High-resolution promoter interaction analysis in Type 3 Innate Lymphoid Cells implicates Batten Disease gene *CLN3* in Crohn's Disease aetiology

Valeriya Malysheva[1,2,3,4,¶,§], Helen Ray-Jones[1,2,3,¶], Tareian A. Cazares[5,*], Owen Clay[6,*,#], David Ohayon[6,*], Pavel Artemov[1,2,*], Joseph A. Wayman[7], Monica Della Rosa[1,2], Carmen Petitjean[1,2,#], Clarissa Booth[8], Joseph I.J. Ellaway[1,2,#], William R. Orchard[1,9,#], Xiaoting Chen[6], Sreeja Parameswaran[6], Takashi Nagano[10,#], Peter Fraser[11], Stefan Schoenfelder[10], Matthew T. Weirauch[6,12,13,14] Leah C. Kottyan[6,12], David F. Smith[12,15,16], Nick Powell[17], Jill Weimer[8], Chris Wallace[18,19^], Emily R. Miraldi[7,12,14,20^], Stephen Waggoner[5,6,12^] and Mikhail Spivakov[1,2,^,§]

[1] Functional Gene Control Group, Epigenetics Section, MRC London Institute of Medical Sciences, London W12 0NN, UK
[2] Institute of Clinical Sciences, Faculty of Medicine, Imperial College, London, UK
[3] VIB, Center for Molecular Neurology, Antwerp, 2610, Belgium
[4] Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Antwerp 2000, Belgium
[5] Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[6] Division of Human Genetics and Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[7] Molecular and Developmental Biology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[8] Pediatrics and Rare Diseases Group, Sanford Research, Sioux Falls, SD 57104, USA
[9] University of Cambridge, Cambridge, UK
[10] The Babraham Institute, Cambridge, CB22 3AT, UK
[11] Florida State University, Department of Biological Sciences, Florida State University, Tallahassee, FL 32306-4295, USA
[12] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[13] Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA
[14] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA.
[15] Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[16] Divisions of Pediatric Otolaryngology and Pulmonary & Sleep Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA.
[17] Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, UK
[18] MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
[19] Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, CB2 0AW
[20] Department of Electrical Engineering and Computer Science, University of Cincinnati, OH 45221, USA

[¶] Joint first authors * Joint second authors ^ These authors jointly supervised the work
[§] Corresponding authors: V.M. (valeriya.malysheva@vib.be) & M.S. (mikhail.spivakov@lms.mrc.ac.uk)
Present addresses: O.C: Wake Forest School of Medicine, Section of Pediatric Rheumatology, Department of Pediatrics. C.P.: British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge UK; Heart and Lung Research Institute, University of Cambridge, Cambridge UK. J.E.: European Bioinformatics Institute, Hinxton, CB10 1SD, UK; T.N.: Laboratory for Nuclear Dynamics, Institute for Protein Research, Osaka University, Osaka 565-0871, Japan; Institute of Medical Science, University of Tokyo, Tokyo, Japan; W.R.O.: Cancer Research UK Cambridge Research Institute, Cambridge CB2 0RE, UK.

Abstract

Innate lymphoid cells (ILCs) are rare tissue-resident “helper” lymphocytes that do not express diversified antigen receptors. Type 3 ILCs (ILC3s) are an important class of these cells enriched in the respiratory and intestinal mucosa, where they regulate inflammation and mucosal homeostasis. To gain insight into the cis-regulatory circuitries underlying ILC3 function, we used high-resolution Capture Hi-C to profile promoter-anchored chromosomal contacts in human primary ILC3s. Combining significant interaction detection with the Activity-By-Contact approach adapted to Capture Hi-C, we reveal a multitude of contacts between promoters and distal regulatory elements and obtain evidence for distinct regulatory wiring of alternative promoters. We find that promoter-interacting regions in ILC3s are enriched for genetic variants associated with multiple immune diseases. Focusing on Crohn’s disease (CD), in which ILC3s are established mediators, we used a Bayesian approach that incorporates multivariate fine-mapping to link CD-associated genetic variants with putative target genes. We identify known and previously unimplicated genes in conferring genetic risk of CD through activity in ILC3s. This includes the C*LN3* gene that is mutated in most cases of the neurodegenerative disorder Batten disease. Using *Cln3* mutant mice, we show that CLN3 is a putative negative regulator of IL-17 production in an inflammatory subset of ILC3s. This finding suggests a functional role for CLN3 in ILC3 biology, with mechanistic implications for Crohn’s and Batten diseases.

Introduction

Innate lymphoid cells (ILCs) play crucial roles in inflammation and immunity, as well as in tissue development and homeostasis[1,2]. These cells develop from common lymphoid progenitors and share many common features with T lymphocytes, but do not express rearranged T cell receptors[3]. Therefore, rather than acting as part of adaptive immunity, ILCs react to cytokines and pathogens from the environment, responding with regulatory cytokine production and immunomodulatory activity[4,5].

Three main types of ILCs have been identified based on their cytokine profile and the transcription factors regulating their development and function[2,3]. The first group includes tissue-resident ILC1s that play a role in immune defence against viruses and some bacteria[6,7]. The second group is represented by ILC2s that regulate airway and skin inflammatory responses and are implicated in disorders such as asthma and atopic dermatitis[6]. Finally, the third group includes lymphocyte tissue-inducer cells that are crucial for lymph node development, as well as ILC3s that are involved in host defence and homeostasis of epithelial barriers[2–4]. The ILC3 population is distributed in tissues across the body, including the gut, where they are responsible for mucosal homeostasis and barrier integrity[8]. ILC3-produced cytokines, such as IL-22, promote epithelial cells to renew and to release antimicrobial peptides that protect against invading pathogens[9]. However, overexpression of gut ILC3-derived cytokines, such as IL-22 and IL-17, has been associated with the establishment or exacerbation of Crohn's disease (CD)[10–12].

Immune disorders, including CD, are known to have a significant genetic component, with genome-wide association studies (GWAS) identifying hundreds of disease susceptibility variants for these conditions. Given the importance of ILCs in immune control, it is highly plausible that some of these variants affect ILC function. However, as most GWAS variants are non-coding and these studies are by design cell-type-agnostic, identifying causal genes and cell types implicated by GWAS variants is often challenging.

GWAS variants are strongly enriched at transcriptional enhancers[13–15], and therefore cell type-specific maps of active enhancers and enhancer-promoter connections provide important clues for the functional interpretation of GWAS findings[16,17]. Recent studies have mapped ILC enhancers by the assay for transposase-accessible chromatin (ATAC-seq) and chromatin immunoprecipitation (ChIP-seq) for the H3K27ac histone mark, identifying putative key regulators of ILC identity and their downstream targets based on proximal gene assignment[18–22]. Enhancers often localise large distances (up to megabases) away from their target gene promoters, physically contacting them in the 3D dimensional space of the nucleus in a cell-type-specific manner. Therefore, robust and sensitive identification of enhancer-promoter contacts requires the analysis of chromatin architecture.

Chromosome conformation capture assays such as Hi-C, which are based on the proximity ligation of cross-linked, digested chromatin, provide powerful tools for connecting enhancers and GWAS variants with their target genes[23,24]. The conventional Hi-C technique theoretically allows the detection of all pairwise chromosomal contacts across the genome. However, the complexity of the resulting sequencing libraries requires extremely high sequencing coverage to achieve the sensitivity and resolution required for detection of specific enhancer-promoter

contacts. This challenge can be addressed by techniques such as Capture Hi-C that selectively enrich Hi-C material for contacts involving, at one end, regions of interest such as gene promoters[25–28]. Over the last years, we and others have demonstrated the power of Promoter Capture Hi-C (PCHi-C) in determining transcriptional regulatory circuitries and in linking enhancers and disease-associated genetic variants with putative target genes[29–35]. In a foundational study by Javierre et al., we applied this approach to 17 abundant human primary blood cell types and developed COGS (Capture Hi-C Omnibus Gene Score), a Bayesian approach for prioritisation of GWAS target genes using statistical fine-mapping and PCHi-C data[29,30]. Results from Javierre et al. have been incorporated into major variant-to-gene resources, including OpenTargets Genetics and Priority Index. However, the PCHi-C protocol used in Javierre et al. required dozens of millions of cells for analysis, precluding the analysis of rare cell types.

Here, we address this limitation by using a high-resolution and efficient PCHi-C protocol to profile the *cis*-regulatory wiring of ILC3s isolated from human tonsils[29]. We detect promoter-enhancer contacts in PCHi-C data using a combination of our established statistical interaction-calling methodology (CHiCAGO)[36,37] and a PCHi-C-adapted Activity-by-Contact (ABC)[13,38] approach devised in this study. We develop a modified PCHi-C-aware GWAS gene prioritisation algorithm that incorporates multivariate statistical fine-mapping (multiCOGS) and use it to prioritise known and novel genes for CD through chromatin contacts in ILC3s, including the *CLN3* gene, mutations in which underpin ~80% cases of the neurodegenerative disorder Batten disease[39,40]. We validate CLN3 function in ILC3s experimentally using *Cln3* mutant mice, proposing this protein as a negative regulator of IL-17 production. Furthermore, ILC3 treatment with the activating cytokine TL1A results in downregulation of *Cln3* expression. These data have mechanistic implications for both Crohn's and CLN3 Batten diseases.

Results

A compendium of promoter contacts in ILC3s

To profile promoter-anchored chromosomal contacts in type 3 Innate Lymphoid cells (ILC3s), we employed a *DpnII*-based PCHi-C protocol on ILC3s extracted from human tonsils (**Fig. 1A**). Significant promoter contacts were detected with CHiCAGO[36] at a single-fragment resolution, as well as after pooling the 'other end' fragments into ~5kb bins, while leaving the baited promoter-containing fragment unbinned[37]. Using this approach, we detected 31,003 contacts between promoters and promoter-interacting regions (PIRs) at a single-fragment resolution and 58,632 contacts in 5kb bins (**Fig. 1B**; **Tables S1-S2**). Binning resulted in the detection of longer-range contacts, as we reported previously in other cell types[37] (**Fig. 1C, D**). A joint UMAP analysis[41] of ILC3 promoter interaction profiles with those detected in 17 abundant blood cell types using *HindIII*-based PCHi-C segregated ILC3s with other lymphoid cell types, consistent with the notion that patterns of promoter interactions reflect the cells' lineage history[29] (**Fig. 1E**; see Methods).

The usage of alternative TSSs was recently reported to be a major driver of gene isoform choice[42], but little is known about the role that chromatin organisation

plays in this process. The increased resolution afforded by using *DpnII* in Hi-C library generation enabled capturing alternative TSSs for 6,789 genes located on separate *DpnII* fragments (**Fig. 2A, B;** hereafter referred to as "alternative promoters"). We looked for shared and independent chromatin interactions among alternative promoters, focusing on 1,661 genes whose 3,997 alternative baited promoters had appreciable signal coverage (CHiCAGO score >= 5; read count >= 5).

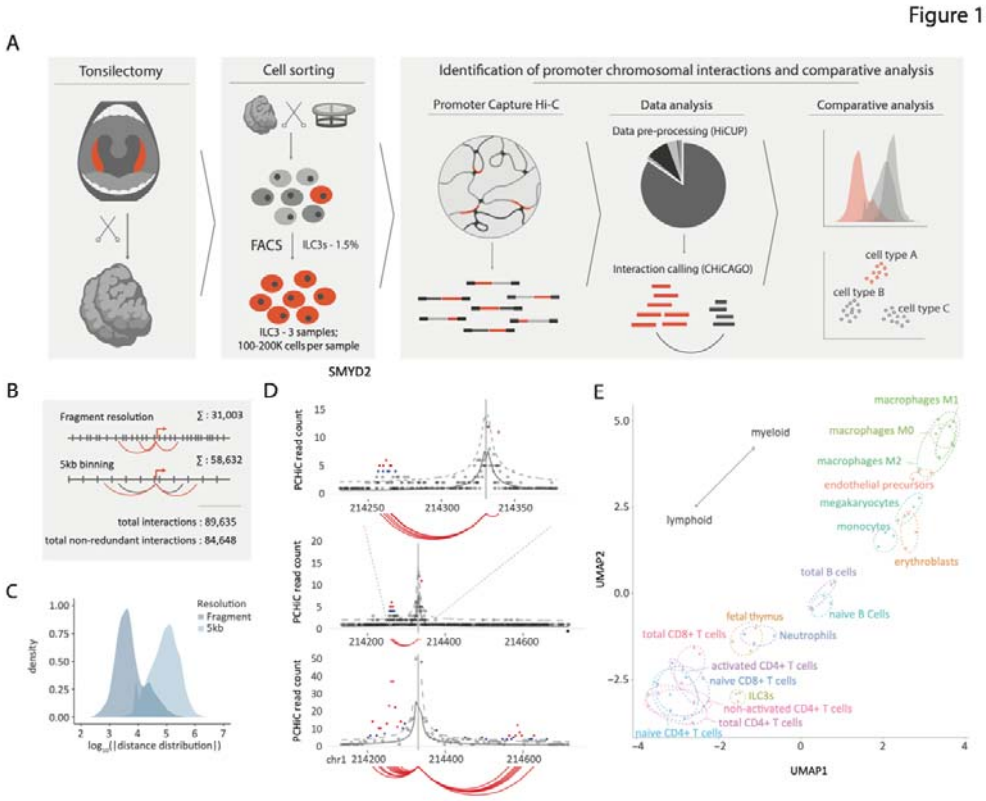


Figure 1. Compendium of promoter-enhancer interactions in ILC3s. A. Scheme of the study from cell extraction to pre-processing and filtering with HiCUP (pie chart), formal characterisation of distal chromatin interactions (calling with CHiCAGO) and comparative analysis (schematic representation of distance comparison and principal component analysis). **B.** Chromosomal interaction binning strategy. The analysis is done in two modes: fragment resolution (no binning) and 5kb binning. The baited fragment remains unbinned, i.e., at fragment resolution, in the 5kb binning mode. Red arcs depict newly detected interactions, grey arcs depict "duplicated" interactions. The numbers of significant interactions are specified for each mode individually and across both modes, accounting for "redundant" interactions (see Methods). **C.** Comparison of promoter-PIR distance distributions for PIRs detected at fragment resolution and 5kb binning resolution. **D**. Example of chromosomal interactions for the *SMYD2* gene at fragment resolution and 5kb binning resolution. The inset shows a zoomed in view at the promoter interactions at fragment resolution. **E.** UMAP plot showing the positioning of ILC3 promoter interaction patterns with respect to those detected in 17 blood cell types and tissues from Javierre et al. The arrow indicates the lymphoid vs myeloid axis.

Overall, 53% of PIRs were shared across all alternative promoters for a given gene, while 19% were shared across some of them and 28% were unique to a single promoter (referred to as "shared", "partially shared" and "distinct", respectively in **Fig. 2C, D**). The three alternative promoters of the *INPP4B* gene provide examples of the multiple degrees of contact sharing across its 14 PIRs included in the analysis (**Fig. 2E**). Each alternative *INPP4B* promoter shared at least two PIRs with another

promoter, while three PIRs were shared between all three promoters. However, each promoter also had at least one distinct interaction (**Fig. 2E**). Since contact frequency generally decreases with interaction distance, we next focused on the short interaction range (<= 50kb), where the sequencing coverage of contacts is the highest, providing for the highest sensitivity of the analysis. “Distinct” PIRs were still abundant (22%, 79/363) among this subset of contacts (**Fig. S1A, B**; see **Fig. S1C** for an example), confirming our general observations.

Overall, our PCHi-C analysis provides a high-resolution compendium of promoter contacts in ILC3s, including divergent contacts across alternative promoters.

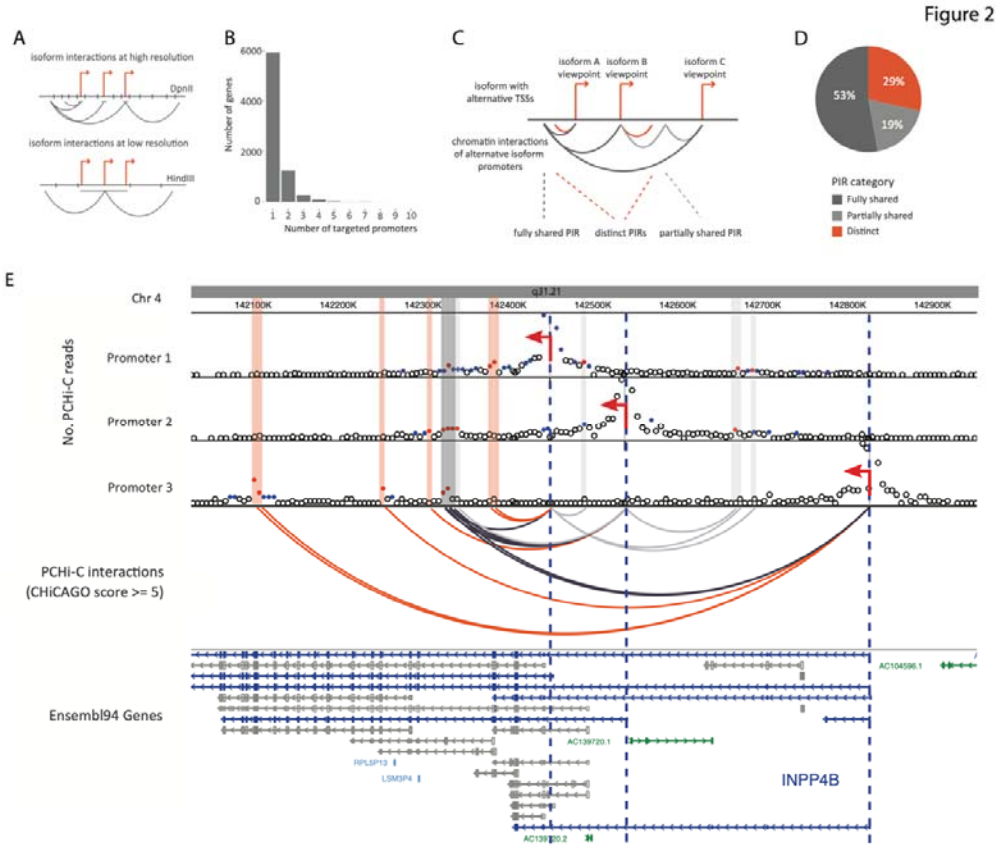


Figure 2. Analysis of chromosomal interactions of alternative promoters. A. PCHi-C at increased resolution allows for the detection of interactions at alternative promoters. **B.** Number of isoforms with alternative promoters targeted for capture in the current study. **C.** Scheme representing the classification of PIRs: fully shared, partially shared and distinct. **D.** Proportion of PIRs falling into each category across all targeted genes. **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 associated transcripts driven by these promoters (Ensembl 94 genes) are shown in blue, whilst non-associated transcripts are shown in grey. Other shown transcripts are for processed pseudogenes (light blue) and lincRNAs (green). PIRs are categorised as fully shared between alternative promoters (dark grey arcs), partially shared (light grey arcs) or distinct (red arcs).

Two complementary analysis approaches link ILC3 enhancers with target genes

As expected, at both fragment and 5kb resolution ILC3 PIRs were enriched for markers of active and/or accessible enhancers (ATAC, H3K27ac) and active transcription (H3K4me3), based on public data in this cell type[20] ("active PIRs", **Fig. 3A**). Therefore, for downstream analysis we decided to focus on a PIR set merged across resolutions, resulting in a total of 18,231 active and/or open PIRs (Methods). To assess the cell-type specificity of these signals, we turned to data from the BLUEPRINT consortium that profiled the chromatin annotations in 88 blood cell types (not including ILCs)[43]. Nearly half of all active and/or accessible ILC3s PIRs (47.5%, 8,658/18,231) overlapped with CTCF-binding signals in at least one of the BLUEPRINT-profiled cell types (**Fig. 3B**), consistent with the key role of CTCF in 3D chromosomal organisation. However, only 3% (636/18,231) contained active enhancer annotations[44] (**Fig. 3B**), while nearly 20% of PIRs (3,425/18,231) did not have any functional annotations in the BLUEPRINT data. In contrast, most of the active and/or accessible PIRs had detectable promoter contacts in the 17 abundant blood cell types profiled in Javierre et al. (~80.4%, 12,409/15,435 at *HindIII* resolution; see Methods). Yet, ~60% of the active PIRs (9,054/15,435) contacted the same gene promoters in both ILC3s and other blood cells, while the rest engaged in cell-type-specific promoter contacts **(Table S3)**. This suggests that patterns of promoter-enhancer contacts are more preserved across related lineages compared with enhancer activity *in cis*, consistent with previous observations[45]. We then probed the relationship between enhancer-promoter connectivity and gene expression. For this, we integrated the promoter-enhancer links with public single-cell gene expression data (RNA-seq) in ILC3s[46]. Consistent with previous observations in other cell types[29] we observed a significant positive correlation between mean gene expression and the number of epigenetic features at PIRs (**Fig. 3C**).

To further increase the sensitivity of detecting functional promoter-enhancer chromosomal interactions from PCHi-C data, we adapted the Activity-by-Contact approach[38] (originally developed for Hi-C) for this assay. In contrast to CHiCAGO detection based on identifying contacts whose frequency significantly exceeds a distance-dependent background, ABC instead considers any observed contact frequency between an enhancer and a promoter as potentially functionally meaningful. In addition, while CHiCAGO scores are independent of enhancer activity levels at the PIRs, ABC incorporates both contact frequency and enhancer activity into the final metric ("ABC score")[38] (**Fig.3D).** To estimate contact frequencies robustly in Capture Hi-C data, we took advantage of CHiCAGO-inferred distance decay and normalisation parameters and validated the algorithm initially using public enhancer perturbation data in K562 cells (see Methods and **Fig. S2**).

We applied the CHi-C-adapted ABC approach to the ILC3 PCHi-C data, detecting 12,386 enhancer-promoter pairs. Enhancer-promoter links detected with this approach replicated the positive association between the number of connected enhancers and gene expression observed with CHiCAGO-detected PIRs above **(Fig. 3E and Fig. S3)**. Notably, the ABC-detected pairs showed only a limited overlap with the CHiCAGO-detected promoter-enhancer contacts, with 9.4% of ABC pairs identified as significant by CHiCAGO (**Fig. 3F**). This limited overlap is consistent with the largely distinct distance distributions of contacts detected using each of these

methods (**Fig. 3G**). As these two approaches appear to be largely complementary (see **Discussion**), we have combined the CHiCAGO and ABC results for downstream analyses, generating a resource presented in **Table S4**.

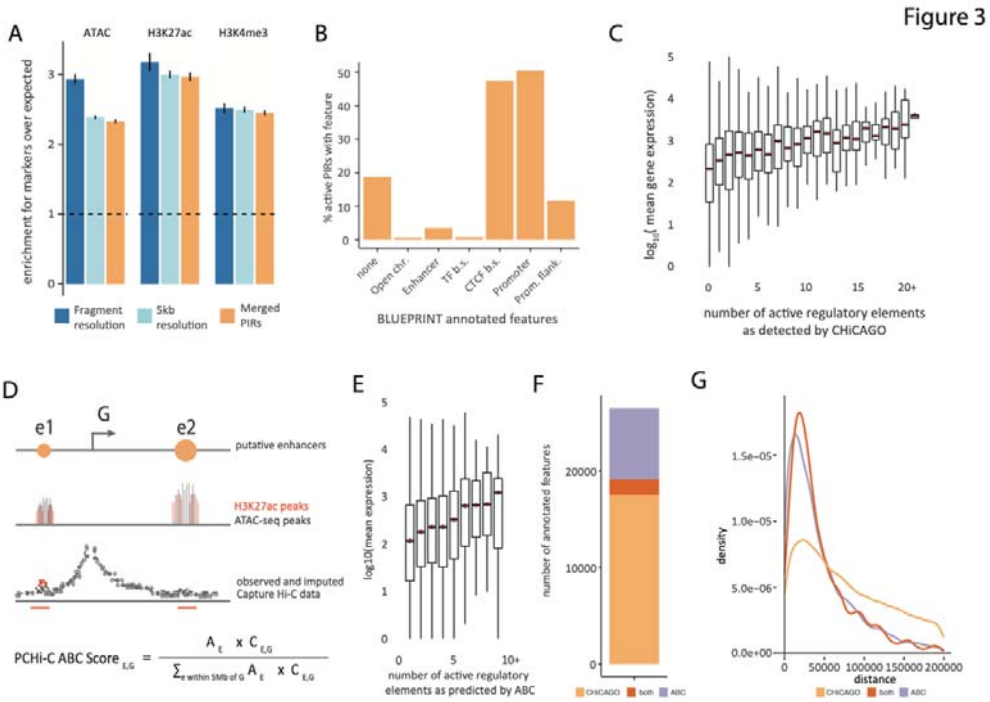


Figure 3. Using CHiCAGO and PCHi-C adapted ABC for linking enhancers with their target genes. A. Enrichment of PIRs for the markers of active enhancers and promoters (H3K27ac and H3K4me3) and accessible chromatin (ATAC). **B.** Characterisation of active and/or open ILC3 PIRs at merged fragment and 5kb resolution as per BLUEPRINT annotations. **C.** Correlation between gene expression and number of regulatory elements identified in CHiCAGO PIRs. **D.** Scheme depicting the adaptation of the ABC model for use with Capture Hi-C data. **D.** Number and overlap of interactions identified with CHiCAGO and ABC. **E.** Correlation between gene expression and number of regulatory elements identified in ABC-predicted enhancers **F.** Overlap between ABC-predicted enhancers and CHiCAGO-detected active and/or open PIRs. **G.** Distance comparison across CHiCAGO-specific, ABC-specific and shared interactions.

Promoter-interacting regions in ILC3s are enriched for genetic variants associated with multiple immune traits

To reveal the potential roles that GWAS variants have in mediating disease susceptibility in ILC3s, we investigated the enrichment of the detected promoter-enhancer pairs in ILC3s for SNPs associated with 763 human traits and diseases using the RELI algorithm[47] (**Fig. 4A**; see Methods). Briefly, RELI determines significantly enriched overlaps between independent genomic risk loci (in our case, PIRs and ABC enhancers detected from PCHi-C data) and trait-associated variants, in comparison to a null distribution of artificially-created variant sets with similar LD characteristics to the trait-associated variants[47].

Overall, promoter-connected regulatory elements in ILC3s showed a significant enrichment at genetic risk loci associated with 29 human traits (**Fig. 4B**, **Table S5**) These traits were enriched in autoimmune disorders (EFO:0005140;

P=8.27e-05, hypergeometric one-tailed test) and affected a broad array of organs and tissues in which ILC3s are known to reside. These included the gut (Crohn's disease, celiac disease, primary sclerosing cholangitis), airways (asthma), blood (lymphocyte count, monocyte count, platelet volume), joints (systemic lupus erythematosus, rheumatoid arthritis) and pancreas (type 1 diabetes). These results highlight the importance of ILC3s in mediating genetically-determined susceptibility to immune diseases and suggest that targeting the genes and pathways implicated by susceptibility variants in these cells may have therapeutic potential.

Linking GWAS variants with PCHi-C-identified target genes prioritises known and novel candidates for Crohn's Disease

For the remainder of the study, we focused on the genetic mechanisms underpinning the role of ILC3s in Crohn's disease (CD). ILC3s have been strongly implicated in this condition[11,20] and, consistent with this, ILC3 PIRs were enriched ~2.3-fold for CD-associated genetic risk loci compared with background variant sets (**Fig. 4B**).

To identify putative CD-associated genes in ILC3s we initially used our previously published Bayesian prioritisation algorithm, COGS. In brief, COGS provides a single measure of support ("COGS score") for each gene's association with a trait of interest[29,30]. This score is calculated by the statistical fine-mapping of GWAS signals to (i) gene coding regions, (ii) promoters, and (iii) promoter-connected regions (in our case, represented by both CHiCAGO PIRs and ABC enhancers) (**Fig. 4C**).

We ran COGS on a CD GWAS meta-analysis of three patient cohorts by de Lange *et al.*[48] using the compendium of PIRs and CHiC-ABC-detected enhancers in ILC3s and prioritised 50 genes at a COGS score cutoff of 0.5 (**Fig. 4D**; **Table S6**). Many of these genes have roles in immune processes that are known to be dysregulated in inflammatory bowel disease[49–51]. For example, several candidate genes were involved in cytokine signalling (*IL1RL1, LTBR, IL2RA, IFNGR2, TNFSF4*), autophagy (*ATG16L1, GPR65*), and antimicrobial processes in the gut (*PTPN2, IRF8*)[52,53]. The prioritised genes also highlighted the importance of IL-23/Th17 signalling in CD (for example, *RORC, CCR6* and *JAK2*). In addition, COGS prioritised several known transcriptional regulators (*FOS*, *TSC22D1* and *RBPJ*).

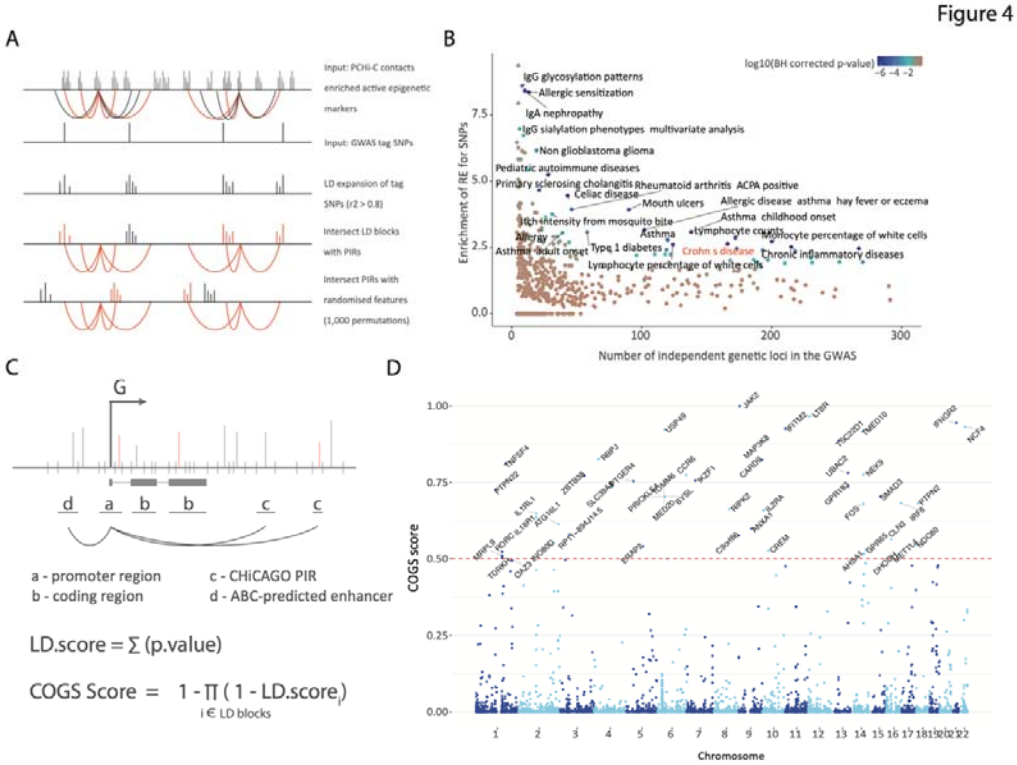


Figure 4. Prioritising candidate genes for Crohn's disease in ILC3 using the COGS pipeline. A. Scheme of RELI algorithm for associating disease phenotype and other traits with the chromosomal interaction profiles. The RELI algorithm analyses the enrichment of trait-specific independent genetic risk loci within the PIRs. **B.** Results of RELI analysis in ILC3s chromosomal interactions across 763 diseases and traits. Traits with enrichment >=2, log10(BH corrected p-value) < 0.01 and the number of independent risk loci in GWAS >=10 are labelled. Crohn's disease trait highlighted in red. **C.** Schematic of the COGS algorithm. **D.** Prioritisation scores for all genes from the COGS analysis of CD GWAS data[48] and promoter interaction data in ILC3s. Genes with COGS scores above 0.5 are labelled.

We assessed the enrichment of annotated biological pathways among the genes with high COGS scores using pre-ranked GSEA (see Methods). Highly ranked genes in standard COGS were significantly enriched in 77 GO terms in hallmark gene sets[54] implicating processes such as lymphocyte differentiation, cytokine production and adaptive immune response (**Fig. S4A**). We also noted an enrichment for cell-cell adhesion, which influences the migration of immune cells into the gut[55], and ERK1/ERK2 (MAPK) signalling, known to be dysregulated in IBD[56].

Overall, this analysis has provided an initial set of credible prioritised CD-associated genes based on ILC3 PCHi-C data using our previously established statistical pipeline.

Multivariate fine-mapping improves the prioritisation of CD-relevant genes

The original version of the COGS algorithm makes a strong assumption of a single causal variant per linkage disequilibrium (LD) block. However, multiple causal variants per LD block are likely present in GWAS data due to a joint contribution of variants from multiple regulatory regions, as well as in some cases of coding variants[57,58]. To address this, we developed multiCOGS, a version of COGS that

overcomes the “single causal variant” assumption by enabling the analysis of data obtained from multivariate fine-mapping algorithms such as SuSiE[59–61] (see Methods).

MultiCOGS analysis produced a markedly increased number of CD-prioritised genes (from 50 to 99) compared with standard COGS on the same GWAS and PCHi-C datasets (**Fig. 5A** and **Table S7**). Examples of newly prioritised genes included the well-characterised immune genes *IL1R1, IL1R2, NFKB1, TNFSF8* and *TYK2,* the latter recently found to harbour rare CD-associated coding variants[62]. In contrast, multiCOGS did not prioritise *JAK2* (multiCOGS score = 0.01, COGS score ~1), due to the shift of the fine-mapped signal from the promoter to a region ~20 kb upstream of it without detectable promoter contacts in multivariate fine-mapping.

GSEA using the genes prioritised by multiCOGS (score>0.5) revealed more than double the number of enriched GO terms compared with standard COGS (77 vs 211), with many of the same biological pathways implicated (top 20 shown in **Fig. S4B**). We then assessed the expression of the prioritised genes in CD patients and in individuals without the disease using data from a transcriptomics IBD database spanning multiple tissues[63]. We found that genes prioritised by multiCOGS were significantly enriched within a set of genes showing differential expression in the rectum between CD patients and healthy donors (padj=0.005) (**Fig. S4C**). In contrast, genes prioritised by standard COGS were not significantly enriched in any differentially expressed gene sets.

Finally, we searched for prior evidence of association of the prioritised genes with CD or IBD using curated gene-to-disease databases, functional studies and a recent exome sequencing study of CD patients (see **Methods**)[62,64–67]. We found that over half of multiCOGS-prioritised genes in ILC3s (56/99) were not previously implicated in these diseases (Table S8). Notable examples of novel CD-associated genes included *RBPJ* (COGS score 0.83), the DNA binding effector of *NOTCH*, which has an important role in gut ILC3 plasticity[68], as well as *CLN3* (COGS score 0.56), the causative gene for a subset of lysosomal storage disorders characterised by progressive neurodegeneration, such as Batten disease[69].

Jointly, these results demonstrate the importance of relaxing the single causal variant assumption for GWAS gene prioritisation based on chromosomal contacts.

The chromatin properties and cell-type specificity of the PIRs driving CD gene prioritisation

We assessed the localisation of fine-mapped CD GWAS signals with the different types of regions considered by multiCOGS, including the PIRs detected in ILC3 PCHi-C and promoter-proximal and coding regions defined on the basis of reference genome annotation. We found that GWAS signals at PIRs were required for the majority of multiCOGS gene prioritisations (70/99). In contrast, only four prioritised genes (*SLAMF8*, *PTPN22*, *IRF6* and *PPP5C*) required signal contribution from fine-mapped SNPs within their coding regions, demonstrating the importance of distal variants and their promoter links for GWAS variant-to-gene assignment.

To gain insight into the putative function of the PIRs containing fine-mapped CD-associated variants, we assessed their patterns of chromatin marks and transcription factor (TF) binding. As TF-binding data for human ILCs are not publicly available and are challenging to generate, we used the Locus Overlap Analysis

(LOLA)[70] tool for this purpose, taking advantage of its in-built database of TF ChIP datasets obtained across multiple cell types, including lymphoblastoid cells. We found that CD

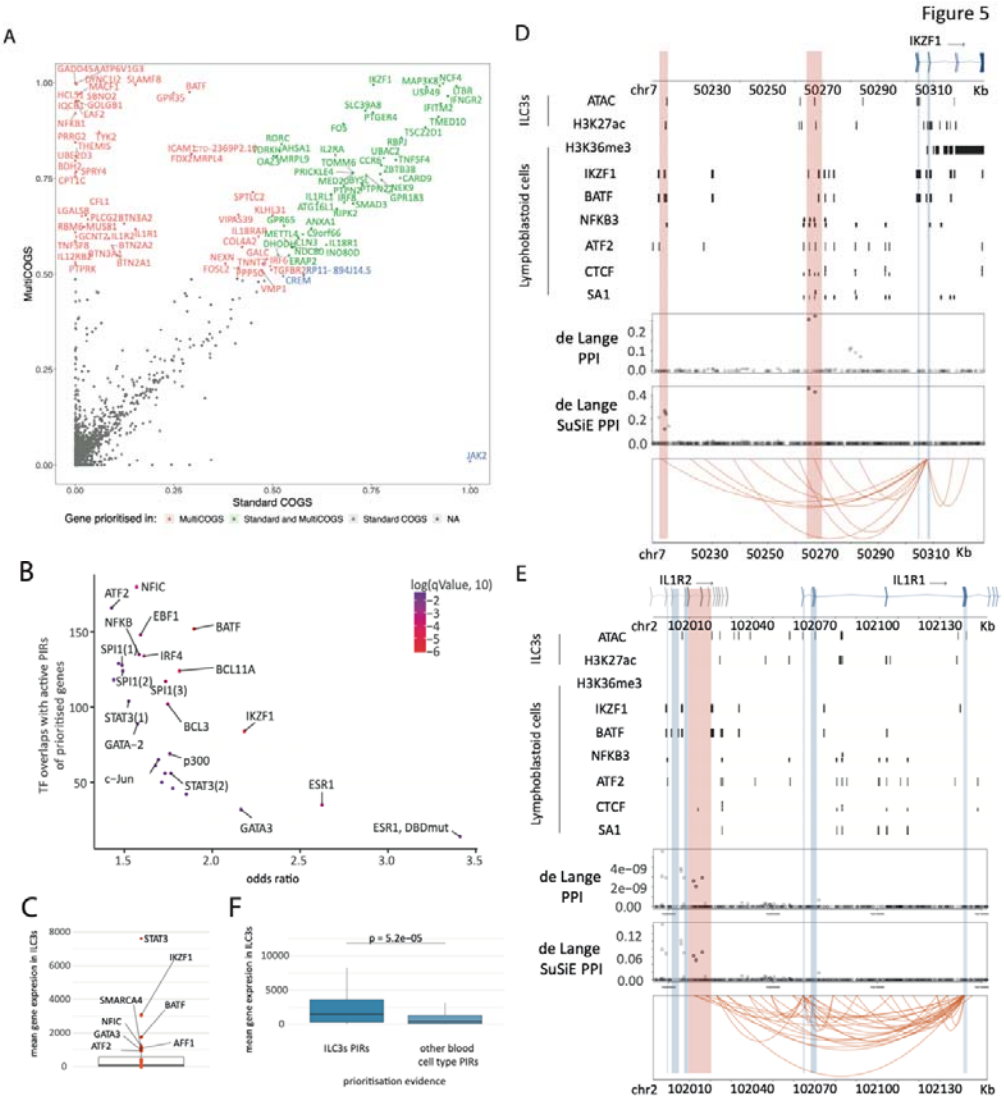


Figure 5. Improving prioritisation of GWAS-associated genes with multivariate fine-mapping. A. Comparison of genes prioritised using standard single-variant COGS and multivariate COGS. **B.** Enrichment analysis for TF binding sites at active PIRs for genes prioritised by COGS vs active PIRs of all genes that entered the COGS analysis. **C.** Expression of TFs enriched at PIRs prioritised genes. **D and E.** Example of genes prioritised by COGS for CD, IKZF1 and IL1R1, with enriched TF binding, standard single-variant COGS and multiCOGS posterior probability profiles. The vertical blue bands highlight the captured loci. The vertical red bands highlight the loci that are harbouring SNPs with high posterior probability of being causal for the observed disease trait. Orange arcs correspond to CHiCAGO interactions at 5kb resolution of IKZF1 (D) and IL1R1 (E).

variant-harbouring PIRs were enriched for the binding sites of TFs previously implicated in inflammatory response and CD (**Fig. 5B**). Several of these TFs, including IKZF1/Ikaros[71], BATF[72] and NFKB3/RELA[73–75] are expressed in ILC3s **(Fig. 5C)** and have established roles in ILC3 biology. For example, the promoter of the

IKZF1 gene contacts two upstream PIRs across ~100kb and ~40kb, both of which harbour fine-mapped CD susceptibility variants (**Fig. 5D**). These PIRs bear the active enhancer mark H3K27ac and ATAC-seq chromatin accessibility signals in ILC3s and, based on data from lymphoblastoid cell lines, recruit multiple TFs, including IKZF1 itself, as well as architectural proteins CTCF and a cohesin complex subunit SA1 potentially mediating their promoter contacts (**Fig. 5D**). Accessible chromatin (detected by ATAC-seq) along with CTCF and SA1 binding was also detected in the variant-harbouring PIR of the *IL1R1* gene (**Fig. 5E**).

Since promoter contacts can vary significantly between cell types, we assessed the cell-type specificity of multiCOGS results for the 70 genes whose prioritisation was driven by ILC3 PIRs. Using the Javierre et al. PCHi-C data in abundant blood cell types[29], we asked whether these genes could also be prioritised based on the PIRs detected in this dataset. Excluding nine genes whose tissue specificity could not be ascertained due to differences between study designs (see **Methods**), we found that over half of the remaining ILC3 PIR-prioritised genes (34/61) could also be prioritised based on Javierre et al. PIRs (Table S9). This result is consistent with the involvement of multiple immune cell types in mediating CD risk[76,77]. In contrast, 27/61 genes could only be prioritised using ILC3 PIRs. These genes included several whose roles are known to affect ILC3 function in the gut, such as *TNFSF4*[78], *RBPJ*[68], *CCR6*[79], *GPR183*[80] and *IL18R1*[11] (Table S9).

Ten of the 27 genes prioritised based on ILC3-specific PIRs, but not the PIRs from Javierre et al. (including *THEMIS*, *CCR6*, *GPR183*, *IL1RL1* and others) had been previously implicated in CD (shown in bold in Table S9). The remaining 17/27 were newly implicated genes, including, among others, *MACF1*, whose product regulates tight junctions and has been linked to experimental colitis susceptibility in mice[81], and *BTN3A2*, encoding a CD277 isoform that is expressed by Natural Killer cells, in which it is thought to inhibit cytokine-induced IFN-γ production[82]. Overall, the genes prioritised based on ILC3-specific PIRs had a considerably higher expression in ILC3s than those prioritised based on PIRs from other blood cells (**Fig. 5F**). Jointly, these results support the notion that PCHi-C-based prioritisation pinpoints potentially causal cell types and highlight the utility of cell-type-specific chromosome conformation data for GWAS target gene prioritisation.

PCHi-C and eQTL data prioritise Batten disease gene *CLN3* as a causal candidate gene for CD

Both COGS and multiCOGS approaches have prioritised the Batten disease gene *CLN3* as a potentially causal candidate underpinning Crohn's disease risk. Close inspection of the SuSIE fine-mapped CD GWAS locus driving this gene's prioritisation reveals a credible set of variants that overlaps a *CLN3* PIR in ILC3s located 14.2 kbps downstream from *CLN3's* main TSS and a fragment adjacent to this gene's internal promoter (**Fig. 6A**). Surprisingly, we found that these regions are devoid of chromatin accessibility and enhancer activity signals in ILC3s, as well as in all other cell types included in the Ensembl Regulatory Build database (**Fig. 6A**). At the same time, data from lymphoblastoid cell lines[83] showed enrichment for the H3K36me3 mark typically associated with transcriptional elongation[84] and facultative heterochromatin[85] (**Fig. 6A**).

To obtain complementary evidence for the potential regulatory function of this locus, we queried the OpenTargets database[86] for possible colocalisation between the CD susceptibility locus and genetic variants (eQTLs) associated with *CLN3* expression. We found that CD risk and *CLN3* expression were likely to have a joint causal genetic signal (posterior probability >= 0.8, as determined by coloc[87] and reported in OpenTargets) in whole blood[88,89], thyroid[89], small intestine[89], brain cerebellum[89] and monocytes.[90,91]

Jointly, these results propose *CLN3* as a causal candidate gene for CD susceptibility, implicating ILC3 and potentially other cell types.

Cln3 is involved in the inflammatory function of mouse ILC3s

CLN3 has not been previously implicated in immune-mediated conditions and its function in immune cells is hitherto poorly understood, which prompted us to explore its function in ILC3s in the context of CD *in vivo*. To this end, we took advantage of a mouse model bearing a deletion of exons 7 and 8 (*Cln3*$^{\Delta ex7/8}$) relevant to clinical mutations in *CLN3* in Batten disease[92]. We compared the number of ILC3s present in the small intestine lamina propria of *Cln3*$^{\Delta ex7/8}$ and wild-type control mice (**Fig. 6B**). Loss of *Cln3* did not affect the frequencies of ILC3 in the small intestine (**Fig. 6C**). However, the CCR6^{neg} subset of intestinal ILC3s from *Cln3*$^{\Delta ex7/8}$ mice spontaneously secreted more IL-17, a key pro-inflammatory cytokine in CD pathophysiology, than ILC3s from control mice (**Fig. 6D**). Elevated IL-17 production was not observed in the CCR6^{+} lymphoid tissue inducer subset of ILC3 (**Fig. 6E**).

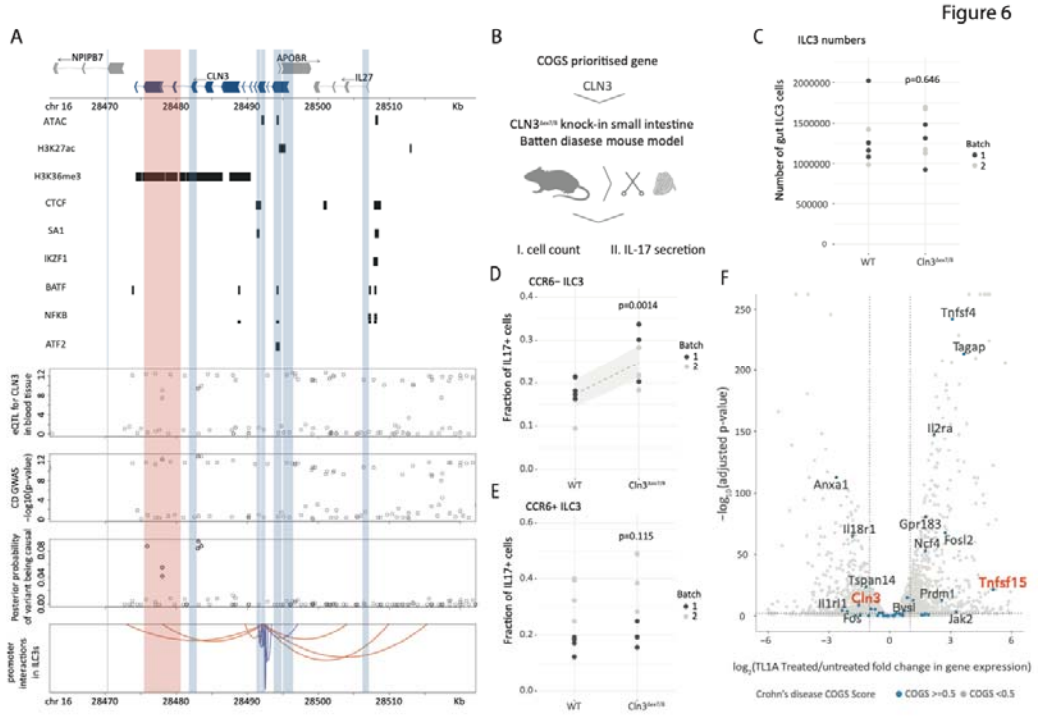


Figure 6. In vivo validation of functional role of CLN3 in ILC3s. A. Interaction profile of CLN3 with enriched TF binding, blood tissue eQTL, CD GWAS and SuSiE posterior probability profiles. The vertical blue bands highlight the captured loci. The vertical red bands highlight the loci that are harbouring SNPs

with high posterior probability of being causal for the observed disease trait. Orange arcs correspond to CHiCAGO interactions at 5kb resolution. Purple arcs correspond to CHiCAGO interactions at fragment resolution. Interactions of all baited regions in the locus are shown. **B.** Scheme of the validation strategy**. C.** The effect of *Cln3*$^{\Delta ex7/8}$on the number of ILC3s in mouse small intestinal lamina propria. **D** and **E.** The effect of *Cln3*$^{\Delta ex7/8}$ on the fraction of IL-17^{+} CCR6^{neg} ILC3s (panel D) and CCR6^{+} ILC3s (panel E) determined by flow cytometry, with significance of effect ascertained using beta regression (see Methods). **F.** Differential gene expression (RNA-seq) analysis in TL1A-treated vs untreated small intestine ILC3 from wild-type mice. Genes prioritised by COGS are highlighted in blue with *Tnfsf15* (TL1A) and *Cln3* highlighted in bold red.

TL1A (*TNFSF15*) is a TNF family cytokine that is a key activator of ILC3 proliferation and cytokine production[93], and is a promising drug target in inflammatory bowel disease[50,94]. We therefore asked whether *Cln3* expression could be controlled by TL1A. For this, we took advantage of a previously published RNA-seq dataset comparing gene expression in TL1A-treated and untreated wild-type mouse ILC3s[95]. Differential analysis of these data showed that TL1A treatment results in a significant reduction of *Cln3* expression (~3-fold, padj = 1.4 x 10^{-9}; **Fig. 6D**). Strikingly, approximately one third of the mouse homologs of the genes prioritised in the CD GWAS using multiCOGS (34/99 at a score cutoff of 0.5) were differentially expressed in this dataset alongside *Cln3* (padj < 0.05; **Table S10**). These included, for example, another TNF family cytokine *TNFSF4* (padj = 7.5 x 10^{-243}), which already has a promising functional link with ILC3s and gut inflammation[78]. Finally, *TL1A* itself has an elevated multiCOGS score of ~0.4 for CD and additionally contains a high-confidence fine-mapped CD-associated variant rs10114470 (SuSIE PIP=0.56) located in its 3' UTR that is not considered by the COGS pipeline.

Taken together, our results suggest that *Cln3* acts downstream of TL1A as an inhibitor of ILC3 inflammatory function and point to a critical role for the TL1A axis in mediating the effects of genetic variation on CD susceptibility in these cells.

Discussion

In this study we presented high-resolution promoter interaction profiling in Type 3 innate lymphoid cells (ILC3s), revealing tens of thousands of promoter-enhancer contacts and GWAS variants associated with multiple immune diseases. Our analysis adds these highly clinically relevant cells to the ever-expanding array of cell types with available promoter interactome maps, including the 17 abundant blood cell types that we profiled previously using high-coverage PCHi-C at a six-cutter enzyme (*HindIII)* resolution[29]. ILC3s are a relatively rare cell type that cannot be easily expanded *in vivo*, which makes their chromosomal interaction profiling challenging. Indeed, this problem precluded ILC3 profiling by standard Hi-C alongside their Type 2 ILC counterparts in a very recent mouse study[96]. Robust Capture Hi-C profiling typically requires even higher cell numbers. Our efficient PCHi-C protocol[97] and the use of a four-cutter enzyme (*DpnII)* have enabled a higher-resolution analysis of human ILC3s in this study, detecting a multitude of promoter contacts and revealing divergent regulation from alternative TSSs.

Unlike in our previous studies, here we have taken advantage of two conceptually different computational analysis strategies for detecting promoter contacts from Capture Hi-C data. The first one is based on our established CHiCAGO pipeline to detect 'significant contacts' - i.e., those whose frequency

significantly exceeds the expectation at a given distance and technical noise levels. The second strategy adapts the "Activity-By-Contact" (ABC) method[13,38] to Capture Hi-C data, which, in contrast, considers the raw contact frequency rather than its significance. As expected from this conceptual difference, ABC prioritises shorter-range contacts compared with CHiCAGO (**Fig. S2**), resulting in the largely non-overlapping sets of identified contacts and GWAS-prioritised genes. Importantly, we found that the longer-range contacts detected using CHiCAGO were also enriched for active enhancers and drove the majority of our identified disease associations. From the practical point of view therefore, these two approaches are largely complementary, and their combined use is warranted. Mechanistically, this suggests that at short linear distances, the background frequencies of promoter-enhancer contacts arising from constrained Brownian motion are sufficient for the functional interaction between these regions. In contrast, at longer ranges, additional factors (e.g., cohesin-mediated loops) are likely required to facilitate the statistically unusual contact frequencies and enable functional interactions.

Human ILC3s in our analysis were derived from tonsillectomy material, but their regulatory elements showed an 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[98,99]. Furthermore, ILC3s from regularly inflamed tonsils were found to have a closer cytokine profile to mucosal-resident ILC3 populations than ILC3s from resting lymph nodes or peripheral blood[100]. Focused studies in relevant physiological contexts and disease models will further establish the role of ILC3s in mediating the effects of genetic variation.

We found a strong enrichment for CD-associated SNPs within the ILC3 PIRs, consistent with the recent findings showing that superenhancers specific to ILC3 or Th17 cells, rather than to ILC1 or Th1 cells, preferentially contain CD-associated variants[20]. Using our multiCOGS strategy that integrates GWAS data processed with multivariate statistical fine-mapping with information on enhancer-promoter links from PCHi-C, we have prioritised a total of 99 genes in ILC3s. Notably, the number of multiCOGS-prioritised genes has increased considerably compared with the results obtained with our previously developed COGS pipeline[29,30] that assumes no more than a single causal variant per linkage disequilibrium (LD) block. At the molecular level, the increased recall of prioritised genes reflects the fact that the same LD block often contains multiple regulatory elements (including promoter-proximal and distal enhancers), variants within each of which may have largely independent effects from those in other elements regulating the same gene[45,57] and from within the gene's coding regions[58]. These results reinforce the notion that the assumption of a single causal variant per LD block used by many established GWAS analysis methods (particularly those based on summary data) is unnecessarily restrictive and may miss key genetic mechanisms underpinning disease processes.

The *CLN3* gene prioritised by our analysis for CD underpins the majority of cases of Batten disease, which sparked our interest to probe this association further. Intriguingly, we found that the region harbouring the fine-mapped CD susceptibility variants in the *CLN3* locus lacks active chromatin signals in ILC3s, as well as in all other cell types with available Ensembl Regulatory Build annotations. It is possible that this region contains gene regulatory module(s) that only become active under

specific inflammatory conditions and/or only in a subset of cells, which could be missed by the analysis of the pooled, steady-state cell populations. With relevance to this hypothesis, the H3K36me3 histone modification covering this region in lymphoblastoid cell lines has recently been proposed as a mark of enhancers that are 'poised' for rapid activation[101]. However, variants in this region may also exert gene control via mechanisms other than enhancer activation. For example, several of the fine-mapped CD-associated variants in the *CLN3* locus have been associated with alternative polyadenylation of the *CLN3* transcript's 3'UTR (apaQTL) in multiple tissues[102,103]. Alternative polyadenylation affects the length of the 3'UTR and thus the mRNA stability and translation efficiency and is emerging as an important potential function for genetic variants conferring complex disease risk[103]. In addition, recent analysis has suggested that *CLN3* is subject to splicing-dependent transcriptional activation[104], adding to the range of possible regulatory mechanisms of variants in this locus. The regulatory complexity of this locus is further augmented by the fact that it is detected as an eQTL for both *CLN3* and a number of neighbouring genes in several other cell types and tissues. For example, in monocytes, this locus is also an eQTL for a known CD target *IL27,* with an opposite direction of allelic effect and a lower significance of association compared with *CLN3* itself[91,105]*.* Elucidating the exact molecular effects of CD susceptibility loci in ILC3s and their cell-type specificity requires perturbation experiments that are currently challenging to perform in this rare cell type.

Structurally, *CLN3* encodes a transmembrane lysosome protein whose C- and N-ends localise to the cytoplasm[106]. While CLN3 function remains not fully elucidated, recent studies have demonstrated that it is required for glycerophospholipid catabolism[107,108]. Glycerophospholipids are key structural components of cell membranes, but also play a variety of regulatory roles, including in innate immunity[109]. It was shown previously that knockdown of *Cln3* in mouse monocytes interfered with LPS-mediated release of the inflammatory cytokine IL-6[109]. Our experiments in a *Cln3*-mutant mouse model suggest this gene as a negative regulator of IL-17 expression in the inflammatory $CCR6^{neg}$ subset of ILC3s, but not in the lymphoid tissue inducer-like $CCR6^{+}$ ILC3 subset. We further show that this gene, alongside several other of our prioritised CD-associated candidates, potentially acts downstream of TL1A (*Tnfsf15*), an important regulator of ILC3 inflammatory function. How exactly CLN3 acts within this pathway at the molecular level to regulate IL-17 production remains the subject of future research, but this result reinforces the importance of the TL1A axis in both CD and ILC3 biology.

While our analysis was necessarily focused on common diseases that are amenable to GWAS methodology, the role of CLN3 in ILC3 biology and its predicted involvement in CD are potentially relevant for the rare Batten disease. This progressive neurodegenerative disease is cytologically characterised by the inclusions of lysosomal storage material observed in neurons. However, such inclusions are also observed outside of the CNS, potentially implicating other cell types in Batten disease aetiology. Neuroinflammation, which may be influenced by the systemic production of autoantibodies, is a known feature of CLN3 Batten disease[110] and is emerging as critical in other neurodegenerative diseases[111]. This raises the possibility that the immune axis of CLN3 function may contribute towards Batten disease aetiology, leading to new questions to be explored in the search for the mechanisms underlying this severe condition.

Methods

Human ILC3 cell isolation

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 with an institutional review board (IRB) approval. Criteria for enrollment in the study included a history of sleep-disordered breathing or recurrent or chronic tonsillitis requiring removal of the tonsillar tissue. Consent was obtained from parents in the perioperative suite on the day of the procedure. 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. Samples were labelled with a de-identified barcode and transferred to the research team for further processing.

Next, tonsils were dissociated into single cell suspension as previously described[112,113]. Briefly, 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. Cell sorting was performed using a FACSAria II sorter (BD Biosciences, Mountain View, CA, USA). Post sorting sorted ILC3 cells were washed with PBS for 5 minutes at 1200rpm, 20°C and then incubated in 100L of 2% formaldehyde (in PBS) for 10 minutes, followed by an addition of 0.125M glycine. Next, cells were centrifuged at 400g for 5 minutes at 4°C, resuspended with cold PBS and centrifuged again at 400g

for 5 minutes at 4°C, supernatant was discarded, and cells were snap frozen in liquid nitrogen and then stored and -80°C prior to PCHi-C analysis.

Promoter Capture Hi-C

Promoter Capture Hi-C was performed as previously described[97]. Cells were fixed in 2% PFA for 10 minutes, lysed in a lysis buffer (30 minutes on ice), and digested with *DpnII* (NEB) overnight at 37C rotating (950rpm). Restriction overhangs were filled-in with Klenow (NEB) using biotin-14-dATP (Jena Bioscience), and ligation was performed in the ligation buffer for 4 hours at 16C (T4 DNA ligase; Life Technologies). After overnight decrosslinking at 65C, the ligated DNA was tagmented to produce fragments of 300-700 bp range. 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). Promoter Capture Hi-C was performed using a custom-design Agilent SureSelect system following the manufacturer's protocol. The PCHi-C libraries were paired-end sequenced (100bp) on an Illumina HiSeq 2500 machine at a sequencing depth of ~400M reads per sample.

PCHi-C data pre-processing and detection of significant interactions

Sequencing data from three ILC3 PCHi-C biological replicates were aligned to the hg38 genome assembly using Bowtie2[114] and quality-controlled using HiCUP[115]. Significant interactions were then detected across the replicates by CHiCAGO[36] as previously described[37] at single *DpnII* fragment resolution and in bins of fragments approximately 5kb in length, with the baited promoter fragments left solitary (unbinned). For CHiCAGO analysis at single-fragment resolution, p-value weights were estimated following our previously described procedure[37] and are listed in **Table S11**; default p-value weights were used for the 5kb analysis. A CHiCAGO score cutoff of >=5 was used for both resolutions. A consensus list of promoter interactions was compiled from non-redundant contacts detected at the fragment and 5kb resolutions.

Integration with *HindIII* Promoter Capture Hi-C data

Our previous PCHi-C study in 17 abundant human primary blood cell types, including both lymphoid and myeloid cells by Javierre et al.[29] was performed using a 6-bp restriction enzyme *HindIII*, unlike the 4-bp cutter enzyme *DpnII* used in the current study. Since restriction fragment size affects the distance distribution of contacts detected in Hi-C-related methods[37,116,117], direct comparison across these two datasets is challenging. To partially address this issue, we pooled the reads in *DpnII*-based ILC3 data into genomic windows corresponding to *HindIII* fragments and re-processed the data with HiCUP using the hg19 genome assembly and *HindIII* parameters. We then identified significant interactions using CHiCAGO[36] with the default *HindIII*-based parameters and integrated them with the significant interactions from the Javierre study[29]. To assess the similarity of promoter-interaction patterns in

ILC3s with the cell types profiled in Javierre et al., we first ran a joint PCA analysis. We noted that PC1 (accounting for <10 % of the variance) clearly segregated the three ILC3 replicates from the remaining cell types, and therefore most likely corresponded to the difference in PCHi-C methods, resolution and sequencing depth. We disregarded PC1, and focused on PC2, PC3 and PC4, accounting for 6.16%, 3.7% and 3.16% of variance across all tissues, respectively (components beyond PC4 accounted for <3.1% of variance each and were disregarded). For visualisation purposes, we combined these three components using the UMAP non-linear dimensionality reduction algorithm implemented in the umap package in R[118], obtaining the plot shown in **Fig 1E.**

Alternative promoter analysis

We used the CHiCAGO results for ILC3 PCHi-C data at 5kb resolution to profile PIR sharing between alternative promoters. First, we identified a set of genes that had more than one baited promoter, with each promoter having at least one significant interaction with a CHiCAGO score of >= 5 with >= 5 reads. We defined fully shared PIRs as those that interacted with all baited alternative promoters for the same gene, and partially shared PIRs as those that interacted with a subset of alternative promoters for the same gene. We defined distinct PIRs as those that only interacted with a single promoter fragment (CHiCAGO score >= 5). To increase the stringency with which we called PIRs “distinct”, we applied two further criteria. First, if a PIR interacted with another alternative promoter at a lenient CHiCAGO score >=3, we defined that PIR as shared. Second, if the adjacent fragment to the PIR in question interacted with another alternative promoter at CHiCAGO score >=3, we also defined that PIR as shared. We note that, under our classification rules, the PIRs of genes with only two alternative promoters included in the analysis can only be classified as “fully shared” or “distinct”, therefore the “partially shared” PIR category is only applicable to the subset of genes with more than two baited alternative promoters

Epigenomic data pre-processing

The SRA accession list was downloaded from the GEO accession GSE77299. The SRA files were converted to fastq files and sequencing adapters were trimmed from reads using *trim galore* (https://github.com/FelixKrueger/TrimGalore). The 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 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequences were mapped to the hg38 reference genome using STAR with modification for aligning ChIP-seq and ATAC-seq reads. Samtools[119] was used to select reads with a MAPQ score of 255 which is the flag for unique mapping reads from STAR[120]. ATAC-seq reads were filtered, retaining properly paired and oriented reads using the samflag = 3. PCR duplicates were removed using samtools. We then removed reads that fell within blacklisted regions using Bedtools[121] intersect. The final filtered BAM file was then converted to a bed file using Bedtools bamtobed. This conversion breaks read-pairing and ensures each read contributes to peak identification with MACS2[122]. The ATAC-seq reads in BED format were shifted by +4 on the (+) strand and -5 on the (-) strand to account for the Tn5 transposase cut site. Peaks were

called using MACS2 using three biological replicates per sample as the treatment group with an input ChIP-seq control sample. The replicate correlation between the ATAC-seq samples was poor with a <10% overlap between biological replicates. This result was consistent with the high level of duplication and low peak count (8,852) in the worst sample (SRR3129112). Thus, our ATAC-seq results are limited to the sample with best quality metrics (SRR3129113). In total we detected 34,077 H3K27ac peaks and 72,825 ATAC-seq peaks.

PCHi-C adapted ABC model

For a given gene-enhancer pair, the ABC score is the normalised product of enhancer Activity (proxied by the levels of chromatin accessibility and relevant histone modifications) and Contact (proxied by 3D contact frequency detected from a chromosome conformation capture assay)[13,38]. In the original implementation of ABC, Activity is estimated as the geometric mean of read counts of DHS/ATAC-seq peaks, and Contact by KR-normalised Hi-C contact frequency between the respective element and gene promoter[38]. The resulting product is divided by the sum of all ABC values for a given gene from enhancers within a 5-megabase window around the transcription start site:

$$ABC\ Score_{E,G} = \frac{A_E \times C_{E,G}}{\Sigma e\ within\ 5\ Mb\ of\ G\ A_e \times C_{e,G}}$$

To adapt ABC for PCHi-C data, we took advantage of the CHiCAGO normalisation algorithm and developed an imputation procedure in the normalised counts space based on the inferred decay of interaction read counts with distance.
As we do not expect the frequency of enhancer-promoter contacts to fall below levels expected due to Brownian collision, for a given pair of fragments involving a baited promoter we select the maximum between the CHiCAGO-normalised observed read counts (N_{obs}) and expected read counts N_{exp} estimated as:

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

where B_{mean} is the CHiCAGO-estimated Brownian noise level and s_i and s_j and 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 due to QC fail, we estimated the expected normalised read count directly from the interaction distance *d*, using the distance function f(*d*) fitted by CHiCAGO. Please refer to Additional File 1 in the publication presenting the CHiCAGO pipeline[36] for the formal definition of these parameters and their estimation procedures.

The imputed normalised read counts were used as Contact data in the ABC pipeline, and the public H3K27ac and ATAC-seq data in ILC3s processed as described above were used to compute Activity. Our forked version of the ABC with additional scripts for imputations is available at https://github.com/pavarte/PCHIC-ABC-Prediction.

To validate the CHiC-ABC approach, we took advantage of the high-throughput CRISPRi-FlowFISH data from Fulco et al.[38], presenting the impact of perturbing ~3,500 enhancer elements on the expression of 30 genes in K562 cells. Since PCHi-C data for K562 cells are not currently available, we used our previously published PCHi-C dataset in the related primary cell type, erythroblasts, to generate

the ABC scores based on these data and the ATAC-Seq and H3K27ac ChIP-Seq datasets for K562 cells from Fulco et al. In comparison with the original ABC scores from Fulco et al. based on pooling conventional Hi-C data from multiple cell types, our approach showed a higher precision (69.1% vs 58.3%) at the same level of recall (58.3%) of CRISPRi-FlowFISH-validated enhancer-promoter pairs (**Fig. S2**). The CHiC-ABC score cutoff for ILC3 data was chosen by balancing the number of detected contacts with the effect size of the relationship between the number of connected enhancers and gene expression (**Fig. S3**), with a cutoff of 0.04 chosen for downstream analyses.

Microarray gene expression data analysis

The microarray CEL files were downloaded from the GEO accession number GSE78896. The CEL files were then analysed using AltAnalyze (http://www.altanalyze.org/). Probes were filtered for a DABG (detection above background) as previously described[123]. Probes were collapsed to the gene level and RMA normalised using the AltAnalyze platform.

RNA-seq data analysis

Human ILC3 RNA-seq data were downloaded from the GEO accession number GSE130775. Salmon[124] was used to quasi-map reads to transcripts. Reads were aligned to the hg38 genome assembly. The transcript counts were then imported and collapsed to gene counts using Tx import.

Mouse ILC3 differential RNA-seq data analysis was performed using DESeq2[125]. In brief, the gene count matrices were downloaded from GEO (GSE120723) and the standard DESeq2 algorithm was run according to the vignette. Low count genes were pre-filtered before running. The following parameters were used to report significantly differential expressed genes: alpha = 0.05 and adjusted p-value < 0.05.

PIR enrichment for epigenomic features

For each gene, sets of adjacent PIRs for each gene (detected at the fragment or 5kb resolution or the merged PIR sets for each gene) were collapsed together to obtain “collapsed PIRs” (cPIRs). Trans-chromosomal PIRs were removed. The observed proportion of cPIRs overlapping epigenomic features of interest (ATAC-seq, H3K27ac or H3K4me3, respectively) was computed using the *foverlaps* function from the *data.table* package in R. To obtain the expectation for this proportion, we repeated this analysis for random cPIRs that were generated by “transplanting” each set of all cPIRs for each gene to randomly selected genes in a manner preserving the size and spatial localisation of the cPIRs with respect to each other and the respective baited promoter fragment. This “transplantation” was repeated 100 times for all genes (baited promoter fragments), and the mean proportion of random cPIRs overlapping epigenomic features of interest (over 100 permutations), as all as a standard deviation of this quantity, were compared with the proportion of overlap for the observed cPIRs. Compared with the PIR enrichment estimation algorithm implemented in CHiCAGO (*peakEnrichment4Features*), this permutation procedure

preserves not only each PIR's distance from bait, but also the spatial relationships between multiple PIRs of the same gene.

RELI analysis

RELI[47] (v 0.1.1a) was used to find enrichment of genetic variants in promoter interacting regions (PIRs) that are accessible and marked with activating epigenetic markers (H3K27ac and ATAC-seq). In brief, RELI tests genomic features such as ATAC-seq, ChIP-seq, or PIRs for statistically significant overlaps with known disease risk variants identified from genome-wide association studies. Risk variants are expanded to linkage disequilibrium blocks (LD blocks) with variants that have an R2 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 performing the intersection with the feature files. A p-value is generated by comparing the observed number of intersections in the test to the null distribution.

The PIR sets were defined as the other ends of the BEDPE files. Interactions had to have a score of ≥ 5. We excluded trans-chromosomal and off target interactions. Promoter to promoter contacts were kept. ATAC-seq, H3K27ac, and H3K4me3 were intersected with each PIR set. The intersection of the features was used. We used UCSC liftover (v. 377) to lift over the genomic coordinates from hg38 to hg19 for use with RELI. The RE features used in this analysis are the union of ATAC-seq data with H3K27ac data for each cell type. All intervals were sorted and merged before input into RELI. Bonferroni and Benjamini-Hochberg p-value correction were performed with the python package statsmodels with alpha=0.05 (family-wise error rate of 5%; the probability that at least one of the predictions is a false positive). Traits with the BH adjusted p-value < 0.01 were defined as significant. For depicting RELI results we labelled only significant traits with enrichment >=2.

Standard COGS

To run standard COGS[29,30], we adapted the code from the R package rCOGS (https://github.com/ollyburren/rCOGS) to use the data.table framework instead of GenomicRanges for optimised speed and enable both the standard COGS and multiCOGS analyses (see Code availability). We used linkage disequilibrium blocks calculated for GRCh38 from https://github.com/jmacdon/LDblocks_GRCh38[126] and minor allele frequencies from the 1000 Genomes Project, European individuals. Protein coding SNPs were identified using VEP version 99.2 (https://github.com/Ensembl/ensembl-vep). We obtained gene transcription start sites (Havana and Ensembl/Havana merge) from Ensembl GRCh38 release 88 (March 2017), matching the version used to design the *DpnII* promoter capture system. We included promoters whether or not they were targeted in the capture system, enabling COGS to prioritise all gene targets where the causal variants fell near the gene promoter (defined as +/- 5 *DpnII* fragments from the transcription start site). We used minimum CHiCAGO interaction scores of 5 (solitary baits and other-ends binned to 5kb) and ABC scores of 0.026 for COGS input. The results for each protein-coding gene were linked across datasets using Ensembl gene IDs as primary identifiers. The Major Histocompatibility Complex was removed (GRCh38 6:28510120-33480577), prior to running COGS.

Sources of prior mechanistic evidence for CD genes

Datasets used to compare the COGS prioritised genes with previously functionally validated genes were: OpenTargets[64] L2G gene prioritisation score > 0.5 for five CD studies[48,127–130], the IBDDB database of functionally validated targets[65], a functional screen of IBD genes[66], experimentally validated IDB and CD genes from DisGeNET[67] that had evidence "AlteredExpression", "Biomarker", "Posttranslationalmodification", or "Therapeutic" or CD-containing exonic variants in a recent IBD exome study[62].

SuSIE and multiCOGS

The Sum of Single Effects (SuSIE) model allows for multiple causal variants within a GWAS locus[59,60]. We downloaded summary data for Crohn's disease[48] from the GWAS Catalog (http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST004001-GCST005000/GCST004132/harmonised/28067908-GCST004132-EFO_0000384.h.tsv.gz) and used LD block data for EUR from lddetect (https://bitbucket.org/nygcresearch/ldetect-data/src/master/) which we liftOvered[131] to hg38 to divide the data into approximately independent blocks. We used EUR samples from phased 1000 Genomes Phase 3 data, downloaded from https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html to generate LD matrices. We used these matrices to first impute the summary statistic data within blocks using the published method[132]. For blocks with appreciable association signals (minimum p < 10-6), we used the susieR package[59,60] 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, described below. For the remaining blocks, or where susieR failed to find any signals meeting our criteria, we fine-mapped using the single causal variable approach, as previously described[29,30], and used the posterior probabilities of association as input for multiCOGS.

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

$$\text{multiCOGSscore}_{gene}=1-\prod(1-\text{Score}_{gene,LDblock,credSet}).$$

To reveal the contributions of the four categories of genomic loci underlying the prioritised genes (PCHi-C PIRs, ABC enhancers, promoter proximal regions and coding SNPs) we ran multiCOGS on each category separately by specifying the *feature.names* argument in the *compute_cogs* function. We considered the "highest-contributing signal" for a prioritised gene as the category whose individual COGS score was greater than for all three other categories. We considered a gene to have contributions from more than one category if the score for each separate category was greater than zero.

Gene Set Enrichment Analysis

All genes were ranked by their COGS score and Gene Set Enrichment Analysis (GSEA) was performed against hallmark biological processes using the function "gseGO" in the R package clusterProfiler[133]. We obtained 496 datasets of differentially expressed (DE) genes within IBD tissues and cell types from The Inflammatory Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) Framework[63]. We made lists of significant DE genes (adjusted P value < 0.05 and absolute log2 fold change >= 2) that were compared across the same tissues and selected only sets with a maximum of 2000 DE genes, to avoid mis-estimation of the normalised enrichment score. We performed GSEA against the resultant 24 gene sets using the function "GSEA".

Comparison of multiCOGS results across PCHi-C datasets

To compare the ILC3 multiCOGS results with more abundant hematopoietic primary cell types, we first obtained the Javierre *et al*.[29] peak matrix at a cutoff of CHiCAGO score 5 from https://osf.io/u8tzp/ and converted it to GRCh38 using bedtools liftOver[131]. We then ran multiCOGS for CD separately on the ILC3 data and the Javierre *et al*. data in the following two settings: 1) considering all regions in the COGS analysis (PIRs, promoter proximal and coding regions) and 2) only considering non-PIRs (promoter proximal and coding regions). In each setting, we specified the required regions using the *feature.names* argument in the *compute_cogs* function. Genes that had a COGS score >=0.5 in the first setting but <0.5 in the second setting were defined as requiring PIRs for prioritisation in either dataset. Focusing on the genes that required PIRs for prioritisation in ILCs, we looked for their PIR-based prioritisation in Javierre *et al*. We could not ascertain the tissue-specificity for 9 genes that were prioritised based on PIRs in ILC3s but on non-PIRs in Javierre *et al*., because these cases were likely due to differences in experimental design, such as the different restriction fragment lengths and discrepancies in the locations of *DpnII* and *Hin*dIII restriction enzyme cut sites, respectively, between the studies.

Isolation and analysis of mouse small intestinal lamina propria ILC

Fifteen centimetres of small intestine were harvested from each mouse. Intestines were sequentially washed in buffered saline (HBSS) with 5mM EDTA to remove waste and epithelial cells then were minced and dissociated in RPMI containing collagenase (100 mg/ml collagenase A; Roche), DNase I (50 µg/ml; Sigma), and 10% FBS with shaking at 37°C for 35 min. Leukocytes were collected at the interface of a 40%/80% Percoll gradient (GE Healthcare). For flow cytometric analysis lamina propria mononuclear cells were first stained for surface markers and a near-infrared live/dead exclusion dye. Cells were then fixed and permeabilized before cytoplasmic staining according to the manufacturer's protocol (FOXP3 staining buffer set from eBioscience) with anti-IL17 (BioLegend, clone TC11-18H10.1). Flow cytometric analysis was performed on an LSRII (BD Biosciences). All data were re-analyzed using FlowJo (Tree Star). The Flow Cytometry analysis panel consisted of the following antibodies: Live/dead fixable near-IR Dead cell stain kit L34975 (Invitrogen); 421 Rat Anti-Mouse CD196 (CCR6) (Clone 140706, BD Biosciences);

BUV395 Hamster Anti-Mouse CD3e (Clone 145-2C11; BD Biosciences); BUV395BUV395 Rat Anti-Mouse CD19 (Clone 1D3; BD Biosciences); BUV395 Hamster Anti-Mouse TCR β Chain (Clone H57-597; BD Biosciences); Brilliant Violet 711 anti-mouse NK-1.1 (Clone PK136; BioLegend); Brilliant Violet 605 anti-mouse CD335 (NKp46) (Clone 29A1.4; Biolegend); FITC Hamster Anti-Mouse/human KLRG1 (Clone 2F1; BioLegend); PE/Cy7 Rat Anti-Mouse CD127 (IL-7Rα) (Clone A7R34; BioLegend); BUV737 Hamster Anti-Mouse CD69 (Y245; BD Biosciences); Brilliant Violet 421 Rat Anti-Mouse CD196 (CCR6; Clone 140706; BD Biosciences); PE Rat anti-mouse IL-17A (clone TC11-18H10.1; BioLegend); CD16/CD32 Monoclonal Antibody (93; eBioscience).

Comparison of the fraction of IL17-producing cells between wild-type and CLN3-mutant mouse ILC3s

The proportion of IL17-producing cells obtained from flow cytometry was regressed against the indicator variable reflecting the wild-type vs CLN3-mutant background and the batch number (n=2, with four mice in each) as a random effect using a mixed-effect Beta regression model implemented the function *glmer* in the *lme4* library in R. The analysis was performed separately for CCR6-positive and negative cells. The results were highly similar to those obtained using a logistic regression with a binary outcome variable reflecting IL-17 production per cell, and the mouse background and hierarchical mixed effects (batch and mouse) as predictors, and the total number of cells as a weight. The analysis was also performed using *glmer* function (not shown).

Animal Ethics Statement

Wild type and homozygous Cln3Δex7-8 (Jackson Laboratory # 017895) mutant mice on C57BL/6J backgrounds were used for all studies. All mice were housed under identical conditions in an AAALAC International-accredited facility with Institutional Animal Care and Use Committee (IACUC) approval (US Department of Agriculture [USDA] license 46-R-0009; protocol #178-02-24D, Sanford Research, Sioux Falls, SD, USA).

Data availability

Raw PCHi-C data generated in this study are deposited in the Gene Expression Omnibus (GEO) under the accession number GSE216267. Processed R data files containing CHiCAGO scores at fragment level and 5kb-binned resolution can be found in the same repository.

Code availability

Most scripts for analyses used in the paper are available at https://github.com/vmalysheva/ILC3 and https://github.com/malyshevalab/hILCs_CHi-C, with the following exceptions: CHiC-ABC (https://github.com/pavarte/PCHIC-ABC-Prediction), RELI (https://github.com/tacazares/spivakov_pchic_ILC_CD4), SuSiE

(https://github.com/chr1swallace/cd-finemapping-scripts), COGS and multiCOGS (https://github.com/FunctionalGeneControl/multiCOGS).

Conflict of interest

P.F., S.S. and M.S. are shareholders of Enhanc3D 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. C.W. is also an employee of GSK; GSK had no input into this piece of work.

Acknowledgements

The authors would like to thank Laurence Game and Ivan Andrew at the LMS Genomics facility for sequencing, Michiel Thiecke for technical advice and Wing Leung for technical assistance. M.S. is core-funded by the Medical Research Council as an MRC Investigator (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). This work was supported by the National Institute of Health (U01AI150748 to E.R.M, M.T.W, L.C.K; R01AI153442 to E.R.M; R21AI156185 to E.R.M; R01HG010730, U01AI130830, R01NS099068, R01GM055479, P01AI150585, R01AI141569 to M.T.W; NIH R01AI024717 to M.T.W and L.C.K; NIH R01AR073228 to M.T.W, L.C.K and S.W; R01DK107502, R01AI148276, U19AI070235, U01HG011172, and P30AR070549 to L.C.K); and 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). M.T.W. S.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 an NIH training grant T32 AR069512. D.F.S acknowledges the support by NIH grant 5K08HL148551-02. D.O. was supported by an American Heart Association fellowship**.** This project was funded in part by a Commercialisation Award from the Babraham Institute to M.S. and P.F. 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.

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* vol. 348 Preprint at https://doi.org/10.1126/science.aaa6566 (2015).
3. Spits, H. *et al.* Innate lymphoid cells — a proposal for uniform nomenclature. *Nature Reviews Immunology* vol. 13 145–149 Preprint at https://doi.org/10.1038/nri3365 (2013).
4. Ebbo, M., Crinier, A., Vély, F. & Vivier, E. Innate lymphoid cells: major players in inflammatory diseases. *Nature Reviews Immunology* vol. 17 665–678 Preprint at

https://doi.org/10.1038/nri.2017.86 (2017).
5. Vivier, E. *et al.* Innate Lymphoid Cells: 10 Years On. *Cell* **174**, 1054–1066 (2018).
6. Meininger, I. *et al.* Tissue-Specific Features of Innate Lymphoid Cells. *Trends Immunol.* **41**, 902–917 (2020).
7. Fuchs, A. *et al.* Intraepithelial type 1 innate lymphoid cells are a unique subset of IL-12- and IL-15-responsive IFN-γ-producing cells. *Immunity* **38**, 769–781 (2013).
8. Fan, H. *et al.* Innate Lymphoid Cells: Regulators of Gut Barrier Function and Immune Homeostasis. *J Immunol Res* **2019**, 2525984 (2019).
9. Saez, A. *et al.* Innate Lymphoid Cells in Intestinal Homeostasis and Inflammatory Bowel Disease. *Int. J. Mol. Sci.* **22**, (2021).
10. 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).
11. Zeng, B. *et al.* ILC3 function as a double-edged sword in inflammatory bowel diseases. *Cell Death Dis.* **10**, 315 (2019).
12. Mazzurana, L. *et al.* Crohn's Disease Is Associated With Activation of Circulating Innate Lymphoid Cells. *Inflamm. Bowel Dis.* **27**, 1128–1138 (2021).
13. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
14. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
15. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
16. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **11**, 424 (2020).
17. Leung, A. K.-Y., Yao, L. & Yu, H. Functional genomic assays to annotate enhancer-promoter interactions genome-wide. *Hum. Mol. Genet.* (2022) doi:10.1093/hmg/ddac204.
18. Robinette, M. L. *et al.* Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat. Immunol.* **16**, 306–317 (2015).
19. Gury-BenAri, M. *et al.* The Spectrum and Regulatory Landscape of Intestinal Innate Lymphoid Cells Are Shaped by the Microbiome. *Cell* **166**, 1231–1246.e13 (2016).
20. Koues, O. I. *et al.* Distinct Gene Regulatory Pathways for Human Innate versus Adaptive Lymphoid Cells. *Cell* **165**, 1134–1146 (2016).
21. Shih, H.-Y. *et al.* Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell* **165**, 1120–1133 (2016).
22. Pokrovskii, M. *et al.* Characterization of Transcriptional Regulatory Networks that Promote and Restrict Identities and Functions of Intestinal Innate Lymphoid Cells. *Immunity* **51**, 185–197.e6 (2019).
23. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
24. 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).
25. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-

resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
26. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
27. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).
28. Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W. & Fraser, P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J. Vis. Exp.* (2018) doi:10.3791/57320.
29. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384.e19 (2016).
30. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* **18**, 165 (2017).
31. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat. Commun.* **6**, 10069 (2015).
32. Baxter, J. S. *et al.* Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.* **9**, 1028 (2018).
33. Ray-Jones, H. *et al.* Mapping DNA interaction landscapes in psoriasis susceptibility loci highlights KLF4 as a target gene in 9q31. *BMC Biol.* **18**, 47 (2020).
34. Shi, C. *et al.* Chromatin Looping Links Target Genes with Genetic Risk Loci for Dermatological Traits. *J. Invest. Dermatol.* **141**, 1975–1984 (2021).
35. Thiecke, M. J., Yang, E. J., Burren, O. S., Ray-Jones, H. & Spivakov, M. Prioritisation of Candidate Genes Underpinning COVID-19 Host Genetic Traits Based on High-Resolution 3D Chromosomal Topology. *Front. Genet.* **12**, 745672 (2021).
36. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
37. Freire-Pritchett, P. *et al.* Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat. Protoc.* **16**, 4144–4176 (2021).
38. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
39. Cotman, S. L. & Staropoli, J. F. The juvenile Batten disease protein, CLN3, and its role in regulating anterograde and retrograde post-Golgi trafficking. *Clin. Lipidol.* **7**, 79–91 (2012).
40. Tang, C. *et al.* A human model of Batten disease shows role of CLN3 in phagocytosis at the photoreceptor–RPE interface. *Communications Biology* **4**, 1–18 (2021).
41. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
42. Reyes, A. & Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **46**, 582–592 (2018).
43. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1145–1149 (2016).
44. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in

blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
45. Ray-Jones, H. & Spivakov, M. Transcriptional enhancers and their communication with gene promoters. *Cell. Mol. Life Sci.* **78**, 6453–6485 (2021).
46. 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).
47. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
48. 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).
49. Naser, S. A. *et al.* Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J. Gastroenterol.* **18**, 412–424 (2012).
50. 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).
51. Assadi, G. *et al.* LACC1 polymorphisms in inflammatory bowel disease and juvenile idiopathic arthritis. *Genes Immun.* **17**, 261–264 (2016).
52. 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).
53. 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).
54. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
55. Keir, M. E. *et al.* Regulation and Role of αE Integrin and Gut Homing Integrins in Migration and Retention of Intestinal Lymphocytes during Inflammatory Bowel Disease. *J. Immunol.* **207**, 2245–2254 (2021).
56. Coskun, M., Olsen, J., Seidelin, J. B. & Nielsen, O. H. MAP kinases in inflammatory bowel disease. *Clin. Chim. Acta* **412**, 513–520 (2011).
57. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254 (2022).
58. Castel, S. E. *et al.* Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).
59. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).
60. 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).
61. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).
62. Sazonovs, A. *et al.* Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. *Nat. Genet.* 1–9 (2022).
63. Massimino, L. *et al.* The Inflammatory Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nature Computational Science* vol. 1 511–515 Preprint at https://doi.org/10.1038/s43588-021-00114-y (2021).

64. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
65. 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).
66. Ntunzwenimana, J. C. *et al.* Functional screen of inflammatory bowel disease genes reveals key epithelial functions. *Genome Med.* **13**, 181 (2021).
67. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2016).
68. Chea, S. *et al.* Notch signaling in group 3 innate lymphoid cells modulates their plasticity. *Sci. Signal.* **9**, ra45 (2016).
69. Mole, S. E. & Cotman, S. L. Genetics of the neuronal ceroid lipofuscinoses (Batten disease). *Biochim. Biophys. Acta* **1852**, 2237–2241 (2015).
70. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).
71. Ueta, M. *et al.* Mucocutaneous inflammation in the Ikaros Family Zinc Finger 1-keratin 5-specific transgenic mice. *Allergy* **73**, 395–404 (2018).
72. Hildner, K., Punkenburg, E., Abendroth, B. & Neurath, M. F. Immunopathogenesis of IBD: Batf as a Key Driver of Disease Activity. *Dig. Dis.* **34 Suppl 1**, 40–47 (2016).
73. 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).
74. Bianchi, E., Rogge, L. & Vecellio, M. *Role of the IL-23/IL-17 Pathway in Chronic Immune-Mediated Inflammatory Diseases: Mechanisms and Targeted Therapies*. (Frontiers Media SA, 2021).
75. Atreya, I., Atreya, R. & Neurath, M. F. NF-kappaB in inflammatory bowel disease. *J. Intern. Med.* **263**, 591–596 (2008).
76. Saul, D. *et al.* Cell Type-Specific Induction of Inflammation-Associated Genes in Crohn's Disease and Colorectal Cancer. *Int. J. Mol. Sci.* **23**, (2022).
77. Neurath, M. F. & Finotto, S. The Many Roads to Inflammatory Bowel Diseases. *Immunity* **25**, 189–191 (2006).
78. Deng, T. *et al.* ILC3-derived OX40L is essential for homeostasis of intestinal Tregs in immunodeficient mice. *Cell. Mol. Immunol.* **17**, 163–177 (2020).
79. Shiu, J. *et al.* Gastric LTi cells promote lymphoid follicle formation but are limited by IRAK-M and do not alter microbial growth. *Mucosal Immunol.* **8**, 1047–1059 (2015).
80. Chu, C. *et al.* Anti-microbial Functions of Group 3 Innate Lymphoid Cells in Gut-Associated Lymphoid Tissues Are Regulated by G-Protein-Coupled Receptor 183. *Cell Rep.* **23**, 3750–3758 (2018).
81. Ma, Y. *et al.* ACF7 regulates inflammatory colitis and intestinal wound response by orchestrating tight junction dynamics. *Nat. Commun.* **8**, 15375 (2017).
82. Messal, N. *et al.* Differential role for CD277 as a co-regulator of the immune signal in T and NK cells. *Eur. J. Immunol.* **41**, 3443–3454 (2011).
83. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–32 (2016).

84. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
85. Chantalat, S. *et al.* Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* **21**, 1426–1437 (2011).
86. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2018).
87. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
88. Lepik, K. *et al.* C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLoS Comput. Biol.* **13**, e1005766 (2017).
89. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
90. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643–656.e17 (2016).
91. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398–1414.e24 (2016).
92. Cotman, S. L. *et al.* Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth. *Hum. Mol. Genet.* **11**, 2709–2721 (2002).
93. Ahn, Y.-O. *et al.* Human group3 innate lymphoid cells express DR3 and respond to TL1A with enhanced IL-22 production and IL-2-dependent proliferation. *Eur. J. Immunol.* **45**, 2335–2342 (2015).
94. Furfaro, F. *et al.* TL1A: A New Potential Target in the Treatment of Inflammatory Bowel Disease. *Curr. Drug Targets* **22**, 760–769 (2021).
95. 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.e5 (2018).
96. Michieletto, M. F. *et al.* Multiscale 3D genome organization underlies ILC2 ontogenesis and allergic airway inflammation. *Nat. Immunol.* 1–13 (2022).
97. Ho, J. S. Y. *et al.* TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal inflammation. *Cell* **184**, 2618–2632.e17 (2021).
98. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
99. 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).
100. 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).
101. Barral, A. *et al.* SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol. Cell* **82**, 816–832.e12 (2022).
102. Bonder, M. J. *et al.* Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat. Genet.* **53**, 313–321 (2021).

103. Cui, Y. *et al.* 3'aQTL-atlas: an atlas of 3'UTR alternative polyadenylation quantitative trait loci across human normal tissues. *Nucleic Acids Res.* **50**, D39–D45 (2022).
104. Uriostegui-Arcos, M., Mick, S. T., Shi, Z., Rahman, R. & Fiszbein, A. Splicing-dependent transcriptional activation. *bioRxiv* 2022.09.16.508316 (2022) doi:10.1101/2022.09.16.508316.
105. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
106. Mirza, M. *et al.* The CLN3 gene and protein: What we know. *Mol Genet Genomic Med* **7**, e859 (2019).
107. Laqtom, N. N. *et al.* CLN3 is required for the clearance of glycerophosphodiesters from lysosomes. *Nature* **609**, 1005–1011 (2022).
108. Brudvig, J. J. *et al.* Glycerophosphoinositol is Elevated in Blood Samples From pigs, Mice, and CLN3-Affected Individuals. *Biomark. Insights* **17**, 11772719221107765 (2022).
109. Marielle S.Köberlin Berend Snijder Leonhard X.Heinz Christoph L.Baumann Astrid Fauster Gregory I.Vladimer Anne-Claude Gavin Giulio Superti-Furga. A Conserved Circular Network of Coregulated Lipids Modulates Innate Immune Responses. *Cell* **162**, 170–183 (2015).
110. IgG entry and deposition are components of the neuroimmune response in Batten disease. *Neurobiol. Dis.* **25**, 239–251 (2007).
111. Wu, J. & Li, L. Autoantibodies in Alzheimer's disease: potential biomarkers, pathogenic roles, and therapeutic implications. *J. Biomed. Res.* **30**, 361–372 (2016).
112. Reighard, S. D. *et al.* Therapeutic Targeting of Follicular T Cells with Chimeric Antigen Receptor-Expressing Natural Killer Cells. *Cell Rep Med* **1**, (2020).
113. Krabbendam, L., Nagasawa, M., Spits, H. & Bal, S. M. Isolation of Human Innate Lymphoid Cells. *Curr. Protoc. Immunol.* **122**, e55 (2018).
114. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
115. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
116. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **18**, 1046–1055 (2021).
117. Su, C., Pahl, M. C., Grant, S. F. A. & Wells, A. D. Restriction enzyme selection dictates detection range sensitivity in chromatin conformation capture-based variant-to-gene mapping approaches. *Hum. Genet.* **140**, 1441–1448 (2021).
118. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018) doi:10.48550/arXiv.1802.03426.
119. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
120. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
121. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
122. Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv* (2018) doi:10.1101/496521.
123. Lockstone, H. E. Exon array data analysis using Affymetrix power tools and R

statistical software. *Brief. Bioinform.* **12**, 634–644 (2011).
124. 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).
125. 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).
126. MacDonald, J. W., Harrison, T., Bammler, T. K., Mancuso, N. & Lindström, S. An updated map of GRCh38 linkage disequilibrium blocks based on European ancestry data. *Biorxiv* (2022) doi:10.1101/2022.03.04.483057.
127. Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
128. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
129. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
130. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518 (2016).
131. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–8 (2006).
132. Rüeger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* **14**, e1007371 (2018).
133. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

List of supplementary materials

Supplementary Tables

Table S1: Significant ILC3 CHi-C interactions at single-fragment resolution

Table S2: Significant ILC3 CHi-C interactions at 5kb resolution

Table S3: Comparison of ILC3 CHi-C interactions at HindIII resolution with those of other cell types profiled in Javierre et al or cell type specific interactions

Table S4: Combination of ABC-predicted enhancers and CHi-C interactions

Table S5: Significant RELI results for genetic traits enriched in ILC3 PIRs

Table S6: Standard COGS results for ILC3s with Crohn's Disease GWAS

Table S7: multiCOGS results for ILC3s with Crohn's Disease GWAS

Table S8: Prior functional evidence in IBD or CD for genes prioritised in ILC3s using multiCOGs

Table S9: Comparison with Javierre results for genes prioritised based on PIRs in ILC3s

Table S10: Prioritised multiCOGS genes for Crohn's Disease that were differentially expressed in TL1A mouse study

Table S11: CHiCAGO weights used for running CHiCAGO on fragment resolution ILC3 PCHi-C data

Supplementary Figures

Figure S1: Analysis of isoform-specific long-range and short-range chromosomal interactions

Figure S2: Validation of the CHiC-ABC method using CRISPR perturbation data in K562 cells.

Figure S3: CHiC-ABC analysis in ILC3 cells.

Figure S4: Biological function of prioritised genes in univariate COGS and multiCOGS for Crohn's Disease, de Lange et al GWAS

Figure 1

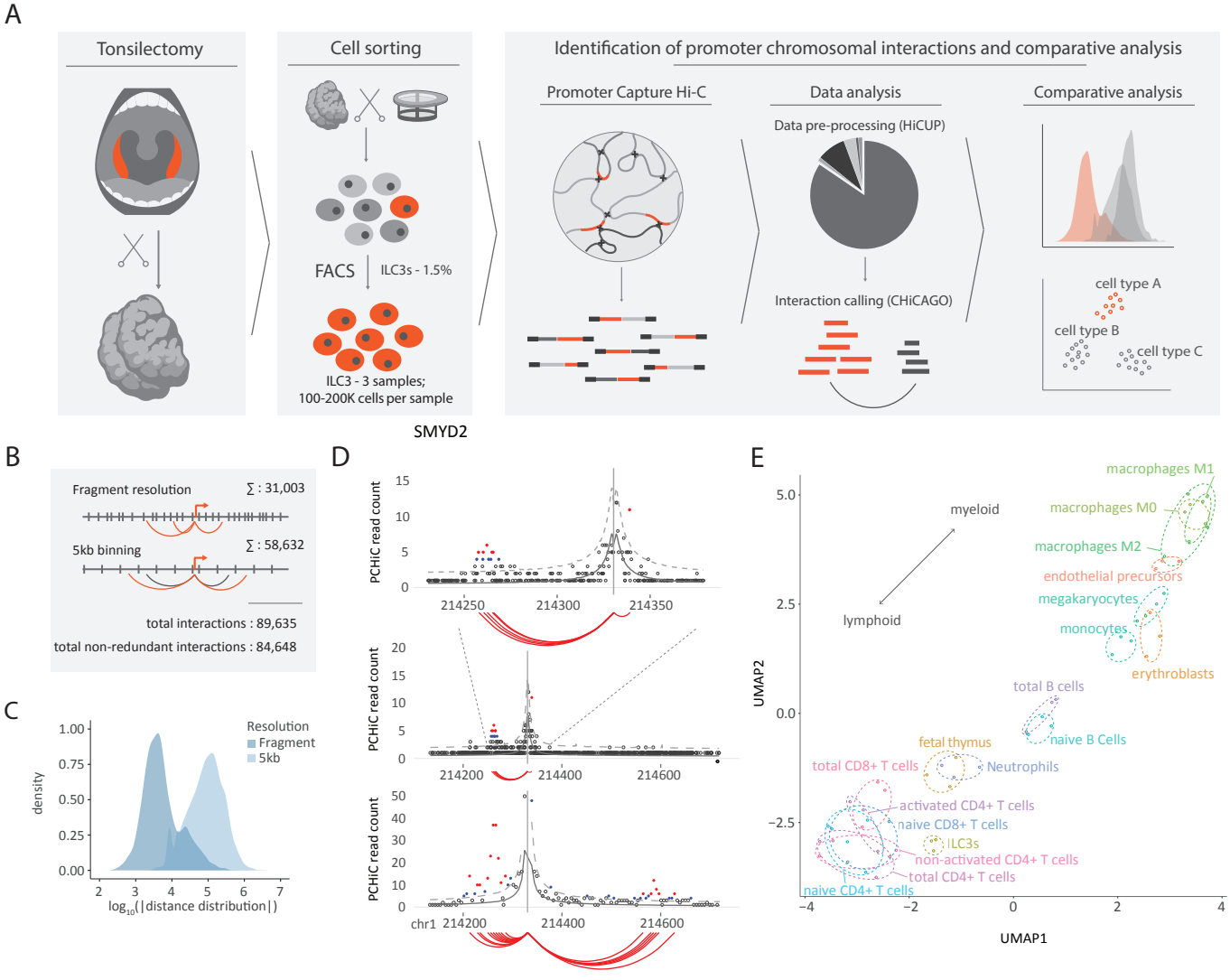

A
Tonsilectomy
Cell sorting
FACS
ILC3s - 1.5%
ILC3 - 3 samples;
100-200K cells per sample
Identification of promoter chromosomal interactions and comparative analysis
Promoter Capture Hi-C
Data analysis
Data pre-processing (HiCUP)
Interaction calling (CHiCAGO)
Comparative analysis
cell type A
cell type B
cell type C
B
Fragment resolution
∑ : 31,003
5kb binning
∑ : 58,632
total interactions : 89,635
total non-redundant interactions : 84,648
C
Resolution
Fragment
5kb
density
log10(|distance distribution|)
D
SMYD2
PCHiC read count
chr1
E
macrophages M1
macrophages M0
macrophages M2
endothelial precursors
megakaryocytes
monocytes
erythroblasts
myeloid
lymphoid
total B cells
naive B Cells
fetal thymus
Neutrophils
total CD8+ T cells
activated CD4+ T cells
naive CD8+ T cells
ILC3s
non-activated CD4+ T cells
total CD4+ T cells
naive CD4+ T cells
UMAP1
UMAP2

Figure 2

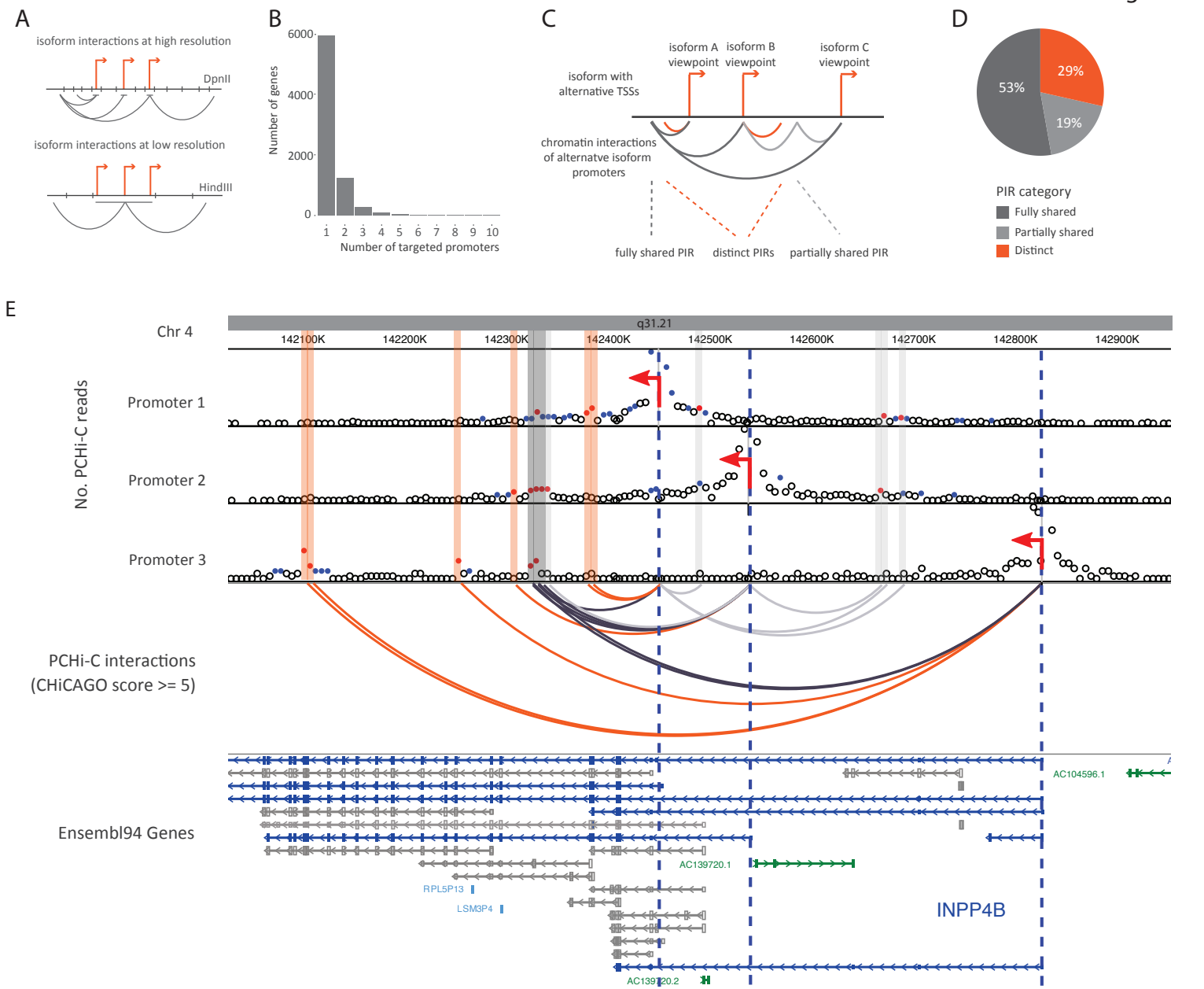

A
isoform interactions at high resolution
DpnII
isoform interactions at low resolution
HindIII
B
Number of genes
6000
4000
2000
0
1 2 3 4 5 6 7 8 9 10
Number of targeted promoters
C
isoform A viewpoint
isoform B viewpoint
isoform C viewpoint
isoform with alternative TSSs
chromatin interactions of alternatve isoform promoters
fully shared PIR
distinct PIRs
partially shared PIR
D
29%
53%
19%
PIR category
Fully shared
Partially shared
Distinct
E
Chr 4
q31.21
142100K
142200K
142300K
142400K
142500K
142600K
142700K
142800K
142900K
No. PCHi-C reads
Promoter 1
Promoter 2
Promoter 3
PCHi-C interactions
(CHiCAGO score >= 5)
Ensembl94 Genes
AC104596.1
AC139720.1
RPL5P13
LSM3P4
INPP4B
AC139720.2

Figure 3

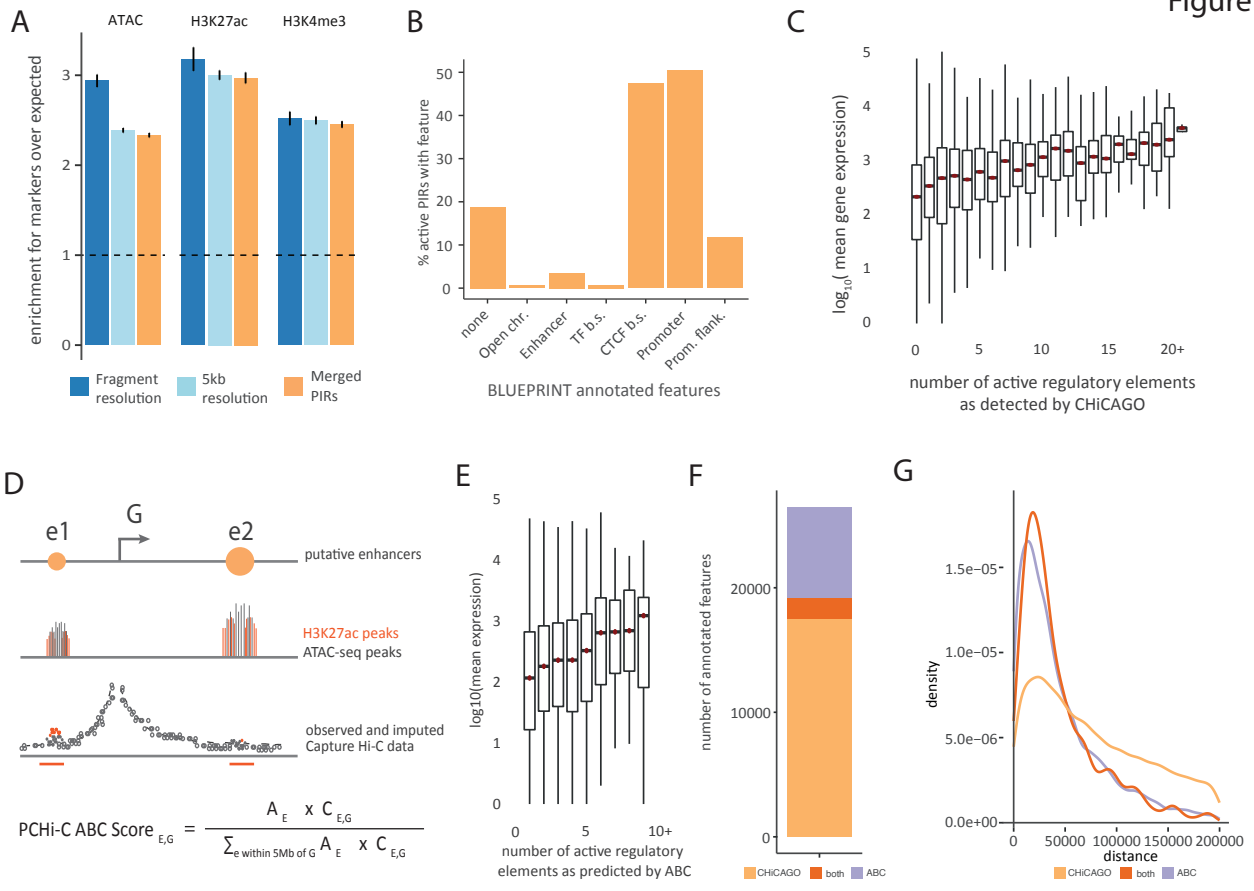
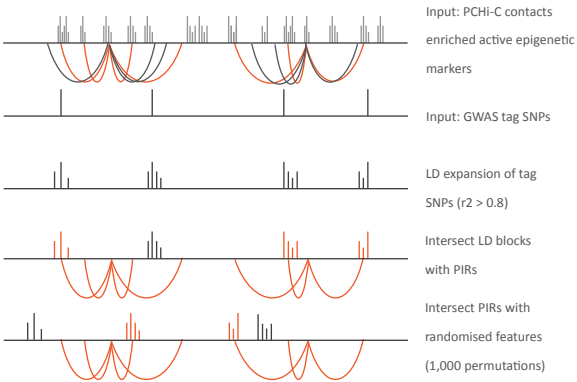
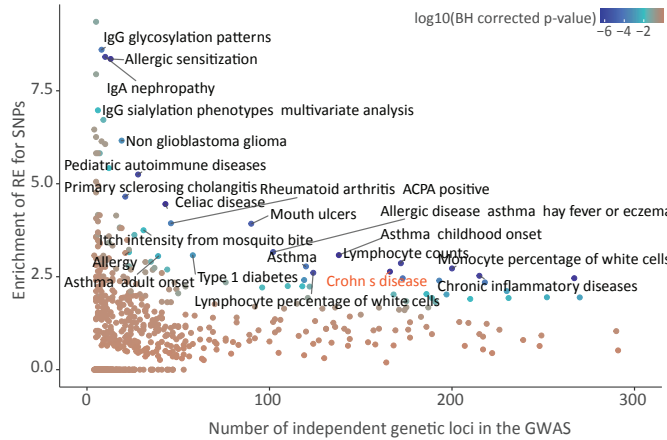

A
ATAC
H3K27ac
H3K4me3
enrichment for markers over expected
Fragment resolution
5kb resolution
Merged PIRs
B
% active PIRs with feature
none
Open chr.
Enhancer
TF b.s.
CTCF b.s.
Promoter
Prom. flank.
BLUEPRINT annotated features
C
log10(mean gene expression)
number of active regulatory elements as detected by CHiCAGO
D
e1
G
e2
putative enhancers
H3K27ac peaks
ATAC-seq peaks
observed and imputed Capture Hi-C data
PCHi-C ABC Score E,G = A E x C E,G / ∑e within 5Mb of G A E x C E,G
E
log10(mean expression)
number of active regulatory elements as predicted by ABC
F
number of annotated features
CHiCAGO
both
ABC
G
density
distance
CHiCAGO
both
ABC

Figure 4

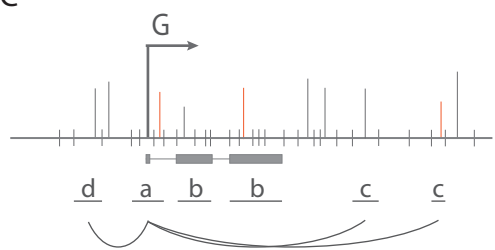
A



B



C



a - promoter region
b - coding region
c - CHiCAGO PIR
d - ABC-predicted enhancer

$$LD.score = \sum (p.value)$$

$$COGS\ Score = 1 - \prod_{i \in LD\ blocks} (1 - LD.score_i)$$

D

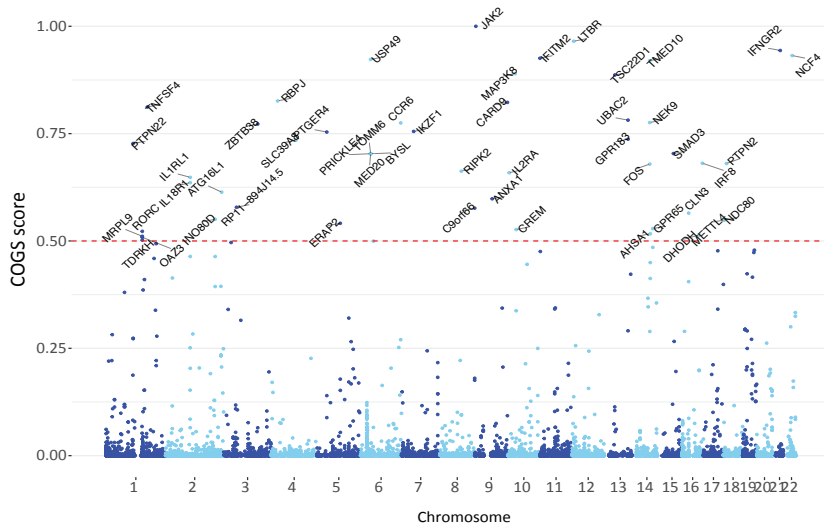


Figure 5

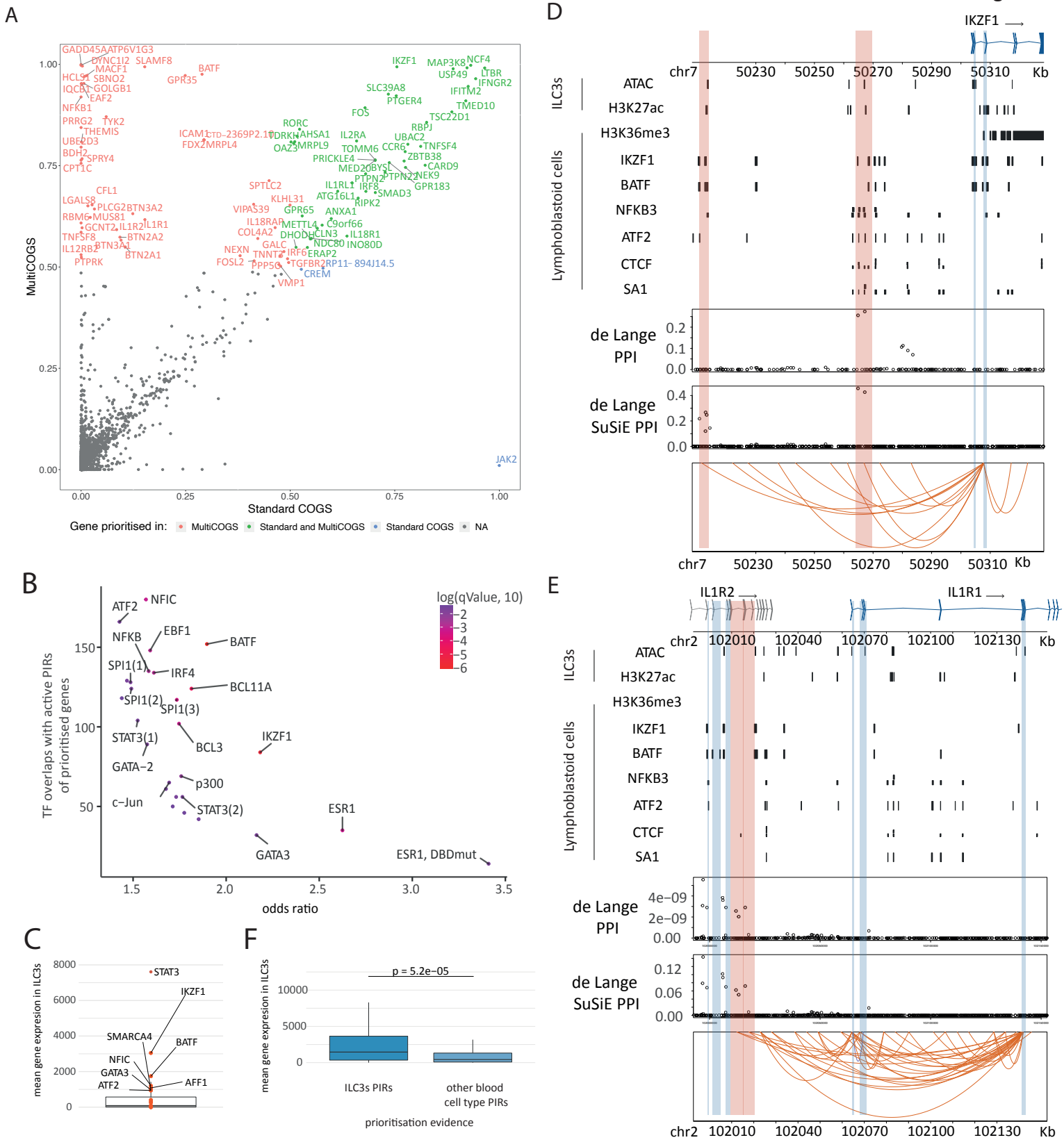
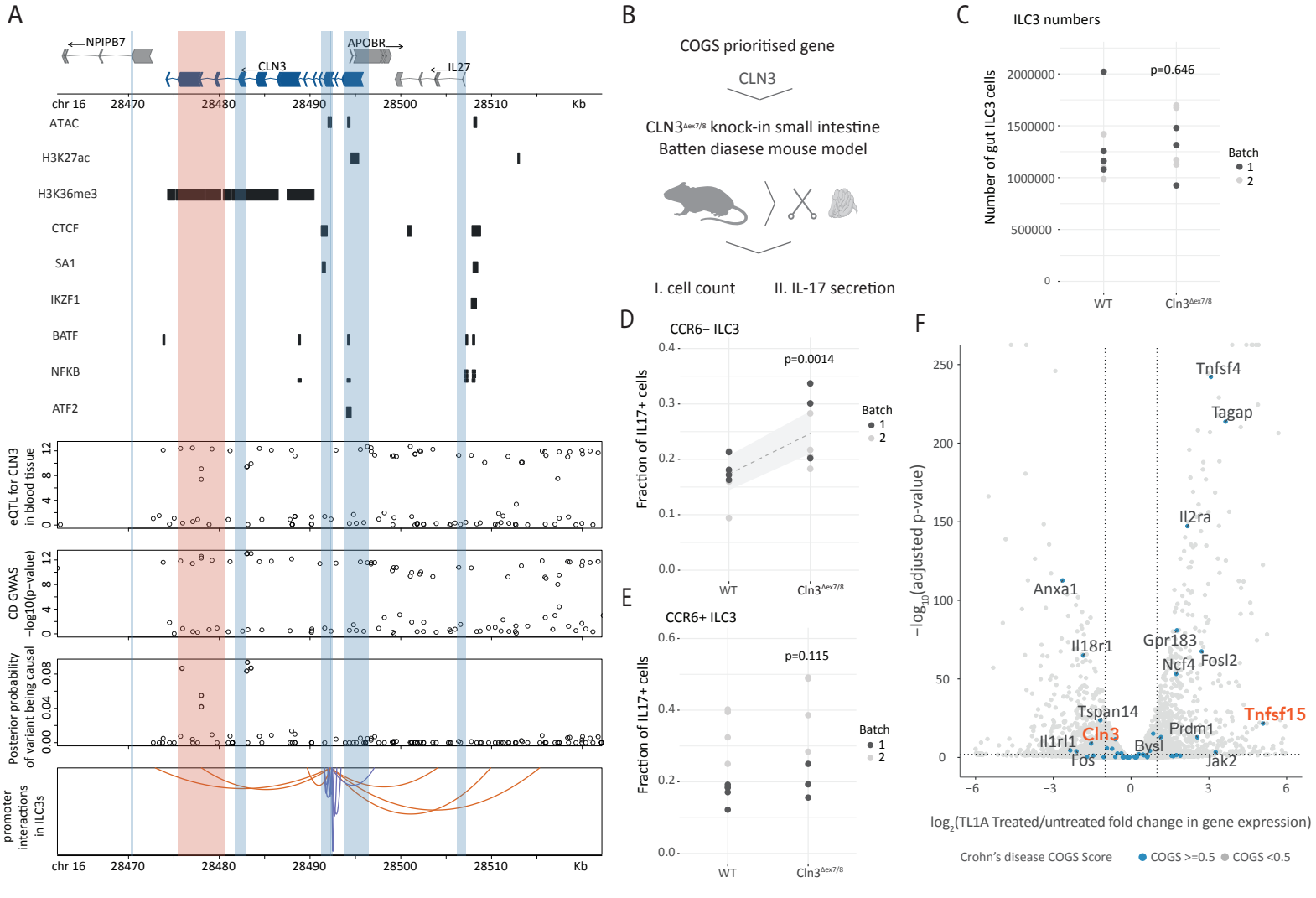

A
MultiCOGS
Standard COGS
Gene prioritised in: MultiCOGS Standard and MultiCOGS Standard COGS NA
B
TF overlaps with active PIRs of prioritised genes
odds ratio
log(qValue, 10)
NFIC
ATF2
EBF1
NFKB
BATF
SPI1(1)
IRF4
BCL11A
SPI1(2)
SPI1(3)
STAT3(1)
BCL3
IKZF1
GATA−2
p300
c−Jun
STAT3(2)
ESR1
GATA3
ESR1, DBDmut
C
mean gene expresion in ILC3s
STAT3
IKZF1
SMARCA4
BATF
NFIC
GATA3
ATF2
AFF1
F
mean gene expresion in ILC3s
p = 5.2e−05
ILC3s PIRs
other blood cell type PIRs
prioritisation evidence
D
IKZF1
chr7
Kb
ILC3s
ATAC
H3K27ac
Lymphoblastoid cells
H3K36me3
IKZF1
BATF
NFKB3
ATF2
CTCF
SA1
de Lange PPI
de Lange SuSiE PPI
E
IL1R2
IL1R1
chr2
Kb
ILC3s
ATAC
H3K27ac
Lymphoblastoid cells
H3K36me3
IKZF1
BATF
NFKB3
ATF2
CTCF
SA1
de Lange PPI
de Lange SuSiE PPI

Figure 6


A
NPIPB7
CLN3
APOBR
IL27
chr 16
28470
28480
28490
28500
28510
Kb
ATAC
H3K27ac
H3K36me3
CTCF
SA1
IKZF1
BATF
NFKB
ATF2
eQTL for CLN3 in blood tissue
CD GWAS −log10(p−value)
Posterior probability of variant being causal
promoter interactions in ILC3s
B
COGS prioritised gene
CLN3
CLN3Δex7/8 knock-in small intestine
Batten diasese mouse model
I. cell count
II. IL-17 secretion
C
ILC3 numbers
Number of gut ILC3 cells
p=0.646
Batch
1
2
WT
Cln3Δex7/8
D
CCR6− ILC3
Fraction of IL17+ cells
p=0.0014
E
CCR6+ ILC3
p=0.115
F
−log10(adjusted p-value)
Tnfsf4
Tagap
Il2ra
Anxa1
Il18r1
Gpr183
Ncf4
Fosl2
Tspan14
Cln3
Tnfsf15
Prdm1
Il1rl1
Bysl
Fos
Jak2
log2(TL1A Treated/untreated fold change in gene expression)
Crohn's disease COGS Score
COGS >=0.5
COGS <0.5