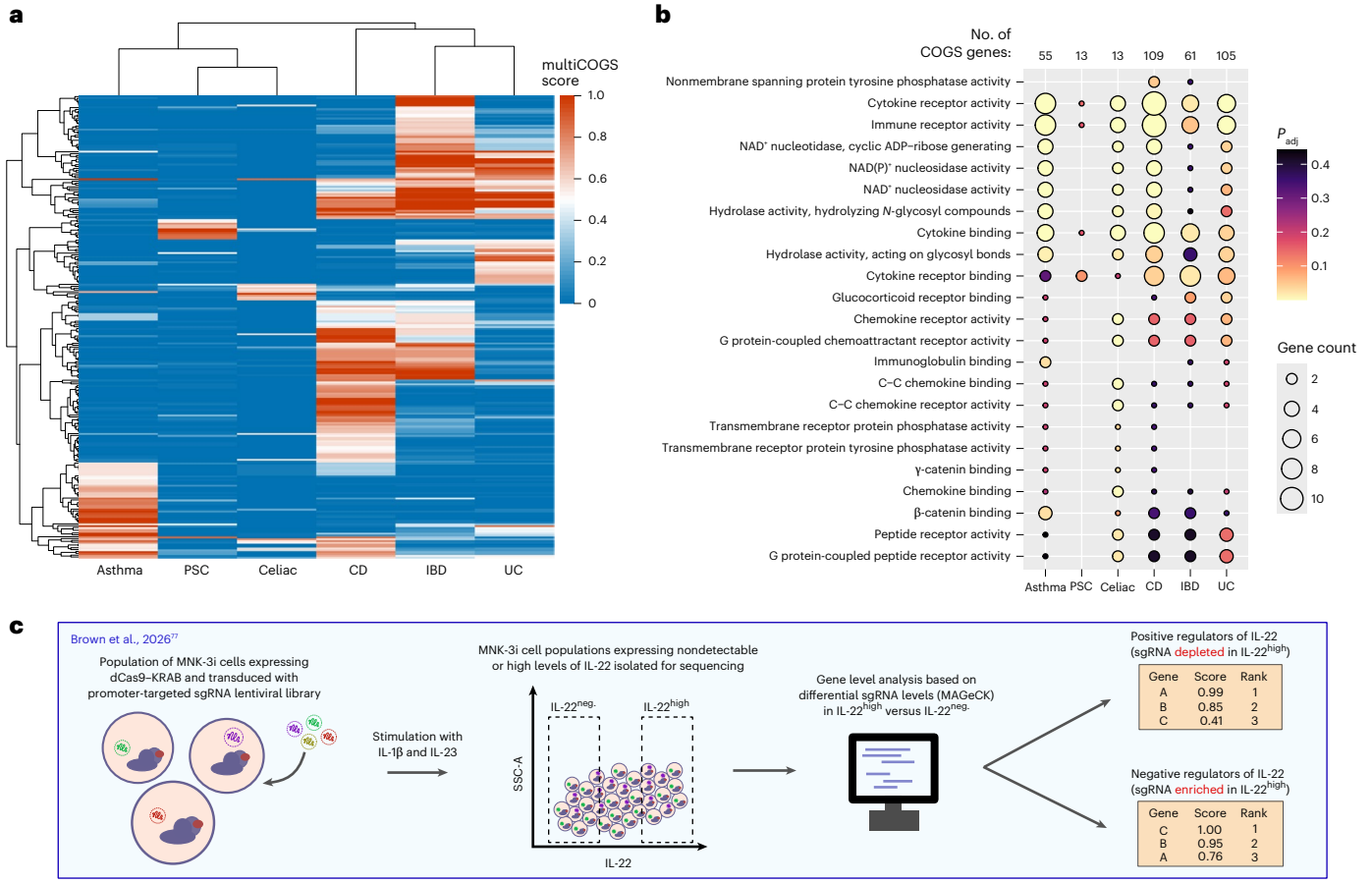


Fig. 7 | A compendium of prioritized genes in ILC3s for six immune-related diseases. a, multiCOGS results across asthma, primary sclerosing cholangitis, celiac disease, Crohn's disease, IBD and ulcerative colitis in ILC3 cells. Rows represent each gene that scored at least 0.5 in one of the traits. Colors show the multiCOGS score in each trait. Clustering on genes (rows) and traits per cell type (columns) was based on Euclidean distance. **b**, Significant hallmark pathways identified in at least one of the traits in ILC3 cells by gene ontology term analysis (hypergeometric overrepresentation test, with *P* values adjusted for multiple testing via the Benjamini–Hochberg method). **c**, Schematic of the MNK-3i CRISPRi screen for detection of genes involved in regulation of IL-22 signaling[76]. **d**, MultiCOGS genes for all six traits visualized among the CRISPRi results, ranked by evidence of positive IL-22 regulation in MNK-3i cells. The gene rankings and gene-level robust rank aggregation analysis (RRA) *P* values were directly obtained from ref. 76. The multiCOGS genes with GSEA adjusted $P < 0.05$ in the screen are labeled in red (GSEA *P* values were adjusted for multiple testing via permutation-derived false discovery rate). multiCOGS genes driving GSEA signal ('leading edge') are labeled in gray. **e**, As in **d**, but for genes ranked by score for negative IL-22 regulation in the MNK-3i screen. Red labels indicate multiCOGS genes significant in the screen at GSEA adjusted $P < 0.05$. As GSEA did not identify a significant enrichment of multiCOGS-prioritized genes among IL-22 repressors, the leading edge genes are not labeled. CD, Crohn's disease; PSC, primary sclerosing cholangitis; UC, ulcerative colitis.

protein previously implicated in limitation of ILC3-driven intestinal inflammation[63].

CLN3 is a lysosomal and endosomal protein with established roles in vesicular trafficking, lysosomal homeostasis and protein turnover[69–71], processes that are central to cytokine storage and secretion. We therefore assessed whether CLN3 overexpression in MNK-3 cells modulated cytokine secretion under basal and inflammatory conditions. As expected, MNK-3 cells constitutively secreted IL-22 and GM-CSF, with further induction of these cytokines upon stimulation, whereas IL-17 production was restricted to stimulated conditions (Fig. 6f, Extended Data Fig. 8g and Supplementary Table 14). CLN3 overexpression, as confirmed by reverse transcription quantitative PCR (RT–qPCR) and immunoblotting (Extended Data Fig. 8e,f), significantly reduced secretion of IL-17, IL-22, and GM-CSF by stimulated MNK-3 cells, with basal IL-22 and GM-CSF secretion also reduced in the absence of stimulation (Fig. 6f and Extended Data Fig. 8g). Viable cell numbers were quantified at the end of cytokine secretion assays and showed no difference under basal conditions, with a modest reduction in *Cln3*-overexpressing cells following stimulation (Extended Data Fig. 8h and Supplementary Table 15).

Collectively, these results highlight CLN3 and the broader *CLN3/APOBR* locus as regulators of ILC3 inflammatory output.

multiCOGS prioritizes candidate effector genes for six immune diseases with potential roles in ILC3 inflammatory function

We extended multiCOGS analysis beyond Crohn's disease to five other immune GWAS datasets with available summary statistics that showed enrichment at active ILC3 and CD4$^+$ T cell PIRs in the RELI analysis, including adult-onset asthma[72], IBD[40], ulcerative colitis[40], primary sclerosing cholangitis[73] and celiac disease[74] (Supplementary Tables 16–18). Across the six traits and two cell types, we detected a total of 332 prioritized disease candidate genes (multiCOGS score > 0.5), of which 251 were prioritized in ILC3 cells (Fig. 7a) and 266 in CD4$^+$ T cells (Supplementary Table 17). As expected from their shared etiology, the three traits relating to Crohn's disease, ulcerative colitis and IBD clustered together with respect to gene-level multiCOGS scores, whereas asthma formed an outgroup (Fig. 7a).

A total of 66 candidate genes were prioritized in ILC3s only and 81 in CD4$^+$ T cells only (Supplementary Table 17). Notable ILC3-specific candidate genes included several cytokines and receptors involved in immune response, such as *CCR2* (celiac disease) and *BCL6* and *IL17A* (asthma), as well as IL-18 receptor gene *IL18R1* (Crohn's disease, celiac disease and asthma). We also noted ILC3-specific prioritization for Crohn's disease and asthma of *BTN3A1* and *BTN3A2*, which encode transmembrane butyrophilins involved in recognition of microbial antigens[75]. Finally, *CLN3* was also prioritized for the broader IBD trait (multiCOGS score 0.538) in addition to Crohn's disease, also selectively in ILC3s.

Pathway analysis of the prioritized genes across all traits revealed shared enrichment for inflammatory gene ontology terms, including cytokine binding and immune receptor activity (Fig. 7b and Supplementary Table 19). To further investigate the roles of these genes in ILC3 inflammatory function, we analyzed a recent CRISPRi screen for regulators of IL-22 expression in IL-23/IL-1β-stimulated MNK-3 cells[76] (Fig. 7c). We found that six multiCOGS-prioritized candidate genes encoded positive regulators of IL-22 production, and five encoded negative regulators (highlighted in red in Fig. 7d,e) (Supplementary Table 20). Genes encoding positive regulators included established IBD-related IL-22 activators *IL23R*, *BATF* and *RORC*, as well as asthma-associated genes *GNA15*, *SESN1* and *GATA3*. Genes encoding negative regulators included *PTPN2*, *NEMF*, *HCLS1*, *PPP5C* and *KXD1*, with previous evidence for IL-22 repression by *PTPN2* through STAT3 dephosphorylation[77]. Overall, multiCOGS-prioritized genes (labeled in Fig. 7d) were significantly enriched among positive regulators of IL-22 production (gene set enrichment analysis $P = 3.64 \times 10^{-2}$; Supplementary Table 21), further implicating dysregulated ILC3 activation as a mechanism linking these genes to immune-related disease risk.

In summary, this analysis expands the compendium of prioritized GWAS gene candidates with potential roles in ILC3s to six immune-related disease traits and demonstrates the potential role of many prioritized genes in ILC3 inflammatory function.

Discussion

Here we report high-resolution promoter interaction profiling in primary human ILC3s, revealing tens of thousands of promoter-anchored contacts, both shared and specific relative to CD4$^+$ T helper cells. We found that ILC3 PIRs were enriched for GWAS variants associated with multiple immune diseases, enabling cell-type-resolved gene prioritization based on these data. Using multiCOGS across six immune-related diseases to prioritize risk genes with potential roles in ILC3s, we produced a compendium of 251 genes, including both known and potentially new candidates. The technical challenges and methodological advances of our GWAS gene prioritization framework and a comparison of complementary approaches are examined in the Supplementary Note.

Integration with a CRISPRi screen for genes affecting ILC3 inflammatory response provided an indication of the potential role of these genes in ILC3 biology. This included 11 prioritized genes that were detected as putative IL-22 activators and repressors in the CRISPRi screen[76]. However, further targeted experiments are needed to gain a deeper understanding of the functional role of the prioritized genes in ILC3 biology and their contribution to immune disease risk.

The *CLN3* gene, prioritized in our analysis for Crohn's disease risk in ILC3s but not in CD4$^+$ T cells, underlies the majority of cases of neurodegenerative disorder Batten disease. Although immune features have been reported in Batten disease and other lysosomal disorders[78,79], the function of CLN3 in the immune system has remained poorly understood. We found that in an ILC3-like mouse cell line, *Cln3* expression was downregulated upon cytokine stimulation of mouse ILC3s, and that *Cln3* overexpression but not knockdown affected stimulation-induced transcriptional programs and cytokine production. These results are supported by the established roles of CLN3 in vesicular trafficking[69–71]. Overall, our findings implicate CLN3 and the broader *CLN3/APOBR* locus in regulation of ILC3 inflammatory function and Crohn's disease risk, raising the possibility that inflammatory processes may contribute to gastrointestinal manifestations observed in *CLN* deficiency[80]. The regulatory complexity of this locus is examined further in the Supplementary Note.

Human ILC3s in our study were derived from tonsillectomy material, but their regulatory elements showed enrichment for variants associated with immunological disorders affecting a broad range of tissues. This is consistent with findings from single-cell genomics suggesting that cell type rather than tissue type is likely to be the driving factor behind variation in chromatin accessibility and gene expression[81,82]. Furthermore, ILC3s from regularly inflamed tonsils have a cytokine profile closer to that of mucosal-resident ILC3 populations than to that of ILC3s from resting lymph nodes or peripheral blood[83]. Focused studies in relevant physiological contexts and disease models are required to further establish the role of ILC3s in mediating the effects of genetic variation. However, such analyses are complicated by the rarity of ILC3s and a lack of robust human cell line models for this cell type, and by the strong influence of organismal and environmental factors on autoimmune disease pathogenesis, which are difficult to reproduce in a laboratory setting either in vitro or in vivo.

In conclusion, our work presents improved methods for profiling and detection of promoter-anchored chromosomal interactions and use of these data to interpret GWAS signals. We used this framework to generate a comprehensive catalog of ILC3 promoter contacts and detect candidate effector genes implicating these cells in immune-related disease risk, providing initial functional validation

of the prioritized candidates. These findings advance our understanding of ILC3 biology and the contributions of this rare cell type to disease, as well as demonstrating the broader utility of our approach for study of regulatory architecture and GWAS interpretation in other rare cell populations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02681-0.

References

1. Sonnenberg, G. F. & Artis, D. Innate lymphoid cells in the initiation, regulation and resolution of inflammation. *Nat. Med.* **21**, 698–708 (2015).
2. Eberl, G., Colonna, M., Di Santo, J. P. & McKenzie, A. N. J. Innate lymphoid cells: a new paradigm in immunology. *Science* **348**, aaa6566 (2015).
3. Spits, H. et al. Innate lymphoid cells — a proposal for uniform nomenclature. *Nat. Rev. Immunol.* **13**, 145–149 (2013).
4. Ebbo, M., Crinier, A., Vély, F. & Vivier, E. Innate lymphoid cells: major players in inflammatory diseases. *Nat. Rev. Immunol.* **17**, 665–678 (2017).
5. Hepworth, M. R. et al. Immune tolerance. Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4$^+$ T cells. *Science* **348**, 1031–1035 (2015).
6. Zeng, B. et al. ILC3 function as a double-edged sword in inflammatory bowel diseases. *Cell Death Dis.* **10**, 315 (2019).
7. Mazzurana, L. et al. Crohn's disease is associated with activation of circulating innate lymphoid cells. *Inflamm. Bowel Dis.* **27**, 1128–1138 (2021).
8. Saez, A. et al. Innate lymphoid cells in intestinal homeostasis and inflammatory bowel disease. *Int. J. Mol. Sci.* **22**, 7618 (2021).
9. Liu, Z. et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat. Genet.* **55**, 796–806 (2023).
10. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
11. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
12. Ho, J. S. Y. et al. TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal inflammation. *Cell* **184**, 2618–2632 (2021).
13. Malysheva, V. et al. Novel method. World Patent WO 2021/064430 (2021).
14. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
15. Malysheva, V., Ray-Jones, H., Lakes, N. & Spivakov, M. ILC3 PCHiC large supplementary data. *OSF* https://doi.org/10.17605/OSF.IO/AQ9FB (2025).
16. Freire-Pritchett, P. et al. Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat. Protoc.* **16**, 4144–4176 (2021).
17. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
18. Koues, O. I. et al. Distinct gene regulatory pathways for human innate versus adaptive lymphoid cells. *Cell* **165**, 1134–1146 (2016).
19. Martin, F. J. et al. Ensembl 2023. *Nucleic Acid Res.* **51**, D933–D941 (2023).
20. Adams, D. et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
21. Cella, M. et al. Subsets of ILC3-ILC1-like cells generate a diversity spectrum of innate lymphoid cells in human mucosal tissues. *Nat. Immunol.* **20**, 980–991 (2019).
22. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
23. Jan-Abu, S. C., Kabil, A. & McNagny, K. M. Parallel origins and functions of T cells and ILCs. *Clin. Exp. Immunol.* **213**, 76–86 (2023).
24. Robinette, M. L. & Colonna, M. Immune modules shared by innate lymphoid cells and T cells. *J. Allergy Clin. Immunol.* **138**, 1243–1251 (2016).
25. von Burg, N. et al. Activated group 3 innate lymphoid cells promote T-cell-mediated immune responses. *Proc. Natl Acad. Sci. USA* **111**, 12835–12840 (2014).
26. Schroeder, J.-H., Howard, J. K. & Lord, G. M. Transcription factor-driven regulation of ILC1 and ILC3. *Trends Immunol.* **43**, 564–579 (2022).
27. Cairns, J., Orchard, W. R., Malysheva, V. & Spivakov, M. Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics* **35**, 4764–4766 (2019).
28. Harley, J. B. et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
29. Cerezo, M. et al. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Res.* **53**, D998–D1005 (2024).
30. Allez, M. et al. CD4$^+$NKG2D$^+$ T cells in Crohn's disease mediate inflammatory and cytotoxic responses through MICA interactions. *Gastroenterology* **132**, 2346–2358 (2007).
31. Pariente, B. et al. Activation of the receptor NKG2D leads to production of Th17 cytokines in CD4$^+$ T cells of patients with Crohn's disease. *Gastroenterology* **141**, 217–226 (2011).
32. Rabe, H. et al. Distinct patterns of naive, activated and memory T and B cells in blood of patients with ulcerative colitis or Crohn's disease. *Clin. Exp. Immunol.* **197**, 111–129 (2019).
33. Ma, C. et al. Critical role of CD6^{high}CD4$^+$ T cells in driving Th1/Th17 cell immune responses and mucosal inflammation in IBD. *J. Crohns Colitis* **13**, 510–524 (2019).
34. Pearson, C. et al. ILC3 GM-CSF production and mobilisation orchestrate acute intestinal inflammation. *eLife* **5**, e10066 (2016).
35. Burren, O. S. et al. Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* **18**, 165 (2017).
36. Rüeger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* **14**, e1007371 (2018).
37. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B* **82**, 1273–1300 (2020).
38. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. *PLoS Genet.* **18**, e1010299 (2022).
39. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).
40. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
41. Sun, R., Hedl, M. & Abraham, C. TNFSF15 promotes antimicrobial pathways in human macrophages and these are modulated by *TNFSF15* disease-risk variants. *Cell. Mol. Gastroenterol. Hepatol.* **11**, 249–272 (2021).

42. Lahiri, A., Hedl, M., Yan, J. & Abraham, C. Human LACC1 increases innate receptor-induced responses and a *LACC1* disease-risk variant modulates these outcomes. *Nat. Commun.* **8**, 15614 (2017).
43. Naser, S. A. et al. Role of *ATG16L*, *NOD2* and *IL23R* in Crohn's disease pathogenesis. *World J. Gastroenterol.* **18**, 412–424 (2012).
44. Richard, A. C. et al. Reduced monocyte and macrophage *TNFSF15*/TL1A expression is associated with susceptibility to inflammatory bowel disease. *PLoS Genet.* **14**, e1007458 (2018).
45. Assadi, G. et al. *LACC1* polymorphisms in inflammatory bowel disease and juvenile idiopathic arthritis. *Genes Immun.* **17**, 261–264 (2016).
46. Ueta, M. et al. Mucocutaneous inflammation in the Ikaros family zinc finger 1-keratin 5-specific transgenic mice. *Allergy* **73**, 395–404 (2018).
47. Chang, K. et al. Multi-omics profiles refine L-dopa decarboxylase (DDC) as a reliable biomarker for prognosis and immune microenvironment of clear cell renal cell carcinoma. *Front. Oncol.* **12**, 1079446 (2022).
48. Ochoa, D. et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
49. Khan, F., Radovanovic, A., Gojobori, T. & Kaur, M. IBDDB: a manually curated and text-mining-enhanced database of genes involved in inflammatory bowel disease. *Database* **2021**, baab022 (2021).
50. Ntunzwenimana, J. C. et al. Functional screen of inflammatory bowel disease genes reveals key epithelial functions. *Genome Med.* **13**, 181 (2021).
51. Piñero, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2016).
52. Sazonovs, A. et al. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. *Nat. Genet.* **54**, 1275–1283 (2022).
53. Schmitz, I. Gadd45 proteins in immunity. *Adv. Exp. Med. Biol.* **793**, 51–68 (2013).
54. Erdenebayar, N., Maekawa, Y., Nishida, J., Kitamura, A. & Yasutomo, K. Protein-tyrosine phosphatase-kappa regulates $CD4^+$ T cell development through ERK1/2-mediated signaling. *Biochem. Biophys. Res. Commun.* **390**, 489–493 (2009).
55. Mole, S. E. & Cotman, S. L. Genetics of the neuronal ceroid lipofuscinoses (Batten disease). *Biochim. Biophys. Acta* **1852**, 2237–2241 (2015).
56. Hildner, K., Punkenburg, E., Abendroth, B. & Neurath, M. F. Immunopathogenesis of IBD: Batf as a key driver of disease activity. *Dig. Dis.* **34**, 40–47 (2016).
57. Griffin, M. J. et al. Early B-cell factor-1 (EBF1) is a key regulator of metabolic and inflammatory signaling pathways in mature adipocytes. *J. Biol. Chem.* **288**, 35925–35939 (2013).
58. Bianchi, E., Rogge, L. & Vecellio, M. Role of the IL-23/IL-17 pathway in chronic immune-mediated inflammatory diseases: mechanisms and targeted therapies. *Front. Immunol.* **12**, 770275 (2021).
59. Atreya, I., Atreya, R. & Neurath, M. F. NF-κB in inflammatory bowel disease. *J. Intern. Med.* **263**, 591–596 (2008).
60. Machado, M. S. G. et al. IL-1 receptor contributes to the maintenance of the intestinal barrier via IL-22 during obesity and metabolic syndrome in experimental model. *Microorganisms* **12**, 1717 (2024).
61. Allan, D. S. J. et al. An in vitro model of innate lymphoid cell function and differentiation. *Mucosal Immunol.* **8**, 340–351 (2015).
62. Longman, R. S. et al. CX_3CR1^+ mononuclear phagocytes support colitis-associated innate lymphoid cell production of IL-22. *J. Exp. Med.* **211**, 1571–1583 (2014).
63. Zhao, X. et al. Elevated nitric oxide during colitis restrains GM-CSF production in ILC3 cells via suppressing an AhR-Cyp4f13-NF-κB axis. *Nat. Commun.* **16**, 5654 (2025).
64. Castellanos, J. G. et al. Microbiota-induced TNF-like ligand 1A drives group 3 innate lymphoid cell-mediated barrier protection and intestinal T cell activation during colitis. *Immunity* **49**, 1077–1089 (2018).
65. Lepik, K. et al. C-reactive protein upregulates the whole blood expression of *CD59* - an integrative analysis. *PLoS Comput. Biol.* **13**, e1005766 (2017).
66. GTEx Consortium The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
67. Quach, H. et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656 (2016).
68. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414 (2016).
69. Lacy, P. & Stow, J. L. Cytokine release from innate immune cells: association with diverse membrane trafficking pathways. *Blood* **118**, 9–18 (2011).
70. Calcagni, A. et al. Loss of the batten disease protein CLN3 leads to mis-trafficking of M6PR and defective autophagic-lysosomal reformation. *Nat. Commun.* **14**, 3911 (2023).
71. Cotman, S. L. & Lefrancois, S. CLN3, at the crossroads of endocytic trafficking. *Neurosci. Lett.* **762**, 136117 (2021).
72. Ferreira, M. A. R. et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am. J. Hum. Genet.* **104**, 665–684 (2019).
73. Ji, S.-G. et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.* **49**, 269–273 (2017).
74. Dubois, P. C. A. et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
75. Arnett, H. A. & Viney, J. L. Immune modulation by butyrophilins. *Nat. Rev. Immunol.* **14**, 559–569 (2014).
76. Brown, R. A. et al. CRISPRi screening identifies SON and MAP4K1 as regulators of type III cytokine expression in innate lymphoid cells. *J. Immunol.* **215**, vkag110 (2026).
77. Zhang, Y., Ding, H., Wang, X. & Ye, S.-D. Modulation of STAT3 phosphorylation by PTPN2 inhibits naïve pluripotency of embryonic stem cells. *FEBS Lett.* **592**, 2227–2237 (2018).
78. Castaneda, J. A., Lim, M. J., Cooper, J. D. & Pearce, D. A. Immune system irregularities in lysosomal storage disorders. *Acta Neuropathologica* **115**, 159–174 (2007).
79. Hersrud, S. L., Kovács, A. D. & Pearce, D. A. Antigen presenting cell abnormalities in the *Cln3*$^{-/-}$ mouse model of juvenile neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* **1862**, 1324–1336 (2016).
80. Ziółkowska, E. A. et al. Gene therapy ameliorates bowel dysmotility and enteric neuron degeneration and extends survival in lysosomal storage disorder mouse models. *Sci. Transl. Med.* **17**, eadj1445 (2025).
81. Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001 (2021).
82. 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).
83. Bar-Ephraim, Y. E. et al. Cross-tissue transcriptomic analysis of human secondary lymphoid organ-residing ILC3s reveals a quiescent state in the absence of inflammation. *Cell Rep.* **21**, 823–833 (2017).
84. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13**, 577–580 (2016).

85. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
86. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
87. Wakefield, J. Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
88. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

[1]MRC Laboratory of Medical Sciences, London, UK. [2]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK. [3]VIB, Center for Molecular Neurology, Antwerp, Belgium. [4]Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Antwerp, Belgium. [5]VIB, Center for AI and Computational Biology, Leuven, Belgium. [6]Trinity Hall, University of Cambridge, Cambridge, UK. [7]Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [8]Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [9]Medical Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [10]Division of Human Genetics and Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [11]Department of Microbial Infection and Immunity, The Ohio State University, Columbus, OH, USA. [12]Pelotonia Institute for Immuno-Oncology, The Ohio State University, Columbus, OH, USA. [13]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [14]Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [15]Center for Genetics and Rare Diseases, Sanford Research, Sioux Falls, SD, USA. [16]University of Cambridge, Cambridge, UK. [17]Department of Haematology, University of Cambridge, Cambridge, UK. [18]National Health Service (NHS) Blood and Transplant, Cambridge, UK. [19]Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter Medical School, Exeter, UK. [20]The Babraham Institute, Cambridge, UK. [21]Department of Biological Sciences, Florida State University, Tallahassee, FL, USA. [22]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [23]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [24]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [25]Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [26]Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [27]Divisions of Pediatric Otolaryngology and Pulmonary & Sleep Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [28]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [29]MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Cambridge, UK. [30]Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK. [31]Present address: Eli Lilly and Company, Indianapolis, IN, USA. [32]Present address: Section of Pediatric Rheumatology, Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC, USA. [33]Present address: Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK. [34]Present address: NeoGenomics, Cambridge, UK. [35]Present address: British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [36]Present address: Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. [37]Present address: European Bioinformatics Institute, Hinxton, UK. [38]Present address: Cancer Research UK Cambridge Research Institute, Cambridge, UK. [39]Present address: University of Kent, Canterbury, UK. [40]Present address: Laboratory for Nuclear Dynamics, Institute for Protein Research, Osaka University, Osaka, Japan. [41]Present address: Institute of Medical Science, University of Tokyo, Tokyo, Japan. [42]These authors contributed equally: Valeriya Malysheva, Helen Ray-Jones, Nora Lakes. [43]These authors jointly supervised this work: Chris Wallace, Emily R. Miraldi, Stephen N. Waggoner, Mikhail Spivakov. ✉e-mail: valeriya.malysheva@vib.be; stephen.waggoner@cchmc.org; mikhail.spivakov@lms.mrc.ac.uk

Methods

Human ILC3 isolation

Three children requiring tonsillectomy (two male, one female) were recruited to a prospective study at a tertiary academic care center through the division of Pediatric Otolaryngology-Head and Neck Surgery at Cincinnati Children's Hospital Medical Center. Consent was obtained from parents in the perioperative suite on the day of the procedure with approval of the institutional review board: 'Tissues for the study of functional genomics and immunity' (no. 2018-3069) at Cincinnati Children's Hospital Medical Center. Criteria for enrollment in the study included a history of sleep-disordered breathing or recurrent or chronic tonsillitis requiring removal of the tonsillar tissue. Individuals were excluded from the study if the tonsillar tissue was acutely infected or if anatomic abnormalities were present that required a more detailed pathologic evaluation after the surgical procedure. Samples were labeled with a deidentified barcode and transferred to the research team for further processing.

Next, tonsils were dissociated into a single-cell suspension as previously described[89,90]. Briefly, human tonsil tissue was processed by mincing with scissors, followed by transfer of up to 4 g of tissue to a gentleMACS C tube (Miltenyi Biotec) containing 8 ml of phosphate-buffered saline (PBS) with 0.5 mg ml^{-1} collagenase D and 3,000 U ml^{-1} DNase I, and then dissociated on a GentleMACS Octo Dissociator (Miltenyi Biotec) using 'program C (Spleen program 2 followed by spleen program 3).' Tissue homogenates were incubated in a 37 °C water bath for 15 min and then dissociated again using 'program C' and transferred through a 100-µm cell strainer into 20 ml RPMI containing 10% human AB serum (Sigma-Aldrich). Next, the cell suspension was overlaid on 10 ml of Ficoll-Paque PLUS (GE Healthcare) and subjected to density-gradient separation via centrifugation for 20 min at 700*g*, 20 °C, slow acceleration and no brake. Leukocytes were collected from the interphase layer and washed with 50 ml of PBS for 6 min at 560*g*, 20 °C.

Single-cell suspensions of tonsil mononuclear cells were subjected to positive selection with anti-human-CD3, anti-human-CD19 and anti-human-CD14 (Miltenyi Biotec) and transferred through LD columns (Miltenyi Biotec) according to the manufacturer's guidelines (Supplementary Fig. 1). The depleted cell suspension flowthrough was collected into a 15-ml conical tube and then centrifuged for 5 min at 300*g*, 20 °C. Subsequently, cells were labeled with LIVE/DEAD Fixable Near-IR dead cell stain kit (Invitrogen). Next, cells were labeled with a sorting antibody cocktail that contained negative-linage (Lin^-) CD19-Brilliant Violet (BV)421 (HIB19), CD14-BV421 (63D3) and CD3-BV421 (OKT3), and the following antibodies: CD45-FITC (HI30), CD94-PerCP-Cy5.5 (DX22), CD127-PE-Cy7 (A019D5), cKit-BV510 (104D2) and NKp44-Alexa Fluor (AF)647 (P44-8), all purchased from BioLegend; and CRTH2-PE (301109) from R&D. ILC3 cells were sorted based on $CD45^+Lin^-CD127^+CD94^-CRTH2^-cKit^+NKp44^+$ expression, similar to the method used by Bar-Ephraim et al.[83]. Cells were sorted using a FACSAria II (BD Biosciences). Sorted ILC3 cells were washed with PBS for 5 min at 300*g*, 20 °C, then incubated in 100 µl of 2% formaldehyde (in PBS) for 10 min, followed by addition of 0.125 M glycine. Next, cells were centrifuged at 400*g* for 5 min at 4 °C, resuspended with cold PBS and centrifuged again at 400*g* for 5 min at 4 °C; then, the supernatant was discarded, and cells were snap-frozen in liquid nitrogen and stored at −80 °C before PCHi-C analysis.

Human $CD4^+$ T cell isolation

Samples were obtained from two male donors after written informed consent had been obtained under studies 'A Blueprint of Blood Cells' (REC reference 12/EE/0040) and 'Genes and mechanisms in type 1 diabetes in the Cambridge BioResource' (REC reference 05/Q0106/20), both of which were approved by the NRES Committee East of England – Cambridgeshire and Hertfordshire. Total $CD4^+$ lymphocytes were obtained from PBMCs from venous blood by negative selection using an EasySep Human $CD4^+$ T Cell Enrichment kit (catalog no. 19052) from STEMCELL Technologies. Purified $CD4^+$ T cells were washed with PBS for 5 min at 300*g*, 20 °C, then incubated in 100 µl of 2% formaldehyde (in PBS) for 10 min, followed by addition of 0.125 M glycine. Next, cells were centrifuged at 400*g* for 5 min at 4 °C, resuspended with cold PBS and centrifuged again at 400*g* for 5 min at 4 °C; then, the supernatant was discarded, and cells were snap-frozen in liquid nitrogen and stored at −80 °C before PCHi-C analysis. Two replicates of 1 million and two more replicates of 50,000 cells were used to generate PCHi-C datasets.

Promoter capture Hi-C

PCHi-C was performed as previously described[12]. Cells were lysed in a lysis buffer (for 30 min on ice) and digested with *DpnII* (NEB) overnight at 37 °C with rotation (200*g*). Restriction overhangs were filled in with Klenow (NEB) using biotin-14-dATP (Jena Bioscience), and ligation was performed in ligation buffer for 4 h at 16 °C (T4 DNA ligase; Life Technologies). After overnight decrosslinking at 65 °C, the ligated DNA was tagmented to produce fragments of 300–700 bp. Ligation products were isolated using MyOne C1 streptavidin beads (Life Technologies), followed by washing with Wash&Binding buffer and nuclease-free water. Isolated Hi-C ligation products on the beads were then used directly for PCR amplification, and the final Hi-C library was purified with AMPure XP beads (Beckman Coulter). PCHi-C was performed using a custom-designed Agilent SureSelect system following the manufacturer's protocol. The PCHi-C libraries were paired-end sequenced (100 bp) on an Illumina HiSeq 2500 machine at a sequencing depth of ~400 million reads per sample (Supplementary Table 1).

PCHi-C data preprocessing and detection of significant interactions

Sequencing data from three ILC3 PCHi-C biological replicates were aligned to the hg38 genome assembly using Bowtie2 (ref. 91) and quality-controlled using HiCUP[92]. Quality metrics for all generated PCHi-C datasets are reported in Supplementary Table 1. Significant interactions were then detected across replicates using CHiCAGO[14] as previously described[16] at single *DpnII* fragment resolution and in bins of fragments approximately 5 kb in length, with the baited promoter fragments left unbinned. Design files for CHiCAGO analysis are available as Data S12 at OSF[15]. Leaving the baited *DpnII* fragment unbinned meant that nearly every baited fragment was occupied by a single protein-coding gene promoter. By contrast, one-third (33%) of baited fragments in the *HindIII*-based Capture Hi-C design (with a median fragment size of 4 kb) contained two or more promoters. Therefore, leaving the baited fragment unbinned substantially improved the resolution and interpretability of analyses such as (multi)COGS.

For CHiCAGO analysis at single-fragment resolution, *P* value weights were estimated following our previously described procedure[16] (weights rounded to 2 decimal digits: weightAlpha = 24.67, weightBeta = −2.29, weightGamma = −16.05, weightDelta = −6.99); CHiCAGO default *P* value weights were used for the 5-kb analysis. A score cutoff of ≥5 was used for both resolutions. A consensus list of promoter interactions was compiled from nonredundant contacts detected at fragment and 5-kb resolutions, with fragment-level contacts taking priority over 5-kb contacts in cases of overlap.

For comparison of *DpnII*-based ILC3 PCHi-C data with *HindIII*-based data from Javierre et al.[17] reported on GRCh37, the ILC3 data were realigned with HiCUP using the GRCh37 assembly and *HindIII* resolution and reanalyzed with CHiCAGO using the *HindIII* PCHi-C design from the same study.

Activity-by-captured-contact

For a given gene–enhancer pair, the ABC score[22,93] is the normalized product of enhancer activity (proxied by levels of chromatin accessibility and relevant histone modifications) and contact (proxied by

three-dimensional contact frequency detected from a chromosome conformation capture assay):

$$\text{ABC score}_{E,G} = \frac{A_E \times C_{E,G}}{\sum_{e \text{ within 5 Mb of } G} A_e \times C_{e,G}},$$

where E and G are an enhancer and gene of interest, respectively; e denotes enhancers in the gene's 5 Mb neighborhood; A is enhancer activity and C is enhancer–promoter contact frequency. To adapt ABC for PCHi-C data, we took advantage of the CHiCAGO normalization algorithm and developed an imputation procedure in the normalized counts space based on the inferred decay of interaction read counts with distance. As we did not expect the frequency of enhancer–promoter contacts to fall below levels expected owing to Brownian collision, for a given pair of fragments involving a baited promoter, we selected the maximum between the CHiCAGO-normalized observed read counts (N_{obs}) and expected read counts N_{exp} estimated as:

$$N_{exp} = B_{mean}/(s_i \times s_j),$$

where B_{mean} is the CHiCAGO-estimated Brownian noise level, and s_i and s_j are the bait- and other end-specific scaling factors. For promoters that could not be baited in the Capture Hi-C design and those that were filtered out owing to failing quality control, we estimated the expected normalized read count directly from the interaction distance d, using the distance function f(d) fitted by CHiCAGO. Given the very large values of f(d) for very-short-range interactions (<1.5 kb), in the imputation procedure we capped them at the value of f(d) for d = 1.5 kb (median fragment length) for candidate enhancers that were less than 1 fragment away from the bait. Please refer to Additional File 1 in the publication presenting the CHiCAGO pipeline[14] for the formal definition of these parameters and their estimation procedures.

The imputed normalized read counts were used as contact data in the ABC pipeline, and public H3K27ac data and ATAC using sequencing (ATAC-seq) data for ILC3s, processed as described above, were used to compute activity. To validate the ABCC approach, we used high-throughput CRISPRi-FlowFISH data from Fulco et al.[22], which showed the impact of perturbing ~3,500 enhancer elements on the expression of 30 genes in K562 cells. As PCHi-C data for K562 cells were not available, we used our previously published PCHi-C dataset in a related primary cell type, erythroblasts[17], to generate ABC scores based on these data and the ATAC-seq and H3K27ac chromatin immunoprecipitation followed by sequencing (ChIP–seq) datasets for K562 cells from Fulco et al. In comparison with the original ABC scores from Fulco et al., which were based on pooling of conventional Hi-C data from multiple cell types, our approach showed higher precision (69.1% versus 58.3%) at the same level of recall (58.3%) of CRISPRi-FlowFISH-validated enhancer–promoter pairs (Extended Data Fig. 2). To select the ABCC score cutoff, we optimized the Pearson correlation between the per-gene ABCC numerator and gene expression ($R_{ABC\text{-}GE}$), using an approach inspired by Xu et al.[94] We used a single ABCC score cutoff of 0.023 for all analyzed cell types, as this was close to both the maximum $R_{ABC\text{-}GE}$ in each cell type and the cutoff of 0.02 that yielded an optimal precision–recall of CRISPRi-FlowFISH-validated enhancer–promoter pairs in K562 cells.

RNA-seq data analysis

Human ILC3 RNA-seq data[21] were downloaded from the Gene Expression Omnibus (GEO; accession GSE130775). Salmon[95] was used to map reads to transcripts on the hg38 and mm10 genome assemblies for human and mouse RNA-seq data, respectively. The transcript counts were imported into DESeq2 objects and collapsed to gene counts using Tximport[96]. Differential analysis was performed using DESeq2 (ref. 88). CRISPRi and CRISPRa data were analyzed separately, with four conditions (baseline versus stimulated; CRISPR-targeted versus scrambled control) included in each analysis and specific comparisons queried using contrasts. Low-count genes were prefiltered before the analysis was run. The following parameters were used to report significantly differentially expressed genes: alpha = 0.05 and adjusted $P < 0.05$. For the TL1A stimulation experiment, gene count matrices from ref. 64 were downloaded from GEO (accession GSE120723) and reanalyzed using DESeq2.

RELI analysis

RELI determines significantly enriched overlaps between selected genomic loci (here PIRs intersecting open chromatin or H3K27ac signals in ILC3s based on public data) and trait-associated genetic variants. This is done by comparing the observed overlaps with a null distribution of artificially created variant sets with LD characteristics similar to those of the trait-associated variants[28]. A practical advantage of RELI over the commonly used stratified LD score regression[97] is that it does not require summary statistics data and can be performed on sets of significant single-nucleotide polymorphisms (SNPs) reported in the GWAS Catalog[29].

RELI[28] (v.0.1.1a) was used to find enrichment of genetic variants in PIRs that were accessible and/or marked with activating epigenetic markers (H3K27ac and ATAC-seq). In brief, RELI tests genomic features such as ATAC-seq or ChIP–seq peaks or PCHi-C PIRs for statistically significant overlaps with known disease risk variants identified in GWAS. Risk variants are expanded to LD blocks with variants that have r^2 value ≥ 0.8. LD blocks are then intersected with the genetic feature BED files. A null distribution is generated using randomly shuffled LD blocks (n = 1,000) and intersection with the feature files. A P value is generated by comparing the observed number of intersections in the test to the null distribution.

Promoter-interacting regulatory elements were determined as input for RELI as follows. The PIR sets were the union of PCHi-C interactions (CHiCAGO score ≥ 5, binned to 5-kb or *DpnII* fragment-level resolution) and ABC enhancers, excluding any *trans*-chromosomal interactions. Regulatory elements were then defined as the union of peaks of open chromatin and H3K27ac in ILC3 and CD4$^+$ T cells (using ATAC-seq and ChIP–seq data as above). The true intersection between these regulatory elements and PIRs in each cell type was then determined using pybedtools intersect. The coordinates for these regions were lifted over from hg38 to hg19 using UCSC liftOver (v.377), then sorted and merged for use with RELI. RELI[28] was run against 763 traits of individuals of European ancestry in the GWAS Catalog v.1.0.2. The traits were filtered to those with ≥10 independent risk loci, resulting in a final set of 495 traits in the results. Results with Bonferroni-adjusted $P < 0.05$ were defined as significant. For depiction of RELI results, we labeled only significant traits with total number of risk loci per trait greater than 10.

Summary-level GWAS imputation and multivariate fine-mapping

We downloaded LD block data for EUR from lddetect[98] (https://bitbucket.org/nygcresearch/ldetect-data/src/master/) and converted it to hg38 using liftOver[99]. We generated LD matrices based on samples from individuals of European ancestry from phased 1000 Genomes Phase 3 data[100] downloaded from https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html.

We first used the LD matrices to impute the summary statistics data within blocks using a published method[36]. In brief, given a vector of observed z-scores, an estimate of the correlation between observed variations S, and an estimate of the correlation between an unobserved variant u and observed markers s, we imputed the z-score at the unobserved variant as:

$$z_u = s^T S^{-1} z.$$

To ensure S was invertible, we pruned observed variants to ensure there were no pairs of variants in high LD ($r^2 > 0.8$).

For blocks with appreciable association signals (minimum $P < 10^{-6}$), we used the susieR package[37,38] to fine-map the data. We defined 'detected signals' as those for which SuSiE could calculate a 95% credible set and used the posterior inclusion probabilities (PIP) for each SNP for each signal thus detected as input for multiCOGS. For the remaining blocks, or for cases in which susieR failed to find any signals meeting our criteria, we fine-mapped using the single causal variable approach, as previously described[17,35], and used the posterior probabilities of association as input for multiCOGS.

multiCOGS

We modified the standard COGS algorithm[17,35] to account for the inclusion of multiple association signals in a region (multiCOGS). Whereas in standard COGS, fragment-level scores are calculated by summing variant-level posterior inclusion probabilities within a given fragment and LD block, multiCOGS considers each credible set within each LD block and forms an overall gene score as the probability that at least one of the multiple fine-mapped signals is linked, through PCHi-C, to the gene of interest:

$$\text{multiCOGS score}_{\text{gene}} = 1 - \prod(1 - \text{score}_{\text{gene, LD block, credible set}}).$$

As input for multiCOGS, we used imputed and fine-mapped summary-level GWAS data alongside CHiCAGO- and ABCC-processed PCHi-C results (thresholded at CHiCAGO score ≥ 5 and ABCC scores ≥ 0.023). In addition, for each transcription start site annotated in Ensembl GRCh38 release 88 (Havana and Ensembl/Havana merge), regardless of whether it was included in the PCHi-C capture design, we defined virtual promoter regions as windows of ±5 *DpnII* fragments surrounding the fragment containing the transcription start site. Finally, we provided the locations of protein-coding SNPs identified using VEP[101] v.99.2 (https://github.com/Ensembl/ensembl-vep). The results for each protein-coding gene were linked across datasets using Ensembl gene IDs as primary identifiers. We removed the major histocompatibility complex locus (GRCh38 6:28510120-33480577) before running multiCOGS.

To determine the contributions of the four categories of genomic loci underlying the prioritized genes (PCHi-C PIRs, ABCC enhancers, virtual promoter regions and coding SNPs), we also ran multiCOGS on each category separately by specifying the feature.names argument in the compute_cogs function.

Cell culture

Mouse MNK-3 cells[61] (Sigma-Aldrich, SCC628) and derived lines were cultured in DMEM with glucose/pyruvate/L-glutamine supplemented with 10% fetal bovine serum, 1× penicillin–streptomycin, 10 ng ml^{-1} mouse recombinant IL-2 and IL-7 (R&D Systems), and 50 µM 2-mercaptoethanol. Media for CRISPRi MNK-3 (MNK-3i) cells contained 10 µg ml^{-1} blasticidin S, and media for CRISPRa MNK-3 (MNK-3a) cells contained 10 µg ml^{-1} blasticidin S and 1,250 µg ml^{-1} hygromycin B. MNK-3i/a cells with single-guide RNA (sgRNA) also received 2 µg ml^{-1} puromycin. MNK-3 activation was induced with 10 ng ml^{-1} IL-1β and 10 ng ml^{-1} IL-23 (R&D Systems).

CRISPRa and CRISPRi

MNK-3i cells were generated as previously described[102] from parental MNK-3 cells (Sigma-Aldrich, SCC628). In brief, MNK-3 cells were transduced with lentivirus containing pLenti CMV rtTA3 Blast (Addgene 26429), selected with blasticidin S, and then infected with TRE3G-dCas9-KRAB-P2A-mCherry lentivirus[102]. Following incubation with doxycycline, mCherry-positive cells were subcloned, and western blot analysis was performed to confirm robust expression of doxycycline-inducible dCas9–KRAB. MNK-3a cells were lentivirally engineered from MNK-3 to constitutively express the dCas9–VP64 fusion gene (Addgene 61425) and the MS2-p65–HSF1 transactivator complex (Addgene 89308), selected with blasticidin S and hygromycin B, and subcloned. All cells were tested for mycoplasma.

Sequences for Cln3-targeting and scrambled gRNAs were based on published sgRNA libraries for MNK-3i[103] and MNK-3a[104] cells and are listed in Supplementary Table 22 alongside RT–qPCR primer sequences. sgRNA sequences and their reverse complements were synthesized by Sigma, annealed, and cloned into a lenti sgRNA(MS2)_puro optimized backbone (Addgene 73797) for MNK-3a or sgOpti (Addgene 85681) for MNK-3i cells using Esp3I digestion as previously described[105]. sgRNA plasmid integration was confirmed by Sanger sequencing (Ohio State Comprehensive Cancer Center Genomics Core). Lentiviral plasmids pMD2.G (Addgene 12259) and psPAX2 (Addgene 12260) were transfected with the sgRNA plasmid into HEK293T cells using Mirus TransIT-293T transfection reagent. Lentivirus media were harvested and filtered 48–72 h posttransfection. Puromycin selection began 36 h after lentiviral guide transduction into MNK-3i/a cells in the presence of polybrene. Bulk transduced populations were used for experiments and maintained in selection antibiotics. RT–qPCR was performed to confirm repression (in MNK-3i lines after 48 h doxycycline incubation) or overexpression (in MNK-3a cells) of target genes relative to *Actb* and the respective scramble control (TRIzol RNA isolation; Verso cDNA synthesis).

To induce CRISPRi guide expression, MNK-3i cells stably expressing Cln3-targeting and scrambled gRNAs were incubated with 2 µg ml^{-1} doxycycline for 48 h. To confirm stimulation, cells were harvested 21 h after cytokine stimulation and stained for intracellular IL-17F and IL-22 (eBioscience IL-22 clone 1H8PWSR and IL-17F clone eBio18F10; BD Life Sciences Cytofix/Cytoperm kit). Expression of IL-17F and IL-22 was assessed on a FACSymphony (BD Life Sciences) and compared against the respective scrambled control.

Design and generation of in vitro-transcribed mRNA

The protein-coding sequence of mouse *Cln3* was based on the longest annotated transcript (NM_001146311.3/ENSMUST00000084589.11). A Myc epitope tag was inserted near the amino terminus, between amino acid residues 3 and 4, within a predicted disordered and cytoplasmic region of the protein. The resulting coding sequence was synthesized and used for in vitro transcription by ApexBio.

In vitro-transcribed mRNA was generated with a cap 1 structure and incorporated N_1-methylpseudouridine. Transcripts contained a poly(A) tail and were supplied in RNase-free sodium citrate buffer (pH 6.4) at a concentration of 1 mg ml^{-1}. Control mRNA encoding GFP was generated using the same chemistry.

mRNA electroporation and cytokine stimulation

MNK-3 cells were electroporated with mRNA using an ATx electroporation system (MaxCyte); 1.0×10^7 cells were electroporated in a 100-µl reaction containing 20 µg of GFP or *Myc*-tagged *Cln3* mRNA (2 µg per 10^6 cells) using the 'Optimization 8' program. Following electroporation, cells were rested for 15 min at 37 °C and then incubated for 15 min at 37 °C in prewarmed medium supplemented with 10 µg ml^{-1} DNase I (Thermo Fisher Scientific), 5 mM $MgCl_2$ and 1 mM $CaCl_2$ before being transferred to complete MNK-3 culture medium.

At 24 h postelectroporation, cells were seeded at 3.0×10^5 cells per well in 24-well plates. Transfected cells were cultured for an additional 24 h in the presence or absence of recombinant mouse 10 ng ml^{-1} IL-1β and 10 ng ml^{-1} IL-23 (R&D Systems). At 48 h postelectroporation, supernatants were collected, clarified by centrifugation and stored at −20 °C. Viable cell numbers were determined by trypan blue exclusion.

Enzyme-linked immunosorbent assay

Cytokines in cell culture supernatants were quantified by enzyme-linked immunosorbent assay using DuoSet kits for mouse IL-17, IL-22

and GM-CSF (R&D Systems) according to the manufacturer's instructions. When necessary, samples were diluted to fall within the dynamic range of the standard curve. Absorbance was measured at 450 nm with wavelength correction at 560 nm using a GloMax Discover microplate reader (Promega). Cytokine concentrations were determined by interpolation from standard curves using a four-parameter logistic fit.

Data were analyzed using GraphPad Prism. Statistical significance was assessed using unpaired Welch's *t*-tests (single experiment) or linear mixed-effects models with genotype as a fixed effect and experiment as a random effect (multiple experiments).

Immunoprecipitation and immunoblotting

MNK-3 cells were electroporated with GFP or Myc-tagged *Cln3* mRNA as described above and harvested 24 h later. Cells were lysed in a non-denaturing buffer containing 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% *n*-dodecyl-β-D-maltoside, 10% glycerol and protease phosphatase inhibitors (Thermo Fisher Scientific). Lysates were clarified by centrifugation at 4 °C.

Myc-tagged proteins were enriched by incubation of clarified lysates with Myc-Trap agarose beads (ChromoTek) for 1 h at 4 °C with rotation. Beads were washed in a buffer containing 0.05% *n*-dodecyl-β-D-maltoside, and bound proteins were recovered for analysis. Input, unbound and bound fractions were quantified by BCA assay (Thermo Fisher Scientific), denatured in LDS sample buffer with reducing agent and resolved by SDS–PAGE on 4–12% Bis-Tris gels (Thermo Fisher Scientific). Proteins were transferred to polyvinylidene fluoride membranes, stained with Revert 700 Total Protein Stain (LI-COR) and imaged before immunoblotting.

Membranes were blocked and probed with antibodies against the Myc tag (Cell Signaling Technology 2278, 1:1,000) or GFP (Invitrogen A-11122, 1:2,000). Fluorescent secondary antibodies were used at 1:10,000, and blots were imaged using an Odyssey DLx Imaging System (LI-COR).

Querying a CRISPRi screen for regulators of ILC3 inflammatory response for multiCOGS-prioritized genes

We downloaded Supplementary Table 5 from Brown et al.[76] and filtered it to include 8,149 genes that had been profiled in the multiCOGS experiment, based on an identical gene name between the mouse and human data. These genes were ranked based on their MAGeCK[106] score for positive or negative regulation of IL-22 production. We then ran gene set enrichment analysis against each of these rankings for the 142 multiCOGS genes for inflammatory traits, using the 'pathway' function in MAGeCK (v.0.5.9.5) with 100,000 permutations. We considered significant CRISPRi genes to be those with adjusted $P < 0.05$ in the gene-level robust rank aggregation analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw PCHi-C data generated in this study for ILC3s have been deposited at GEO under accession code GSE216267. Processed R data files containing CHiCAGO scores at fragment-level and 5-kb binned resolution can be found in the same repository. PCHi-C data for $CD4^+$ T cells were deposited to the European Genome-Phenome Archive under managed access in accordance with the conditions of donor consent, under accession code EGAS50000001316. Raw RNA-seq reads and counts for the *Cln3*-CRISPRi and *Cln3*-CRISPRa experiments in MNK-3 cells have been deposited at GEO under accession code GSE313942. Large data files S1–S12, including significant CHiCAGO interactions at fragment-level and 5-kb resolution in ILC3 and $CD4^+$ T cells, ABCC pairs in both cell types, and DESeq2 objects for the *Cln3*-CRISPRi and *Cln3*-CRISPRa experiments, were deposited at OSF[15] (https://osf.io/aq9fb).

Code availability

All code repositories generated in this study, including ABCC, RELI, multiCOGS, fine-mapping and imputation, and custom analysis scripts, are available via GitHub at https://github.com/topics/malysheva-et-al, and their snapshots at the time of acceptance of this paper have been archived at Zenodo[107–112].

References

89. Reighard, S. D. et al. Therapeutic targeting of follicular T cells with chimeric antigen receptor-expressing natural killer cells. *Cell Rep. Med.* **1**, 100003 (2020).
90. Krabbendam, L., Nagasawa, M., Spits, H. & Bal, S. M. Isolation of human innate lymphoid cells. *Curr. Protoc. Immunol.* **122**, e55 (2018).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
92. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
93. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
94. Xu, Z. et al. Structural variants drive context-dependent oncogene activation in cancer. *Nature* **612**, 564–572 (2022).
95. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
96. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).
97. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
98. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).
99. Hinrichs, A. S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
100. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
101. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
102. Saini, A. et al. Cell type-specific enhancers regulate IL-22 expression in innate and adaptive type 3 lymphoid cells. *Nat. Commun.* **17**, 3960 (2026).
103. Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
104. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
105. Moyer, T. C. & Holland, A. J. Generation of a conditional analog-sensitive kinase in human cells using CRISPR/Cas9-mediated genome engineering. *Methods Cell Biol.* **129**, 19–36 (2015).
106. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
107. Artemov, P., Wayman, J. A. & Spivakov, M. FunctionalGeneControl/ABCC. *Zenodo* https://doi.org/10.5281/ZENODO.20625808 (2026).
108. Ray-Jones, H., Spivakov, M. & Burren, O. FunctionalGeneControl/multiCOGS. *Zenodo* https://doi.org/10.5281/ZENODO.20446607 (2026).
109. Ray-Jones, H., Yang, Z. F., Cazares, T. & Miraldi, E. R. FunctionalGeneControl/RELI_analysis_malysheva_etal. *Zenodo* https://doi.org/10.5281/ZENODO.20626094 (2026).

110. Wallace, C. FunctionalGeneControl/imputation_finemapping_scripts_malysheva_etal. *Zenodo* https://doi.org/10.5281/ZENODO.20626081 (2026).
111. Ray-Jones, H. FunctionalGeneControl/malysheva_etal_downstream_analysis. *Zenodo* https://doi.org/10.5281/ZENODO.20629757 (2026).
112. Malysheva, V. & Spivakov, M. FunctionalGeneControl/malysheva_etal_additional_analysis. *Zenodo* https://doi.org/10.5281/ZENODO.20629884 (2026).
113. Chicco, D. & Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min.* **16**, 4 (2023).

Acknowledgements

We thank L. Game and I. Andrew at the LMS Genomics facility for sequencing, M. Thiecke and O. S. Burren for technical advice and W. Leung for technical assistance. We also thank the UK's National Institute for Health and Care Research (NIHR) BioResource volunteers for their participation and acknowledge NIHR BioResource Centres. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. For the purpose of open access, the authors have applied a CC-BY public copyright licence to any author accepted manuscript version arising from this submission.

Author contributions

Conceptualization: V.M., H.R.-J., C.W., S.N.W. and M.S. Data curation: V.M., H.R.-J., P.A., J.A.W., X.C., S.P. and M.S. Formal analysis: V.M., H.R.-J., N.L., T.A.C., P.A., J.A.W., Z.F.Y., X.C., S.P., C.W. and M.S. Funding acquisition: V.M., P.F., E.R.M., S.N.W. and M.S. Investigation: V.M., H.R.-J., N.L., R.A.B., T.A.C., O.C., D.E.O., P.A., J.A.W., C.P., J.R.B., X.C., S.P., N.P., C.W., S.N.W. and M.S. Methodology: V.M., H.R.-J., N.L., R.A.B., M.D.R., C.P., J.I.J.E., W.R.O., T.N., P.F., S.S., M.T.W., L.C.K., C.W., S.N.W. and M.S. Project administration: V.M., H.R.-J. and M.S. Resources: R.A.B., T.A.C., C.B., X.C., S.P., A.W.D., A.S., F.B., M.F., D.F.S., N.P., J.M.W., E.M.O., C.W. and M.S. Software: H.R.-J., P.A., J.I.J.E., M.T.W., L.C.K., C.W. and M.S. Supervision: V.M., H.R.-J., P.A., M.F., M.T.W., L.C.K., J.M.W., E.M.O., C.W., E.R.M., S.N.W. and M.S. Validation: H.R.-J., N.L. and R.A.B. Visualization: V.M., H.R.-J., N.L., P.A. and M.S. Writing—original draft: V.M., H.R.-J. and M.S. Writing—review and editing: V.M., H.R.-J., N.L., S.N.W. and M.S., with contributions from all authors. Joint first authors: V.M., H.R.-J. and N.L. Joint second authors: R.A.B., T.A.C., O.C., D.E.O. and P.A. Joint principal supervisors: C.W., E.R.M., S.N.W. and M.S.

Funding

M.S. is core-funded by the Medical Research Council as an MRC Programme Leader (MC-A652-5QA20). C.W. is funded by the Wellcome Trust (WT220788), the MRC (MC_UU_00002/4), GSK and MSD, and supported by the NIHR Cambridge BRC (BRC-1215-20014). Research activities related to this manuscript that were conducted at Cincinnati Children's Hospital Medical Center were supported by the NIH (U01AI150748 to E.R.M., M.T.W. and L.C.K.; R01AI153442 to E.R.M.; R21AI156185 to E.R.M.; R01HG010730, U01AI130830, R01NS099068, R01GM055479, P01AI150585 and R01AI141569 to M.T.W.; R01AI024717 to M.T.W. and L.C.K.; R01AR073228 to M.T.W., L.C.K. and S.N.W.; and R01DK107502, R01AI148276, U19AI070235, U01HG011172 and P30AR070549 to L.C.K.); Cincinnati Children's Research Foundation (ARC Award to E.R.M., M.T.W. and L.C.K.; Center for Pediatric Genomics grants to E.R.M., M.T.W. and L.C.K.); and the L.B. Research and Education Foundation (to N.L.). M.T.W., S.N.W. and E.R.M. are partially supported by the Pilot Grant from the Cincinnati Digestive Health Center (P30 DK078392). O.C. was supported by NIH training grant T32 AR069512. D.F.S. acknowledges support by NIH grant 5K08HL148551-02. D.E.O. was supported by an American Heart Association fellowship. Research activities related to this manuscript that were conducted at Ohio State University's Medical Center were supported by the Pelotonia Scholars Graduate (R.A.B.) and PostDoctoral (A.S.) Fellowships. This project was funded in part by a Commercialisation Award from the Babraham Institute to M.S. and P.F. and by VIB core funding to V.M. M.F. was supported by the British Heart Foundation (BHF; FS/18/53/33863), the BHF Cambridge Centre for Research Excellence (RE/18/1/34212) and the NIHR Exeter Biomedical Research Centre.

Competing interests

P.F., S.S. and M.S. are shareholders of Enhanced Genomics Ltd. J.M.W. is an employee of Amicus Therapeutics, Inc., and holds equity in the company in the form of stock-based compensation; Amicus had no input into this work. C.W. is also an employee of GSK; GSK had no input into this work. C.P. is an employee of Vitaccess; Vitaccess had no input into this work. M.D.R. is an employee of NeoGenomics Laboratories Inc.; NeoGenomics Laboratories had no input into this work. V.M., M.S., S.S., T.N. and P.F. are coinventors on a patent (WO2021064430A1) filed by the Babraham Institute that describes the low-input PCHi-C technology. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02681-0.

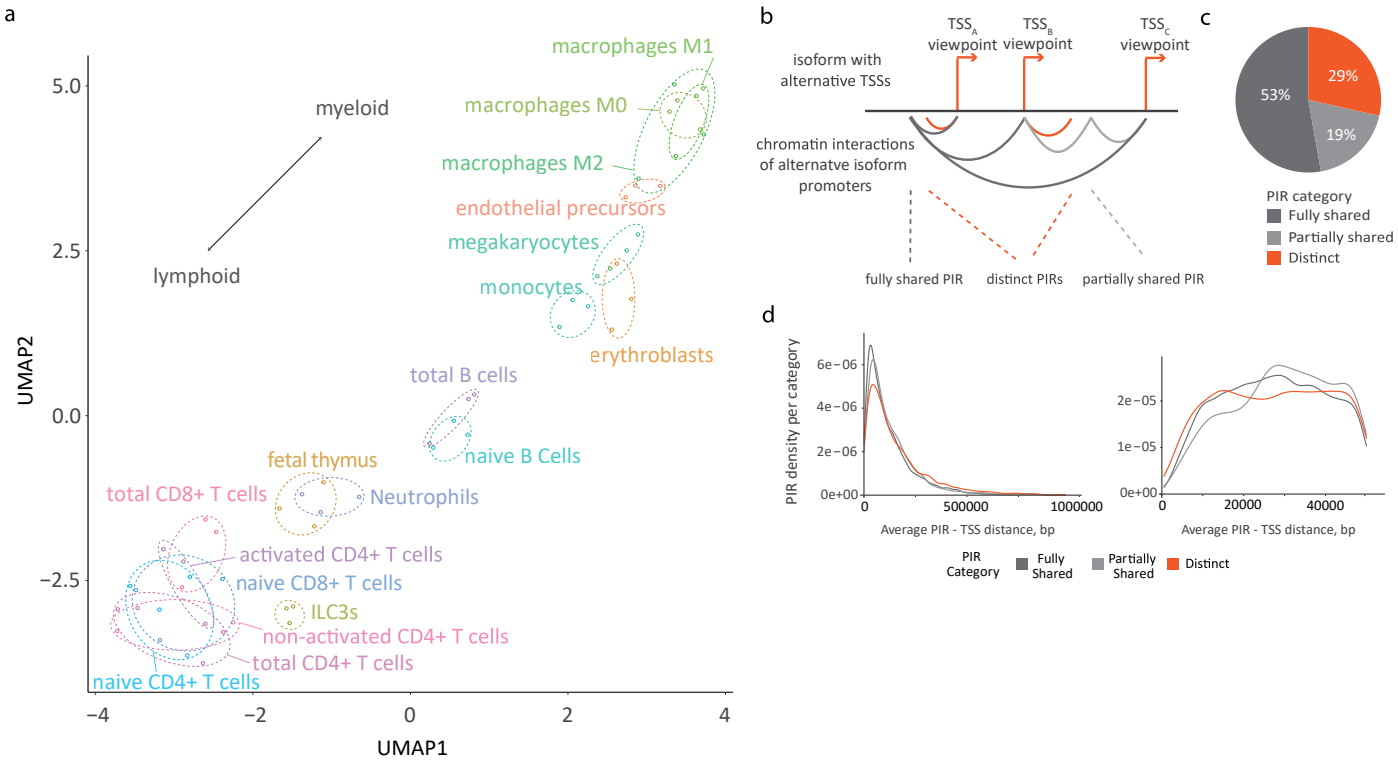
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02681-0.

Correspondence and requests for materials should be addressed to Valeriya Malysheva, Stephen N. Waggoner or Mikhail Spivakov.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

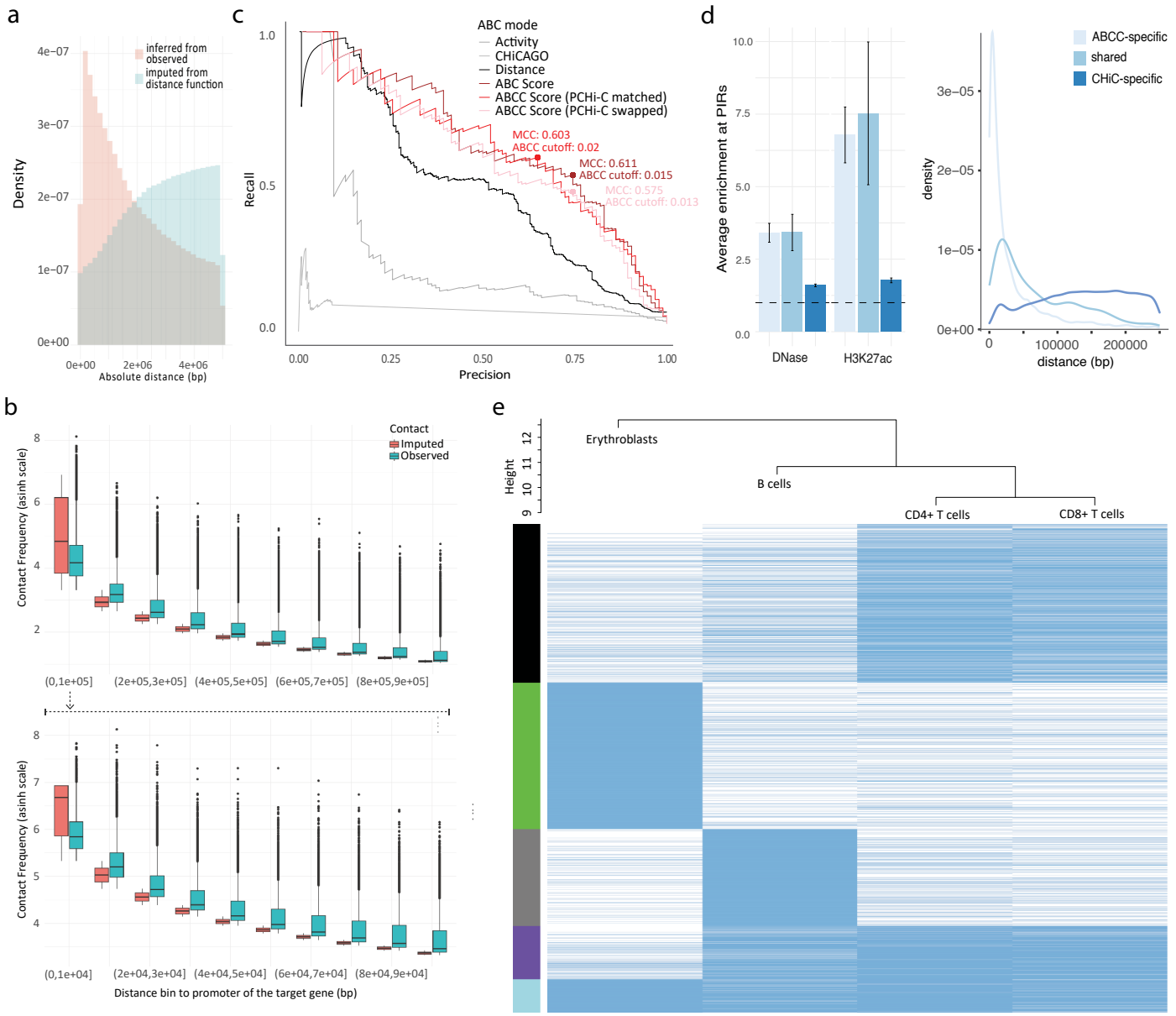
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41588-026-02681-0



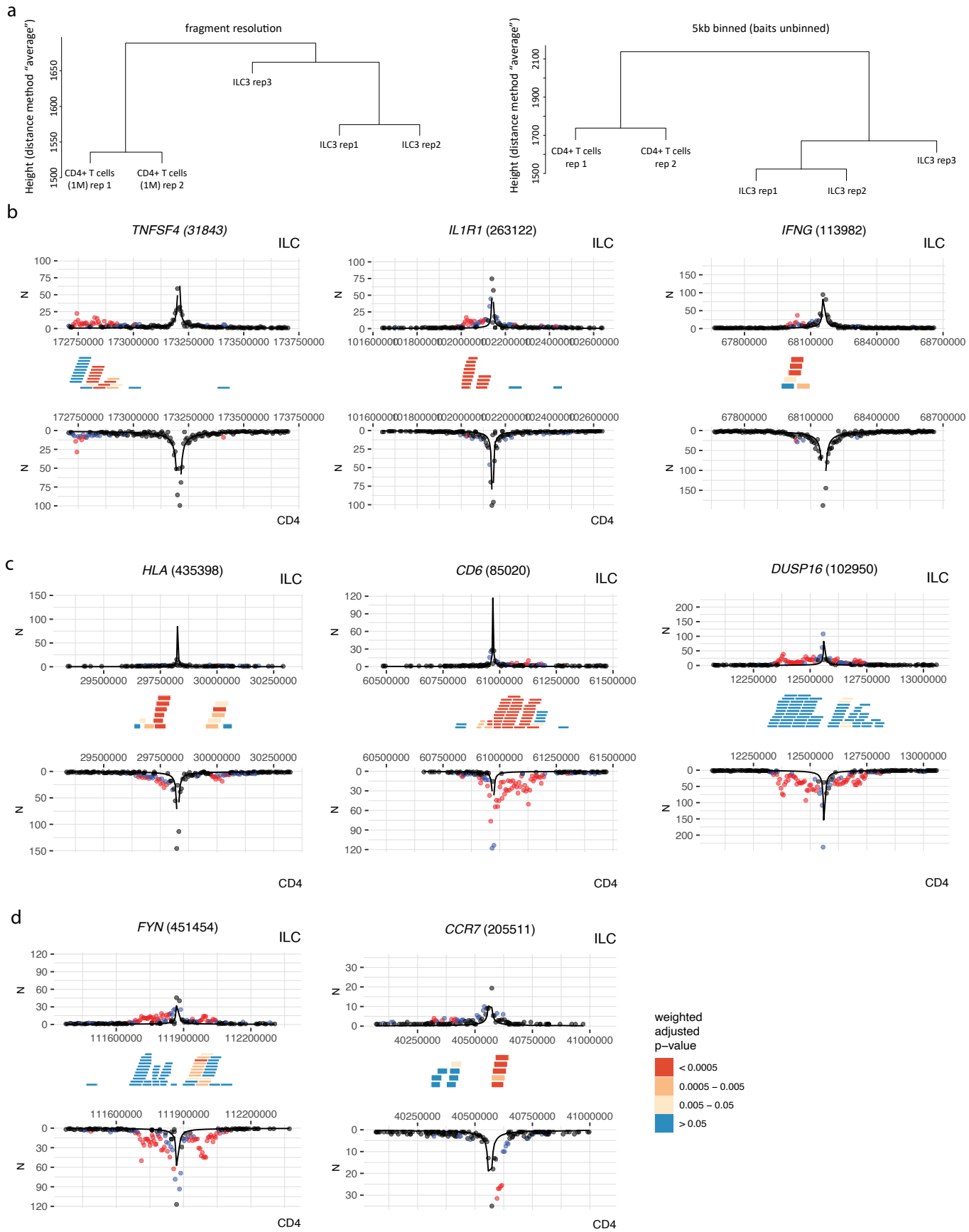
Extended Data Fig. 1 | Compendium of promoter-enhancer interactions in ILC3s. a, UMAP of CHiCAGO scores detected for PCHi-C in ILC3s versus public data in 17 primary human blood cell types[17]. **b**, Scheme representing the classification of PIRs detected at alternative transcription start sites (ATSSs) of the same gene: 'fully shared' (shared across all captured ATSSs), partially shared and distinct (unique to a single ATSS). **c**, Pie chart showing the degree of enhancer sharing across alternative transcription start sites (ATSSs) for short-range contacts. **d**, Distance distribution of ATSS-specific and shared PIRs at 5-kb binned (baits unbinned) resolution. Left plot shows interactions up to 1 Mb (Kruskal-Wallis test $P < 2.22 \times 10^{-16}$; pairwise Wilcoxon test $P = 8.68 \times 10^{-6}$ (partially shared vs. fully shared), $P = 4.46 \times 10^{-8}$ (partially shared vs. distinct) and $P < 2.22 \times 10^{-16}$ (fully shared vs. distinct)). Right plot shows interactions up to 50 kb (Kruskal-Wallis test $P = 7.65 \times 10^{-5}$; pairwise Wilcoxon test $P = 9.8 \times 10^{-5}$ (partially shared vs. fully shared), $P = 6 \times 10^{-4}$ (partially shared vs. distinct) and $P = 1$ (fully shared vs. distinct)).

https://doi.org/10.1038/s41588-026-02681-0



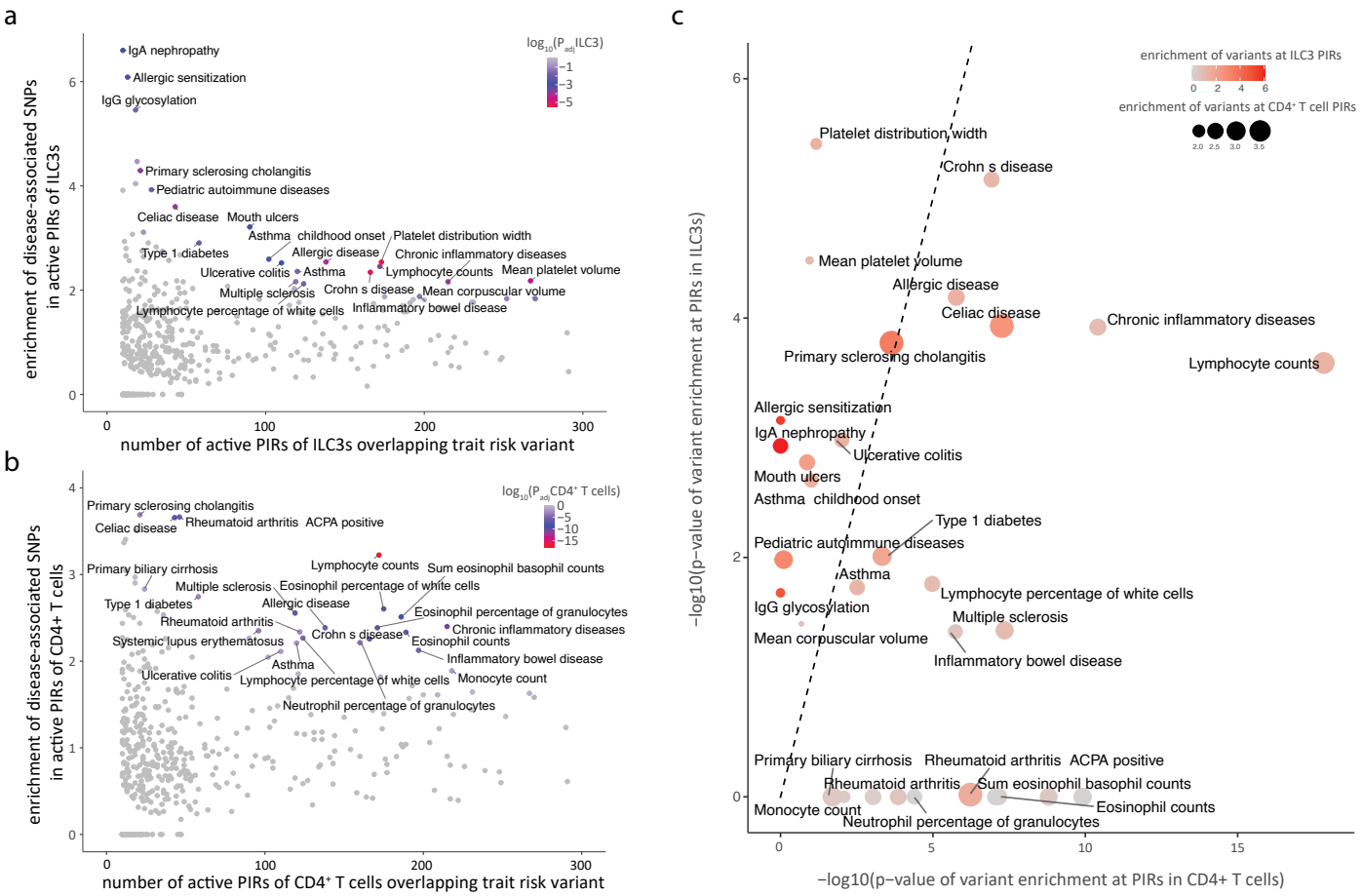
Extended Data Fig. 2 | Benchmarking of the ABCC approach with public data. a, Density distribution of promoter interactions inferred from observed PCHi-C contact frequencies (pink) and those imputed using the CHiCAGO distance function (cyan) across genomic interaction distances. **b**, Contact frequency distributions stratified by distance. Observed PCHi-C contacts are shown in green; imputed contacts (using expected frequencies estimated using the CHiCAGO distance function) are shown in blue. Similarly to standard ABC, frequency capping is introduced for short-range imputed contacts (<5 kb). **c**, Precision–recall curves benchmarking the predictive performance of different scoring approaches for enhancer–promoter interactions in erythroblasts. Curves compare the scoring across: CHiCAGO-detected contacts, Activity alone, Distance alone, the conventional ABC score, and PCHi-C-based ABCC score in two modes: 'matched' (using PCHi-C cell-type specific profile for erythroid cells) and 'swapped' (in which a PCHi-C dataset with a similar read coverage from a different cell type, CD4 + T cells, is used instead). MCC, Matthews correlation coefficient, an alternative to the AUC metric that is more informative under class imbalance and more sensitive to performance at a fixed decision threshold[113]. **d**, Enrichment of epigenetic marks at PIRs: DNase marking chromatin accessibility and H3K27ac marking active enhancers (left plot) and distance distribution of ABCC-specific, PCHiC-specific and shared enhancer-promoter links (right plot) in K562 cells for 0.023 ABCC threshold. **e**, Hierarchical clustering heatmap of enhancer–promoter interactions predicted with ABCC across cell types (erythroblasts, B cells, CD4 + T helper cells, CD8 + T cells).

https://doi.org/10.1038/s41588-026-02681-0



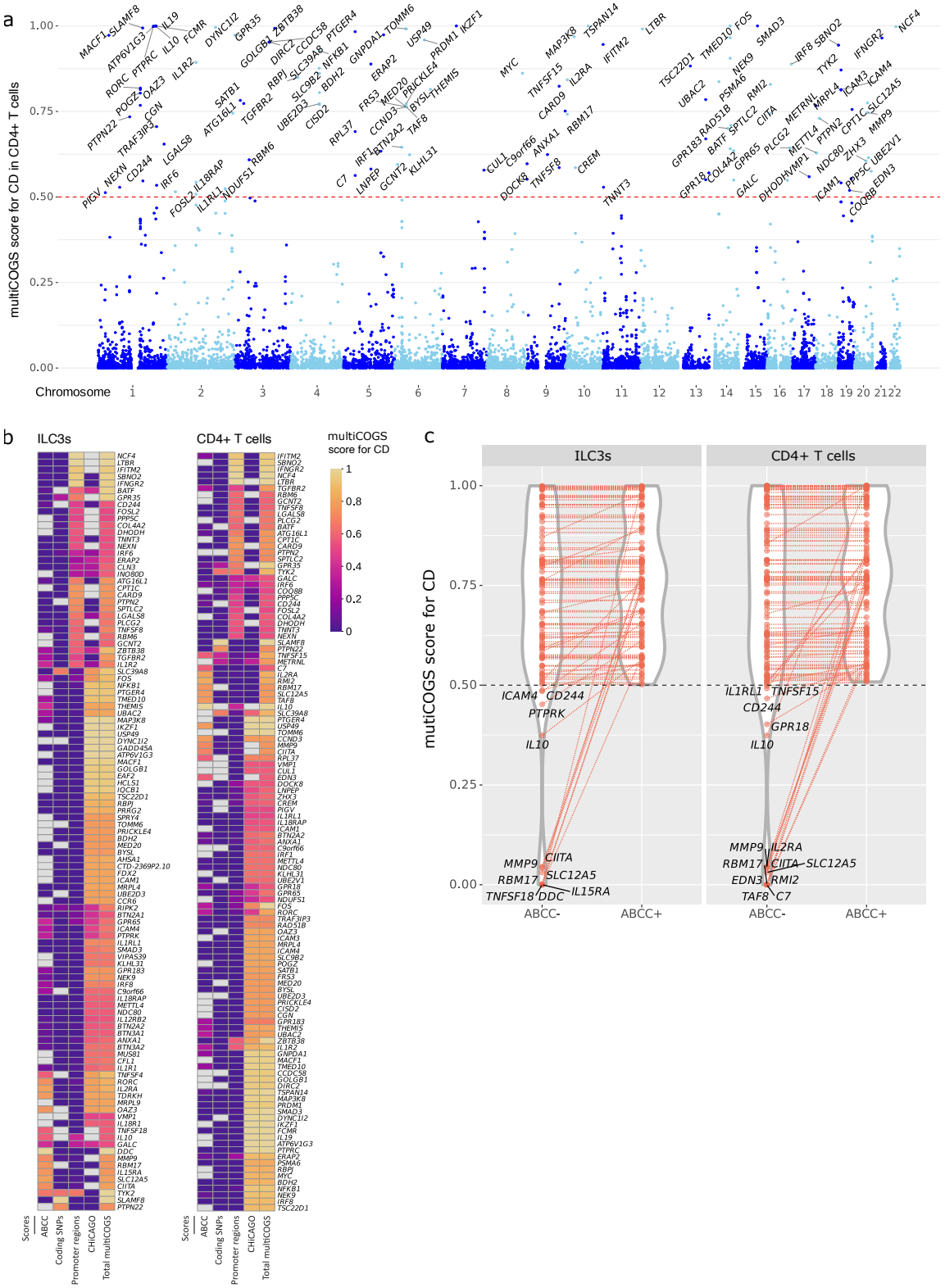
Extended Data Fig. 3 | Genes with differential contacts in ILC3s and CD4 + T cells. a, Hierarchical clustering of CHiCAGO signals based on data in ILC3s and CD4 + T cells. **b**-**d**, Examples of captured promoters with differential wiring between ILC3s and CD4 + T cells: promoters with stronger (**b**) and weaker (**c**) contacts in ILC3s compared with CD4 + T cells, as well as with contacts in both cell types (**d**).

https://doi.org/10.1038/s41588-026-02681-0



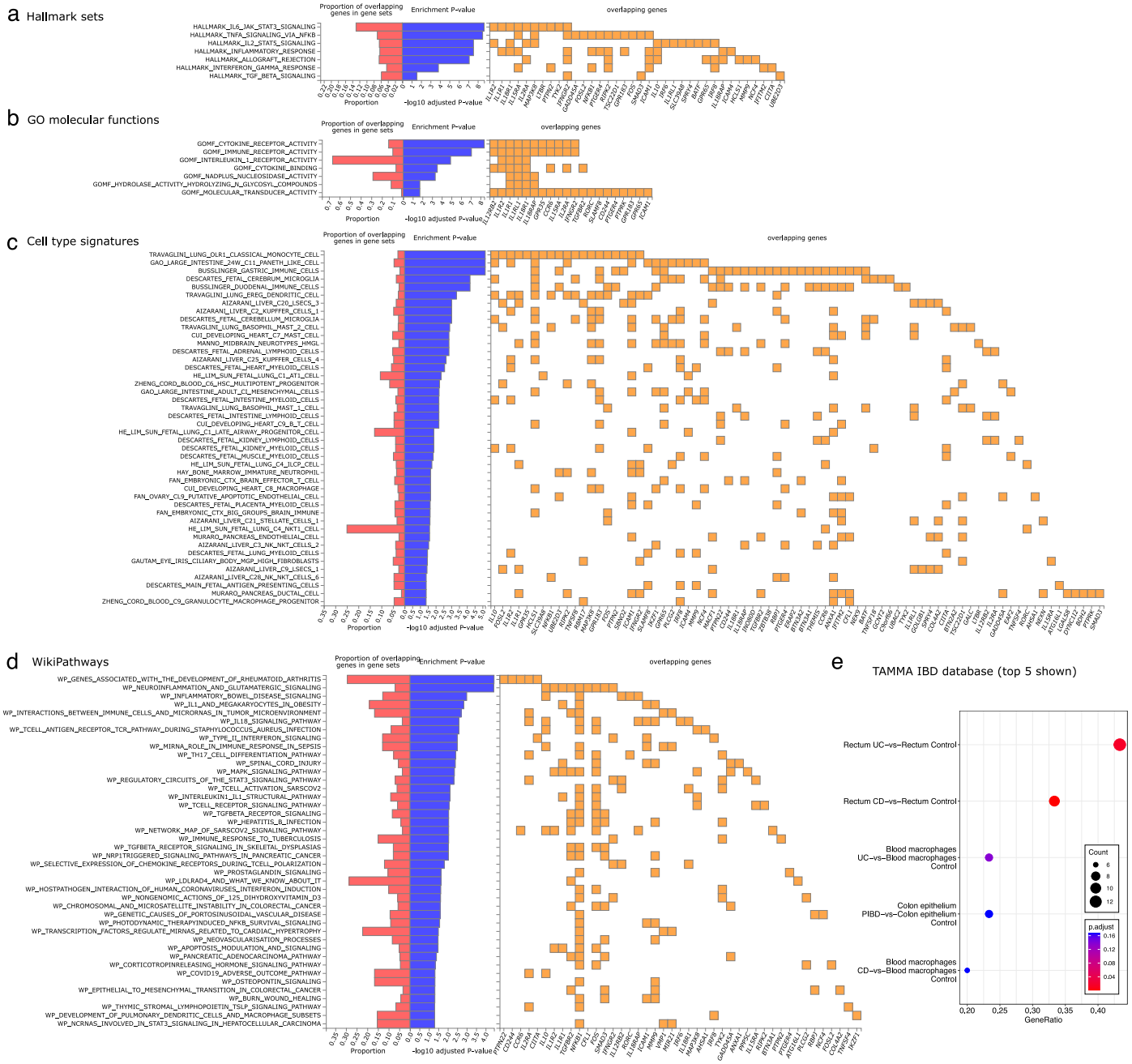
Extended Data Fig. 4 | Additional information for the RELI analysis of risk loci enriched in ILC3 and CD4 + T-cell PIRs. a,**b**, RELI enrichment of risk variants in ILC3s (**a**) and CD4s active PIRs (**b**) across 495 diseases and traits. **c**, Adjusted *P*-value of RELI enrichment of risk variants ILC3s vs. CD4s active PIRs across 495 diseases and traits. In all panels, traits with Bonferroni-adjusted *P*-value < 0.05 in the respective cell type(s) are labelled.

https://doi.org/10.1038/s41588-026-02681-0



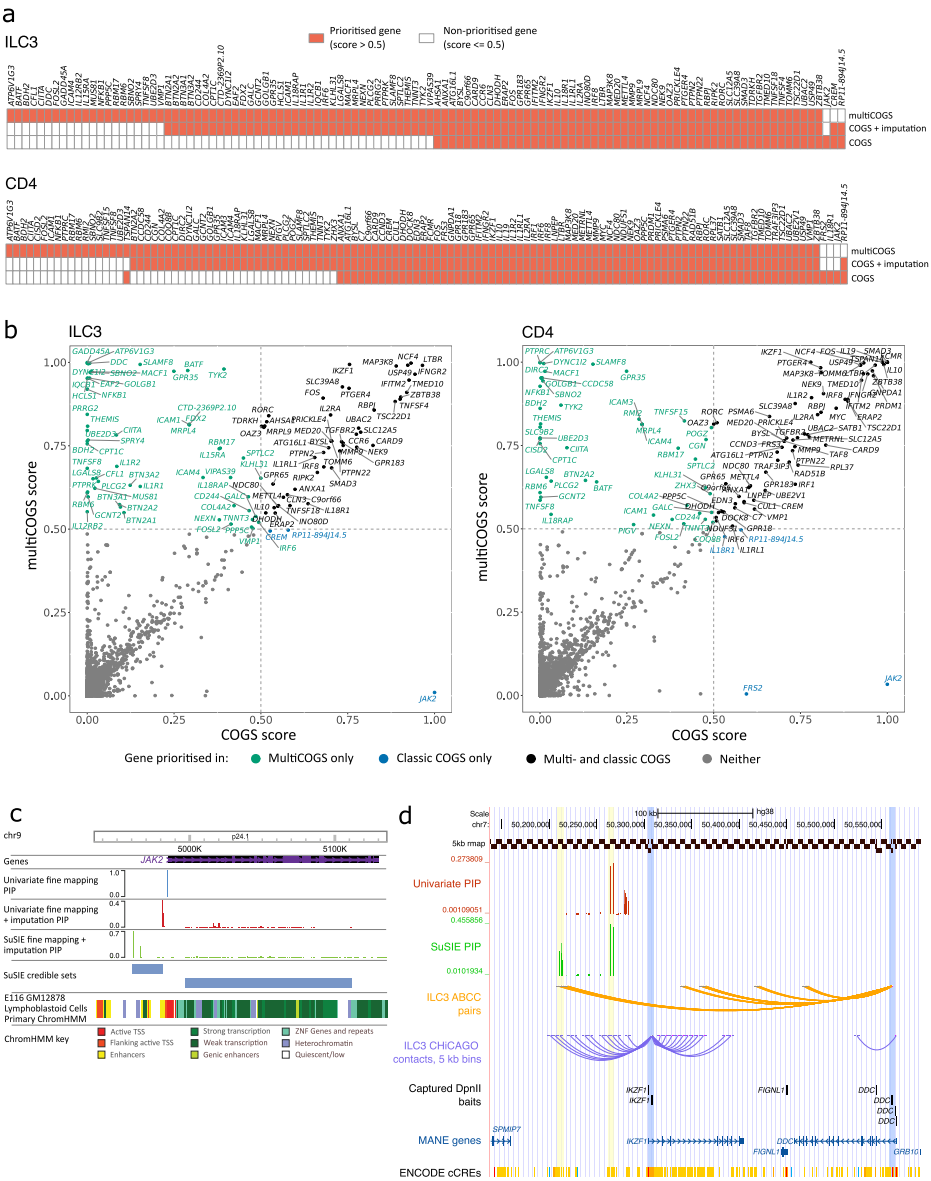
Extended Data Fig. 5 | MultiCOGS prioritizes candidate gene sets for Crohn's disease risk. a, Manhattan plot showing multiCOGS scores for Crohn's disease (CD) risk based on promoter contacts in CD4 + T cells. **b**, Heatmaps of region contributions to multiCOGS scores in ILC3s and CD4 + T cells in CD. **c**, Illustration of genes that were only prioritized for CD with the addition of ABCC, in ILC3s and CD4 + T cells. In each graph, the multiCOGS score with and without ABCC is plotted for all genes that were prioritized in the full multiCOGS run (score > 0.5 with ABCC).

https://doi.org/10.1038/s41588-026-02681-0



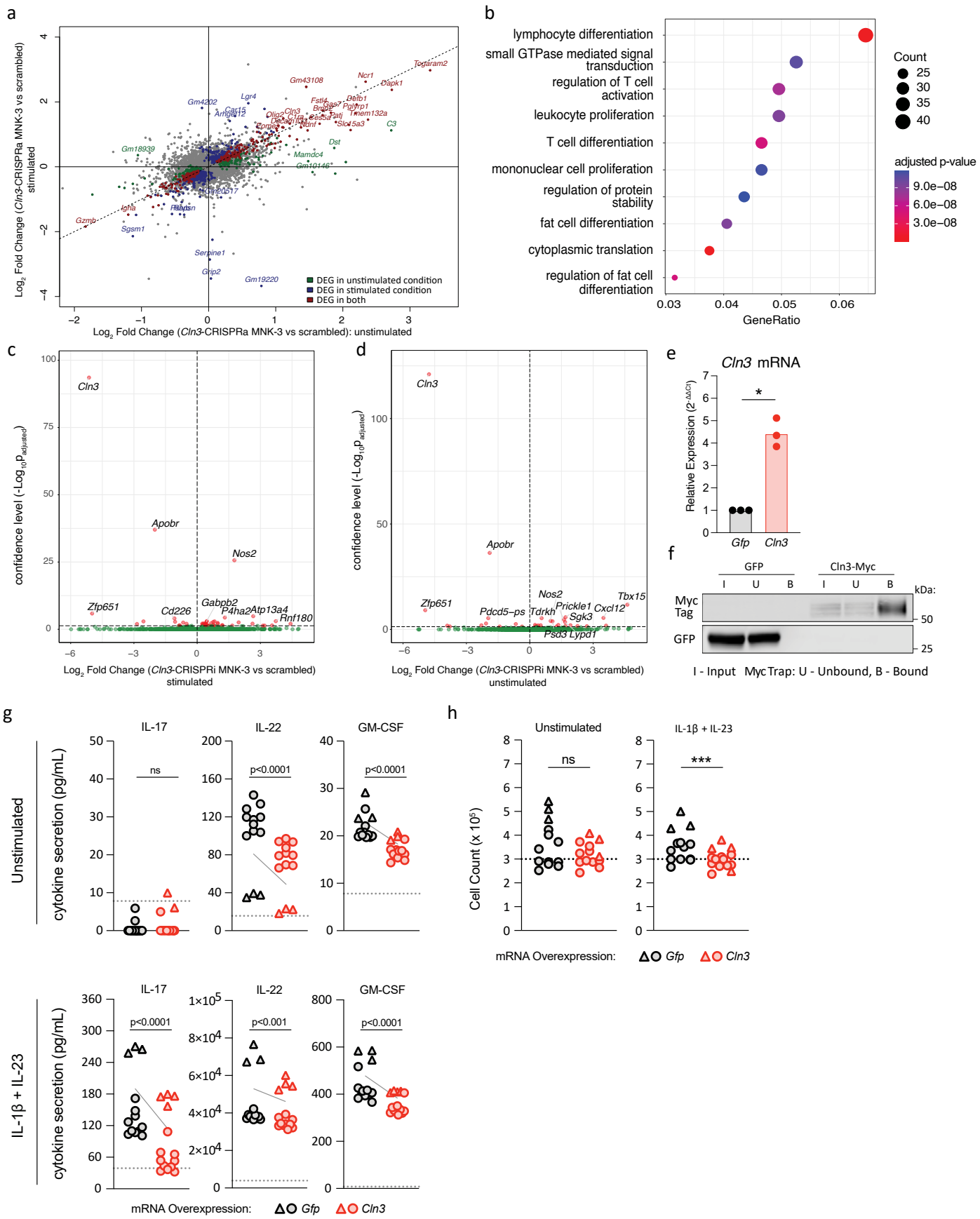
Extended Data Fig. 6 | Biological annotation of multiCOGS-prioritized Crohn's disease effector genes in ILC3s. a-d, Enriched gene sets among multiCOGS genes detected using the GENE2FUNC pipeline in FUMA[86] for the following databases: MSigDB hallmark sets (**a**), GO molecular functions (**b**), MSigDB cell type signatures (**c**), MSigDB WikiPathways (**d**). **e**, Enrichment analysis for differentially expressed gene sets among multiCOGS genes in the TAMMA IBD database.

https://doi.org/10.1038/s41588-026-02681-0



Extended Data Fig. 7 | Comparison of gene prioritization for Crohn's disease by classic COGS versus multiCOGS. a, Comparison of prioritized gene sets between classic COGS, classic COGS plus imputation, and multiCOGS (that is, imputation plus multivariate fine mapping, processed via the multiCOGS algorithm) for Crohn's disease (CD). Shown for ILC3s and CD4 + T cells. **b**, Comparison of COGS scores and multiCOGS scores for genes in ILC3 cells (left) and CD4 + T cells (right) for Crohn's disease. Green labels indicate genes prioritized in multiCOGS only, blue in classic COGS only, and black in both. **c**, Plot of the *JAK2* locus showing the shift of the most likely causal variant from the promoter of *JAK2* to a region around 20 kb upstream of the promoter upon multivariate fine mapping, leading to a lower multiCOGS vs. classic COGS score. No chromosomal interactions were observed between this region and the *JAK2* promoter in ILC3 or CD4 + T cells. **d**, Illustration of multiCOGS prioritization of *IKZF1* and *DDC* in ILC3s at the chr7p locus. At this locus, multivariate fine mapping identifies two credible sets of variants (yellow vertical bars), whereas univariate fine mapping only detects one. PCHi-C interactions connect these likely causal variants to the *IKZF1* promoter (first blue vertical bar). However, ABCC interactions also connect one of the credible sets to the DDC promoter (second blue vertical bar). Thus, multiCOGS prioritises both genes, whereas classic COGS prioritises only *IKZF1*. PIP; posterior inclusion probability.

https://doi.org/10.1038/s41588-026-02681-0



Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41588-026-02681-0

Extended Data Fig. 8 | Additional results on the role of CLN3 in ILC3 inflammatory function. a, Comparison of differentially expressed genes in IL-23/IL-1β-stimulated vs unstimulated *Cln3*-CRISPRa cells (relative to scrambled gRNA controls). **b**, GO term enrichment analysis for genes differentially expressed upon *Cln3*-CRISPRa stimulation. **c**, Differential expression of genes in IL-23/IL-1β-stimulated *Cln3*-CRISPRi MNK-3 cells vs. scrambled gRNA controls. Red, differentially expressed genes (DESeq2 adjusted *P*-value < 0.05); green, all other genes. **d**, Differential expression of genes in unstimulated *Cln3*-CRISPRi MNK-3 cells vs. scrambled gRNA controls. Red, differentially expressed genes (DESeq2 adjusted *P*-value < 0.05); green, all other genes. **e**, *Cln3* expression in MNK-3 cells electroporated with *Cln3*-myc mRNA or GFP mRNA. Transcript abundance was quantified by qPCR, normalized to *Hprt*, and expressed relative to the GFP mRNA control. Each point represents an independent experiment. Statistical significance was assessed using a paired *t*-test, $P < 0.05$ (*). **f**, Verification of CLN3-Myc protein expression and Myc tag-dependent pulldown. MNK-3 cells were lysed, subjected to immunoprecipitation using Myc-Trap agarose, and resolved by reducing SDS-PAGE. I, input lysate; U, unbound fraction; B, bead-bound fraction. Immunoblotting with anti-Myc tag antibodies detected a ~ 65–80 kDa CLN3-Myc species selectively enriched in the bound fraction. The uncropped Western blot is shown in Supplementary Fig. 2. **g**, Cytokine secretion upon *Cln3* overexpression across independent experiments. MNK-3 cells were electroporated with GFP mRNA (black) or *Cln3-Myc* mRNA (red) and cultured for 24 h under unstimulated (top row) or IL-1β + IL-23–stimulated (bottom row) conditions. Cytokine concentrations (IL-17, IL-22, GM-CSF) in culture supernatants were quantified by ELISA. Each symbol represents a biological replicate from two independent experiments (triangles vs. circles). The dotted horizontal line indicates the lower limit of quantification for each assay. Statistical significance was assessed using a linear mixed-effects model with experiment as a random effect and transfection as a fixed effect ($n = 13–14$ per condition). Solid grey lines indicate group means, with dotted grey bands indicating 95% confidence intervals of the fixed-effect. **h**, Cell numbers upon *Cln3* overexpression with and without inflammatory stimulation. MNK-3 cells were electroporated with GFP mRNA (black) or Cln3-myc mRNA (red) and cultured for 24 hr in unstimulated (left) or IL-1β + IL-23–stimulated (right) media. Each symbol represents a biological replicate from two independent experiments (triangles vs. circles). Lines connect the experiment-specific means. Viable cell numbers were quantified by trypan blue exclusion. The dotted line indicates the number of cells seeded at 0 h. Statistical significance was assessed using a linear mixed-effects model with experiment as a random effect and transfection as a fixed effect ($n = 13–14$ per condition). ns, not significant.

nature portfolio

Corresponding author(s): Valeriya Malysheva
Stephen Waggoner
Mikhail Spivakov

Last updated by author(s): 10th June 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	PCHi-C data was pre-processed and analysed using Bowtie2, HiCUP and CHiCAGO. ChIP-seq data was pre-processed using Trim galore, Fastqc, STAR, Samtools and Bedtools and peaks were called using MACS2. Microarray CEL files were analysed using AltAnalyze. Human RNA-seq was processed using Salmon and Tx import. Mouse RNA-seq was analysed using kallisto (v0.48.0), tximport (v1.30.0) and DESeq2 (v1.38.0). Enrichment of TFs in active and/or open regulatory elements of multiCOGS-prioritised genes was deteced using LOLA v1.18. Enrichment of trait-associated variants within open + active PIRs in ILC3s and CD4+ T cells was detected using RELI v0.1.1a. Liftover of genomic regions was performed using UCSC LiftOver v377. The mutiCOGS algorithm is based on the R package rCOGS, leveraging the R package data.table. To identify coding SNPs in the workflow, we used VEP version 99.2. The multivariate fine mapping integrated into multiCOGS was performed using the susieR package. Pathway analysis in the multiCOGS vs CRISPRi screen genes was performed using MAGeCK v0.5.9.5. The scripts used to generate the data in this study are freely available on Github. All repositories containing code used in this study are listed under the topic "malysheva et al", available at: https://github.com/topics/malysheva-et-al

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw Promoter Capture Hi-C data generated in this study are deposited in the Gene Expression Omnibus (GEO) under the accession number GSE216267 (ILC3s) and in the European Genome-Phenome Archive (EGA) under managed access, accession number EGAS50000001316 (owing to conditions of donor consent). Raw RNA-seq reads and counts for the CRISPRi/a experiments at CLN3 in MNK3 cells are deposited in GEO under the accession number GSE313942. Supplementary Data files, including significant CHiCAGO interactions at fragment-level and 5kb resolution in ILC3 and CD4+ T cells, ABCC pairs in both cell types and DESeq2 objects for the CLN3 CRISPRi/a experiments, were deposited to Open Science Framework (https://osf.io/aq9fb).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Two male and one female child participated in our study for ILC3s and two adult male donors for CD4+ T cells
Population characteristics	ILC3s: Age: 6, Non-Hispanic European American Age:5, Non-Hispanic African American Age:6, Non-Hispanic European American CD4+ T cells: Two adult male donors of European ancestry (age not disclosed due to consent conditions)
Recruitment	ILC3: Three children requiring tonsillectomy were recruited to a prospective study at a tertiary academic care center through the division of Pediatric Otolaryngology-Head and Neck surgery at Cincinnati Children's Hospital Medical Center. Criteria for enrollment in the study included a history of sleep-disordered breathing or recurrent or chronic tonsillitis requiring removal of the tonsillar tissue. Subjects were excluded from the study if the tonsillar tissue was acutely infected or if anatomic abnormalities were present requiring a more detailed pathologic evaluation post the surgical procedure. CD4+ T cells: Donors were recruited through NIHR Cambridge BioResource (UK) to donate blood for research purposes at an NHS Blood and Transplant facility at the University of Cambridge. We have not identified any bias in the recruitment of participants.
Ethics oversight	ILC3: Consent was obtained from parents in the perioperative suite on the day of the procedure with approval of the institutional review board (IRB) “Tissues for the study of functional genomics and immunity” (no. 2018-3069) at the Cincinnati Children's Hospital Medical Center. CD4+ T cells: The samples were obtained after written informed consent under studies “A Blueprint of Blood Cells,” REC reference 12/EE/0040, and “Genes and mechanisms in type 1 diabetes in the Cambridge BioResource,” REC reference 05/Q0106/20; both NRES Committee East of England – Cambridgeshire and Hertfordshire. Both: Samples were labelled with a de-identified barcode and transferred to the research team for further processing.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To study chromatin organistaion we generated Promoter Capture Hi-C data for two (CD4+ T cells) or three (ILC3 cells) biological replicates. For the CRISPR activation/interference experiments, we included three biological replicates per condition.

	For the CLN3 overexpression experiments, we performed two independent experiments; the first with three biological replicates and the second with 10 biological replicates.
Data exclusions	No data was excluded from the analysis.
Replication	The number of replicates was determined based on material availability and common practice for each experiment type. For Promoter Capture Hi-C, the use of two biological replicates is standard in the field and sufficient for detecting significant interactions using the CHiCAGO methodology (Freire-Pritchett/Ray-Jones et al., Nature Protocols 2021). For the other experiments, we did not pre-determine sample sizes based on statistical methodologies. We have not reported any findings that failed replication.
Randomization	Randomization is not relevant for this study, due to the focus on a specific cell type in control samples.
Blinding	Blinding is not relevant for this study, due to the focus on a specific cell type in control samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Human
CD19 (HIB19) BV421 1:50 Biolegend #302234
CD14 (63D3) BV421 1:50 Biolegend #367144
CD3 (OKT3) BV421 1:50 Biolegend #317344
CD45 (HI30) AF488 1:200 Biolegend #304017
CD94 (DX22) PerCP-Cy5.5 1:50 Biolegend #305514
CD127 (IL-7Rα) (A019D5) PE-Cy7 1:50 Biolegend #351320
cKit (CD117) (104D2) BV510 1:50 Biolegend #313220
NKp44 (P44-8) AF647 1:50 Biolegend #325112
CRTH2 (301109) PE 1:50 R&D Systems #FAB33381P
Live/dead fixable near-IR Dead cell stain kit 1:1000 Invitrogen #L34993
Human TruStain FcX™ 1:50 Biolegend #422302

Mouse
CCR6 (29-2L17) BV421 1:100 Biolegend #129818
CD3 (17A2) BUV395 1:200 BD Biosciences #740268
CD19 (1D3) BUV395 1:200 BD Biosciences #563557
TCR β Chain (H57-597) BUV395 1:200 BD Biosciences #742485
NK1.1 (PK136) BV711 1:50 Biolegend #108745
NKp46 (CD335) (29A1.4) 1:50 Biolegend #137619
KLRG1 (2F1) FITC 1:100 Biolegend #138410
CD127 (IL-7Rα) (A7R34) PE/Cy7 1:100 BioLegend #135014
CD69 (Y245) BUV737 1:100 BD Biosciences #612793
IL-17A (TC1118H10.1) PE 1:50 BioLegend #506904
CD16/CD32 (2.4G2) 1:200 Tonbo #70-0161-M001
ELISA DuoSet kits for mouse IL-17, IL-22, and GM-CSF (R&D Systems).
Live/dead fixable near-IR Dead cell stain kit 1:1000 Invitrogen #L34993

Validation

Antibody validation is provided on the supplier website(s). All antibodies were validated with proper FMO controls using PBMCs or leukocytes derived from human tonsils.
Supplier validation from Biolegend can be found at: https://www.biolegend.com/en-us/quality/quality-control.
R&D: https://www.rndsystems.com/products/antibodies.
BD Biosciences: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/quality-and-reproducibility.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	MNK-3a and MNK-3i cells were generated in the Oltz lab based on the MNK-3 male mouse cell line (Sigma-Aldrich, SCC628).
Authentication	The cells were not authenticated.
Mycoplasma contamination	The cell lines were mycoplasma-negative at the time of analysis.
Commonly misidentified lines (See ICLAC register)	None

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	No raw ChIP-seq data was generated in this study. Publicly available ChIP-seq data was downloaded from GEO, accession GSE77299. Processed ChIP-seq files are available on: https://github.com/tacazares/spivakov_pchic_ILC_CD4,
Files in database submission	No raw ChIP-seq generated in this study
Genome browser session (e.g. UCSC)	No raw ChIP-seq generated in this study

Methodology

Replicates	Three biological replicates for each assay
Sequencing depth	as described in the original study GSE77299
Antibodies	H3K27ac, H3K4me3
Peak calling parameters	Peaks were called using MACS2 using three biological replicates per sample as the treatment group with an input ChIP-seq control sample. Deafult parameters.
Data quality	The sequencing reads were filtered by PHRED score >=30 and examined for proper pairing with a mate (when paired-end). The sequencing quality and duplication level were checked using Fastqc. Sequences were mapped to the hg38 reference genome using STAR with modification. Samtools was used to select reads with a MAPQ score of 255 which is the flag for unique mapping reads from STAR. PCR duplicates were removed using samtools. We then removed reads that fell within blacklisted regions using Bedtools intersect. The final filtered BAM file was then converted to a bed file using Bedtools bamtobed.
Software	MACS2

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	ILC3 isolation: Human tonsil tissue was processed by mincing with scissors followed by transfer of up to 4g of tissue to a gentleMACS C tube (Miltenyi Biotec) containing 8 mL of phosphate-buffered saline (PBS) with 0.5 mg/mL collagenase D and 3000 U/mL DNaseI, then dissociated on a GentleMACS Octo Dissociator (Miltenyi Biotec) using “program C (Spleen program 2 followed by spleen program 3).” Tissue homogenates were incubated in a 37°C water bath for 15 minutes, then dissociated again using “program C” and transferred through a 100 µm cell strainer into 20mL RPMI containing 10% human AB serum (Sigma Aldrich). Next, cell suspension was overlaid on 10mL of Ficoll-Paque PLUS (GE Healthcare) and subjected to density-gradient separation via centrifugation for 20 min at 1800rpm, 20°C, slow acceleration, no brake. Leukocytes were collected from the interphase layer and then washed with 50mL of PBS for 6 minutes at 1600rpm for 7 minutes, 20°C.

	Single cell suspensions of tonsil mononuclear cells were subjected to positive selection with anti-human-CD3, anti-human-CD19 and anti-human-CD14 (Miltenyi Biotec) and transferred through LD columns (Miltenyi Biotec) according to the manufacturer guidelines. The depleted cell suspension flowthrough was collected into 15mL conical and then centrifuged for 5 minutes at 1200rpm, 20°C. Subsequently, cells were labeled with LIVE/DEAD™ Fixable Near-IR dead cell stain kit (Invitrogen). Next, cells were labeled with sorting antibody cocktail which contained negative linage (Lin-) CD19 Brilliant Violet (BV)421 (HIB19), CD14-BV421 (63D3) and CD3-BV421 (OKT3), and the following antibodies: CD45- FITC, (HI30), CD94-PerCP-Cy5.5 (DX22), CD127-PE-Cy7 (A019D5), cKit-BV510 (104D2) and NKp44-Alexa Fluor (AF)647 (P44-8) all purchased from Biolegend (San Diego, CA), CRTH2-PE (301109, R&D). ILC3 cells were sorted based on the expression of CD45+Lin-CD127 +CD94-CRTH2-cKit+NKp44+, similarly to Bar-Ephraim et al.
Instrument	Human ILC3s: Cell sorting was performed using a FACSAria II sorter (BD Biosciences, Mountain View, CA, USA).
Software	All data were analyzed using FlowJo (Tree Star).
Cell population abundance	cKit+/- NKp44+ sorted cells ranged between 0.3% to 2.9% from total amount of sorted cells. The amount of sorted cells ranged between 100,000 to 216,000 per sort.
Gating strategy	ILC3s were sorted based on the expression of CD45+Lin-CD127+CD94-CRTH2-cKit+NKp44+

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.