High-resolution promoter interaction analysis implicates genes involved in the activation of Type 3 Innate Lymphoid Cells in autoimmune disease risk

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

[1] MRC Laboratory of Medical Sciences, London W12 0HS, UK
[2] Institute of Clinical Sciences, Faculty of Medicine, Imperial College, London W12 0HS, 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] VIB, Center for AI and Computational Biology, Leuven, 2610, Belgium
[6] Trinity Hall, University of Cambridge, Cambridge, UK
[7] Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[8] Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[9] Medical Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[10] Division of Human Genetics and Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[11] Department of Microbial Infection and Immunity, The Ohio State University, Columbus, OH, 43210, USA.
[12] Pelotonia Institute for Immuno-Oncology, The Ohio State University, Columbus, OH 43210, USA.
[13] Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[14] Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[15] Center for Genetics and Rare Diseases, Sanford Research, Sioux Falls, SD 57104, USA
[16] University of Cambridge, Cambridge, UK
[17] Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
[18] National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK
[19] Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter Medical School, Exeter, UK
[20] The Babraham Institute, Cambridge, CB22 3AT, UK
[21] Florida State University, Department of Biological Sciences, Florida State University, Tallahassee, FL 32306-4295, USA
[22] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[23] Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, USA
[24] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[25] Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[26] Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[27] Divisions of Pediatric Otolaryngology and Pulmonary & Sleep Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA.
[28] Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, UK
[29] MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
[30] Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, CB2 0AW

¶ Joint first authors * Joint second authors ^ These authors jointly supervised the work
§ Corresponding authors: V.M. (valeriya.malysheva@vib.be), S.W. (stephen.waggoner@cchmc.org) & M.S. (mikhail.spivakov@lms.mrc.ac.uk)

Present addresses: O.C: Wake Forest School of Medicine, Section of Pediatric Rheumatology, Department of Pediatrics. T.A.C.: Eli Lilly and Company, IN; P.A.: Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, UK; 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. M.D.R.: Cyted, Cambridge UK; J.I.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; F.B.: University of Kent, Canterbury, UK`

Abstract

Innate lymphoid cells (ILCs) are rare, tissue-resident innate lymphocytes that functionally mirror CD4+ T helper cell lineages but lack antigen receptors. Type 3 ILCs (ILC3s) are enriched in the gut, airways, and mucosal lymphoid tissues, where they regulate inflammation and promote barrier integrity. To define the regulatory architecture of primary human ILC3s, we map promoter-anchored chromosomal contacts using high-resolution, low-input Promoter Capture Hi-C (PCHi-C) in these cells alongside CD4+ T cells. By combining statistical detection with a PCHi-C-adapted Activity-by-Contact approach, we link promoters to distal regulatory elements, identifying hundreds of ILC3-specific contacts. We use these maps to connect genome-wide association study (GWAS) risk variants for Crohn's disease to target genes using multiCOGS, a Bayesian framework that integrates PCHi-C with summary-statistic imputation and multivariate fine-mapping. This analysis highlights both known and unanticipated candidates, including *CLN3*, a causal gene for the neurodevelopmental Batten disease. Using a mouse ILC3-like cell line, we show that *Cln3* is downregulated upon cytokine stimulation, and *Cln3* overexpression alters stimulation-induced transcriptional programmes and cytokine secretion. Extending this approach, we generate a catalogue of ILC3-linked risk genes for five additional autoimmune conditions and show that they are enriched for regulators of the ILC3 inflammatory response identified in a CRISPR interference screen. Together, these findings illuminate long-range gene control in ILC3s and prioritise known and newly implicated autoimmune risk genes with potential roles in this clinically important cell type.

Introduction

Innate lymphoid cells (ILCs) play crucial roles in inflammation and immunity, as well as in tissue development and homeostasis[1,2]. ILCs develop from common lymphoid progenitors and share many features with CD4+ T lymphocytes, but do not express rearranged T cell receptors[3]. Therefore, rather than acting as part of the adaptive immune system, ILCs respond to cytokines and pathogens from the environment by producing regulatory cytokines and exerting immunomodulatory activity[4,5].

Three main types of ILCs have been identified based on their cytokine profiles 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 certain bacteria[6,7]. The second group consists of ILC2s, which regulate airway and skin inflammatory responses and are implicated in disorders such as asthma and atopic dermatitis[6]. Finally, the third group includes lymphoid tissue-inducer cells, which are involved in lymph node development, and ILC3s, which participate in host defence and the maintenance of epithelial barrier homeostasis[2–4]. The ILC3 population is distributed across multiple tissues, including the gut, where they are essential for mucosal homeostasis and barrier integrity[8]. ILC3-derived cytokines such as IL-17 and IL-22 promote epithelial cell renewal and release of antimicrobial peptides[9]. However, overexpression of these cytokines in the gut has been associated with the development 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 associated with these conditions[13]. 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[14–16], and therefore, cell type-specific maps of active enhancers and enhancer-promoter connections provide important clues for the functional interpretation of GWAS findings[17,18]. 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[19–23]. However, enhancers often localise large distances (up to megabases) away from their target gene promoters, physically contacting them in the 3D space of the nucleus in a cell-type-specific manner. Therefore, robust and sensitive identification of enhancer-promoter contacts, which is instrumental for inferring the effector genes of non-coding GWAS variants, requires robust and sensitive profiling of chromosomal 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 target genes[24,25]. 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 needed for the detection of specific enhancer-

130 promoter contacts. This challenge can be addressed by techniques such as Capture Hi-C that
131 selectively enrich Hi-C material for contacts involving, at one end, regions of interest such as
132 gene promoters[26–29]. Over the last decade, we and others have demonstrated the power of
133 Promoter Capture Hi-C (PCHi-C) in determining transcriptional regulatory circuitries and in
134 linking enhancers and disease-associated genetic variants with putative target genes[30–36]. In
135 foundational studies[30,31], we applied this approach to 17 abundant human primary blood cell
136 types and developed COGS (Capture Hi-C Omnibus Gene Score), a Bayesian approach for
137 prioritisation of GWAS target genes using statistical fine-mapping and PCHi-C data. Results
138 from this work were incorporated into major variant-to-gene resources, including OpenTargets
139 Genetics[37] and Priority Index[38]. However, the PCHi-C protocol used in these studies required
140 dozens of millions of input cells, precluding the analysis of rare cell types.
141
142 Here, we address this limitation by using a high-resolution and efficient PCHi-C protocol to
143 profile the *cis*-regulatory wiring of ILC3s isolated from human tonsils[30]. We detect promoter-
144 enhancer contacts in PCHi-C data using a combination of our established statistical
145 interaction-calling methodology (CHiCAGO)[39,40] and a newly developed adaptation of the
146 Activity-by-Contact[14,41] (ABC) approach to PCHi-C data that we term Activity-by-Captured-
147 Contact (ABCC). We develop a modified PCHi-C-aware GWAS gene prioritisation algorithm,
148 multiCOGS, that incorporates summary statistics imputation and multivariate statistical fine-
149 mapping, and use it to prioritise known and novel genes for CD through chromatin contacts.
150 Several of the genes are uniquely prioritised using PCHi-C data from ILC3s but not CD4+ T
151 cells, including the *CLN3* gene, mutations in which underpin ~80% of cases of the
152 neurodegenerative disorder Batten disease[42,43]. We show that this gene is downregulated
153 upon cytokine stimulation of mouse ILC3s, and *Cln3* overexpression in an ILC3-like mouse
154 cell line influences stimulation-responsive transcriptional programmes and cytokine
155 production. Finally, expanding multiCOGS to five additional autoimmune conditions, we
156 generate a catalogue of effector genes implicating ILC3s and show that they are enriched
157 among putative regulators of ILC3 inflammatory function. Together, our results shed light on
158 ILC3 cis-regulatory circuitries and prioritise autoimmune risk effector genes with potential roles
159 in this clinically important cell type.

160 Results

161 A compendium of promoter-anchored chromosomal contacts in human 162 ILC3s

163 To profile promoter-anchored chromosomal contacts in type 3 innate lymphoid cells (ILC3s),
164 we employed our low-input *DpnII*-based PCHi-C protocol[44,45] on ILC3s extracted from human
165 tonsils (**Fig. 1A**). Significant promoter contacts were detected with CHiCAGO[39] at a single-
166 fragment resolution, as well as after pooling the 'other end' fragments into ~5 kb bins, while
167 leaving the baited promoter-containing fragment unbinned (Methods)[40]. Using this approach,
168 we detected 31,003 contacts between promoters and promoter-interacting regions (PIRs) at
169 a single-fragment resolution and 58,632 contacts in 5 kb bins (**Fig. 1B**; **Table S1**; **Data S1-S2**
170 at https://osf.io/aq9fb). Binning resulted in the detection of longer-range contacts, as we
171 reported previously in other cell types[40] (**Fig. 1C, D**). A joint dimensionality reduction analysis[46]
172 of ILC3 promoter interaction profiles with those detected in 17 abundant blood cell types using

173 *HindIII*-based PCHi-C segregated ILC3s with other lymphoid cell types, consistent with the
174 notion that patterns of promoter interactions reflect the cells' lineage history[30] (**Fig. S1A**; see
175 Methods).
176
177 The increased resolution afforded by using *DpnII* in Hi-C library generation enabled capturing
178 alternative transcription start sites (ATSSs) for 6,789 genes located on separate *DpnII*
179 fragments. Remarkably, genes with captured ATSSs displayed distinct interaction landscapes
180 across isoforms (**Fig. S1B, C, D**). The three ATSSs of the *INPP4B* gene provide examples of
181 the multiple degrees of contact sharing across its 14 PIRs included in the analysis (**Fig. 1E**).
182
183 Next, we explored the epigenetic status of detected PIRs and compared the chromatin profile
184 of ILC3s with those of 88 other blood cell types detected by the Ensembl regulatory build[47].
185 As expected, at both fragment and 5-kb resolution ILC3 PIRs were enriched for markers of
186 accessible and/or active enhancers (ATAC, H3K27ac) and active transcription (H3K4me3),
187 based on public data in this cell type isolated from tonsils of pediatric donors[21] ("active PIRs",
188 **Fig. 1F**). Nearly half of all accessible and/or active ILC3 PIRs (47.8%, 8,718/18,231)
189 overlapped with annotated CTCF motifs or CTCF binding events in at least one of the
190 Ensembl-profiled cell types (**Fig. 1G**), consistent with the key role of CTCF in 3D chromosomal
191 organisation. However, only 3% of active/open regions in ILC3s (636/18,231) contained
192 Ensembl enhancer annotations[48], while nearly 20% of accessible and/or active PIRs
193 (3,411/18,231) did not have any functional annotations in the Ensembl data (**Fig. 1G)**.
194
195 We then considered the overlap of the active and/or accessible PIRs in ILC3s with those in 17
196 abundant blood cell types profiled with PCHi-C at *HindIII* resolution[30]. In contrast to chromatin
197 annotations, the majority of active/accessible PIRs in ILC3s also had promoter contacts in
198 these blood cell types (~80.4%, 12,409/15,435). Furthermore, ~60% of the active PIRs
199 (9,054/15,435) contacted the same gene promoters in both ILC3s and other blood cells **(Data**
200 **S3** at https://osf.io/aq9fb**)**. Consistent with previous observations, this result confirms that
201 patterns of promoter-enhancer contacts are more preserved across related lineages
202 compared with enhancer activity *in cis*[49]. We then probed the relationship between enhancer-
203 promoter connectivity and gene expression. For this, we integrated promoter-enhancer
204 interactions detected here with publicly available single-cell gene expression data (scRNA-
205 seq) in human mucosal tissue ILC3s[50]. In agreement with epigenetic studies in other cell
206 types,[30] we observed a significant positive correlation between the number of active and/or
207 open PIRs and gene expression (**Fig. S1E**).
208
209 Overall, our analysis provides a high-resolution compendium of promoter contacts in ILC3s,
210 including novel ILC3-specific regulatory elements and divergent contacts at ATSSs.

211 Inference of enhancer-promoter interactions using Activity-by-Captured-
212 Contact (ABCC) complements significant interaction detection

213 To further increase the sensitivity of detecting functional promoter-enhancer chromosomal
214 interactions from PCHi-C data, we adapted the Activity-by-Contact (ABC) approach[41] originally
215 developed for Hi-C. In contrast to CHiCAGO, which detects significant interactions relative to
216 a distance-dependent background, ABC considers any observed contact frequency between
217 a chromatin region and a promoter as potentially functionally meaningful, irrespective of

218 whether this frequency exceeds that expected by chance. In addition, while CHiCAGO scores
219 are independent of enhancer activity levels at the PIRs, ABC incorporates both contact
220 frequency and enhancer activity into the final metric ("ABC score")[41].
221
222 In our adaptation of ABC, which we term 'Activity-by-Captured Contact' (ABCC), we estimated
223 contact frequencies from imputed PCHi-C data, leveraging the statistical modelling of these
224 data produced by CHiCAGO for the imputation task (**Fig. 2A, Fig. S2A, S2B**, see Methods).
225 To validate the ability of the ABCC algorithm to detect functional enhancer-promoter pairs, we
226 took advantage of CRISPR interference (CRISPRi) enhancer perturbation data in K562 cells,
227 which was generated to validate the original ABC approach[14]. As inputs for ABCC, we used
228 public epigenetic annotations in K562 cells and our previously generated high-coverage PCHi-
229 C data in their physiological counterparts, erythroblasts[30]. These analyses demonstrated the
230 power of ABCC to predict functional enhancer-promoter links from lineage-relevant PCHi-C
231 and chromatin readouts (**Fig. S2C**). In contrast, using PCHi-C data from lymphoid cells at an
232 equivalent coverage reduced ABCC performance (**Fig. S2C)**. In addition, joint clustering of
233 the ABCC profiles generated for four primary blood cell types successfully reconstructed the
234 lineage relationships between them (**Fig. S2D**). These results highlighted the potential of
235 ABCC to infer lineage-specific *cis*-regulatory architecture. In comparison with CHiCAGO,
236 ABCC generally detected shorter-range promoter interactions, which was expected due to its
237 reliance on raw contact frequencies (**Fig. S2E**). Both ABCC- and CHiCAGO-detected contacts
238 were enriched for markers of accessible (DNase-seq) and/or active (H3K27ac) enhancers,
239 with regions called by both approaches showing the highest enrichment for these marks (**Fig.
240 S2F**). Taken together, these results suggest that ABCC and CHiCAGO detect complementary
241 subsets of regulatory promoter contacts.
242
243 Applying the ABCC algorithm to ILC3 PCHi-C data produced 18,877 putative enhancer-
244 promoter pairs across 17,690 genes (**Fig. S2F; Data S4** at https://osf.io/aq9fb). Similarly to
245 CHiCAGO-detected PIRs, there was a positive association between the number of ABCC
246 enhancers and gene expression **(Fig. 2B**). However, ABCC-detected interactions generally
247 spanned shorter distances than CHiCAGO-detected pairs (median distance ~69 kb vs ~108
248 kb, respectively, p-value < 2.2e-16, Wilcoxon rank-sum test) (**Fig. 2C**), and the two sets of
249 contacts showed only a limited overlap (8.4%; **Data S5** at https://osf.io/aq9fb). Nonetheless,
250 as expected, both CHiCAGO PIRs and ABCC enhancers were enriched for active and open
251 chromatin features, as well as CTCF binding sites and/or annotated motifs (**Fig. 2D**).
252 Representative examples of jointly detected regulatory landscapes are shown in **Fig. 2E**. We
253 combined ABCC- and CHiCAGO-detected promoter contacts for downstream analyses,
254 referring to them collectively as PIRs hereafter.
255

256 Comparative analysis of promoter interactomes between ILC3 and CD4+ 257 T cells identifies shared and differential regulatory circuitries

258 ILC3s share developmental similarities[51,52] and common "immune modules" with CD4+ T
259 cells[52–54], prompting us to use this abundant cell type for comparative analysis and
260 identification of ILC3-specific regulatory circuits. To this end, we generated and processed
261 high-resolution PCHi-C data for CD4+ T cells using the same protocol, identifying 31,252 and
262 87,348 interactions at single-fragment and 5 kb resolution, respectively (**Data S6** and **S7** at

263 https://osf.io/aq9fb). In addition, we detected 30,258 enhancer-gene pairs with ABCC across
264 16,956 genes (**Data S8** and **S9** at https://osf.io/aq9fb), 30% of which were shared with ABCC
265 pairs identified in ILC3s. Differential analysis of chromatin interactions between ILC3s and
266 CD4+ T cells with Chicdiff[55] revealed a total of 19,038 cell-type-specific interactions (1,818 at
267 fragment resolution and 17,220 at 5 kb resolution) across 3,664 genes (weighted adjusted p-
268 value <0.05) (**Fig. 3A**). As expected, we also detected a significant association between
269 differential interactions and differential expression (chi-squared = 23.938, df = 1, p-value =
270 9.948×10^{-7}) (**Fig. 3B**; **Data S10** at https://osf.io/aq9fb).
271
272 Genes with increased ILC3-specific chromatin contacts were enriched for annotation terms
273 such as "regulation of innate immune response," including *NFKB1* (NF-κB signaling), *TLR3*
274 (innate immune receptor), and *IFNG* (effector cytokine), and "regulation of immune effector
275 process", including *IL23R* (controlling ILC3 activation and cytokine production), *IL1R1*,
276 *TNFSF4*, and *SOCS5* (negative feedback on cytokine signalling) (**Fig. 3C; Fig. S3A; Table**
277 **S2**). In contrast, genes with CD4+ T cell-specific contacts were involved in "regulation of T cell
278 activation" (e.g. *CD3E, CD86, CTLA4, IL6, FOXN1*) and "negative regulation of the MAPK
279 cascade" (e.g. *DUSP14, DUSP16, PTPN6*) (**Fig. 3C; Fig. S3B; Table S3**).
280
281 We also identified 194 genes with differential contacts between ILC3s and CD4+ T cells,
282 including *BCL2, FYN, CD226* (activating receptor on T and NK/ILC3-like cells), and *CCR7*
283 (guiding ILC3 positioning and migration) (**Fig. 3C; Fig. S3C; Table S4**). Notably, many genes
284 with ILC3- and/or CD4+ T cell-specific contacts converged on pathways such as TCR
285 signalling and T cell activation (e.g. *IL23R, RORC, NFKB1, CD300A, PIK3R1, ZAP70, CTLA4,*
286 *CD3E, CD226, ITK, CD28, CCR7*), indicating differences in the regulatory wiring of these
287 genes in ILC3s and their adaptive immune counterparts. In contrast, genes with similar contact
288 profiles across both cell types were associated with processes such as histone modification,
289 chromatin remodelling, and lymphocyte proliferation and differentiation (**Fig. S3; Table S5**),
290 reflecting their shared functionality in both cell types.
291
292 In conclusion, our comparative chromosomal interaction analysis highlights both shared and
293 distinct regulatory wiring of ILC3s and CD4+ T cells, reflecting their specialised roles in innate
294 versus adaptive immune responses and coordinated regulation of immune activation
295 pathways.

296 Promoter-interacting regions in ILC3s and CD4+ T cells are enriched for 297 genetic variants associated with autoimmune disorders

298 Genetic risk variants for complex diseases are strongly enriched at transcriptional
299 enhancers[14–16]. Therefore, we investigated whether regulatory elements interacting with gene
300 promoters in ILC3s and CD4+ T cells were enriched for genetic susceptibility to human traits
301 and diseases, using the RELI algorithm[56] (**Fig. 4A**; see Methods). Briefly, RELI determines
302 significantly enriched overlaps between selected genomic loci (here, promoter-interacting
303 regions intersecting open chromatin or H3K27ac signals in ILC3s based on public data) and
304 trait-associated genetic variants. This is done by comparing the observed overlaps with a null
305 distribution of artificially created variant sets with similar linkage disequilibrium (LD)
306 characteristics to the trait-associated variants[56]. A practical advantage of RELI over the

307 commonly used stratified LD score regression[57] is that it does not require summary statistics
308 data and can be performed on sets of significant SNPs reported in the GWAS Catalog[58].
309
310 Out of the 495 analysed traits and diseases tested from the GWAS Catalog, genetic risk loci
311 for 21 human traits were significantly enriched at promoter-linked putative regulatory elements
312 in ILC3s (BH adjusted p-value < 0.05; **Fig. S4A**, **Table S6**; see Methods). Autoimmune
313 diseases were overrepresented among these traits (according to the ontology EFO:0005140;
314 p-value = 1.077×10^{-5}, hypergeometric one-tailed test), affecting a broad array of organs and
315 tissues that ILC3s are known to reside in. These included the gut (CD, celiac disease,
316 ulcerative colitis, primary sclerosing cholangitis), airways (asthma, hay fever), and the central
317 nervous system (multiple sclerosis). We also noted several traits of peripheral blood cells,
318 including platelet width, lymphocyte count, and corpuscular volume (**Table S6**).
319
320 In CD4+ T cells, 22 traits were significantly enriched at promoter-interacting regulatory
321 elements of CD4+ T cells (BH adjusted p-value < 0.05; **Fig. S4B**), with significant correlation
322 between the two cell types (R^2 = 0.845822, df = 10, 95% CI (0.5284, 0.9558), p = 0.00052;
323 **Fig. 4B**), in line with the assumption that CD4+ T cells and ILC3 cells share many cis-
324 regulatory circuits. However, several traits displayed cell-type specificity, such as allergic
325 sensitisation, mouth ulcers, and IgG glycosylation in ILC3s, and primary biliary cirrhosis,
326 rheumatoid arthritis, and systemic lupus erythematosus in CD4+ T cells (**Table S6**).
327
328 Among the autoimmune disorders, CD risk variants were particularly highly enriched within
329 the active PIRs of both ILC3s and CD4+ T cells (~2.3-fold enrichment in both cell types, p-
330 value = 1.41×10^{-8} in ILC3s and p-value = 2.41×10^{-10} in CD4+ T cells). We confirmed this
331 observation using stratified LD score regression (**Fig. 4C**). While the critical role of CD4+ T
332 cells in CD is well-established[59–62], the connection between ILC3s and CD pathogenesis is
333 more recent. ILC3s are thought to influence inflammatory processes in CD, such as GM-CSF
334 signalling and overexpression of the cytokines IL-22, IL-17, and IFN-γ[11,63]. We next sought to
335 leverage PCHi-C data to prioritise genes linked to CD risk variants in these cell types.
336

337 MultiCOGS prioritises genes linked to Crohn's disease risk based on 338 multivariate fine-mapping of imputed GWAS signals and promoter 339 contacts in ILC3 and CD4+ T cells

340 To identify putative causal variants and genes for CD in ILC3s and CD4+ T cells, we extended
341 our previously published Bayesian prioritisation algorithm, COGS[30,31], which provides a single
342 measure of support ("COGS score") for each gene's association with a trait of interest,
343 calculated based on the location of fine-mapped GWAS signals within (i) gene coding regions,
344 (ii) gene promoters, and (iii) promoter-interacting regions.
345
346 Despite its demonstrated utility in prioritising gene candidates in a range of human
347 traits[30,31,64,65], we identified areas for improvement in COGS. First, if the summary statistics
348 underlying the trait-associated loci are too sparse, COGS may miss likely causal variants
349 intersecting promoter-interacting regions. To mitigate this, we imputed additional trait-
350 associated variants using an established summary statistics-based methodology[66]. Second,
351 the original statistical fine-mapping approach utilised in COGS assumes at most a single

352 causal variant per linkage disequilibrium (LD) block, whereas the latest evidence suggests
353 that trait-associated LD blocks can contain multiple causal variants[67]. To address this, we
354 updated the COGS algorithm to enable integration with recently developed multivariate fine-
355 mapping approaches, such as SuSiE[68–70] (**Fig. 4D**; see Methods). Finally, we accounted for
356 both CHiCAGO- and ABCC-detected promoter-interacting regions. We refer to the updated
357 version of COGS as "multiCOGS".
358
359 We ran multiCOGS on the CD GWAS meta-analysis by de Lange *et al.*[71] using the
360 compendium of CHiCAGO- and ABCC-detected promoter-interacting regions in ILC3s or
361 CD4+ T cells. At the previously established COGS score cutoff of 0.5[30], we prioritised 109
362 genes in ILC3s (**Fig. 4E**) and 118 genes in CD4+ T cells (**Fig. S5A**; **Table S7**). The majority
363 of genes were prioritised based on 3D proximity of non-coding trait-associated variants to gene
364 promoters, either by PCHi-C or ABCC (**Fig. S5B**). ABCC contributed to around 11% of the
365 prioritised genes in both cell types (**Fig. S5C**). At first examination, we noted many candidate
366 genes with roles in immune processes already known to be dysregulated in inflammatory
367 bowel disease (IBD)[72–74]. Examples include cytokine signalling (*IL10*, *IL1RL1, LTBR, IL2RA,*
368 *IFNGR2, TNFSF8*), autophagy (*ATG16L1, GPR65*), and antimicrobial processes in the gut
369 (*PTPN2, IRF8*)[75,76]. The prioritised genes also highlighted IL-23/Th17 signalling (for example,
370 *RORC, NFKB1, IL2RA,* and *TYK2*), a known immune axis in CD pathology[77], and known
371 transcriptional regulators (*FOS*, *TSC22D1*, *RBPJ*). In several loci, multiCOGS prioritised
372 several compelling gene candidates, based on multiple credible sets. For example, in ILC3s,
373 two credible sets of variants in chr7p implicated the *IKZF1* gene (encoding the Ikaros
374 transcription factor) by PCHi-C interactions, and the *DDC* gene (encoding dopamine regulator
375 L-dopa decarboxylase) by ABCC pairing (**Fig. S5D**). Ikaros, an established critical regulator
376 of immune cell development[78], also scored highly in the original COGS algorithm. However,
377 the more distal *DDC* gene, which has recently emerged as a potential regulator of immune
378 cell infiltration[79], scored well below the prioritisation threshold (**Table S7**). This demonstrates
379 the potential of multiCOGS and ABCC for highlighting previously missed gene candidates.[79]
380
381 We next explored more closely how the results of multiCOGS compared with those from our
382 previously published COGS pipeline, which used univariate fine mapping without imputation
383 and was based purely on CHiCAGO results without ABCC (hereafter referred to as "classic
384 COGS"). Classic COGS resulted in substantially smaller prioritised gene sets (55 genes in
385 ILC3 cells and 75 genes in CD4+ T cells with COGS score > 0.5) (**Table S7**). As examples,
386 we note that compelling candidate genes such as *IL12RB2* and *IL15RA* (in ILC3s), *TNFSF15,*
387 and *ICAM3* (in CD4s), and *NFKB1*, *BATF*, *ICAM1* and *TNFSF8* (in both cell types) were only
388 prioritised in multiCOGS (**Table S7**). Moreover, we discovered that both of the novel aspects
389 of multiCOGS (imputation and multivariate fine mapping) contributed substantially to the
390 increased number of genes prioritised in comparison with classic COGS (**Fig. S6A**). For the
391 majority of genes, multiCOGS prioritisation scores were similar or higher than in conventional
392 COGS in both ILC3s and CD4s (**Fig. S6B**). Only five genes prioritised by conventional COGS
393 had sub-threshold scores in multiCOGS, including *JAK2* (see **Fig. S6C** and **Supplementary**
394 **Note 1**).
395
396 Next, we searched for prior evidence of association of all multiCOGS-prioritised genes with
397 CD (or IBD, more broadly) by querying the top CD genes in OpenTargets, curated gene-to-
398 disease databases, and functional studies[37,80–83]. We found that over half of multiCOGS-

399 prioritised genes in ILC3s (61/109) and CD4s (67/118) were not previously implicated in these
400 databases (**Table S8**). These newly prioritised genes included compelling candidates such as
401 ubiquitin-specific peptidase 49 (*USP49*), adding to the existing evidence for the role of protein
402 ubiquitination in IBD development[84], and lymphotoxin beta receptor (*LTBR*), known to be
403 important for gut epithelial cell IL-23 production[85]. In particular, 23 genes selectively prioritised
404 in ILC3s **(Fig. 4F)** were not previously linked to CD in the studied datasets. These included
405 genes with unexpected functions, such as the neurotransmitter DOPA decarboxylase (*DDC*),
406 and a lysosomal/endosomal transmembrane protein (*CLN3*). *CLN3* is involved in lipid
407 trafficking and catabolism[86,87], and mutations in this gene cause Batten disease, a group of
408 lysosomal storage disorders characterised by progressive neurodegeneration[88].
409
410 Taken together, by accounting for imputed variants and multiple causal variants per locus,
411 multiCOGS expands the ability to discover candidate genes in complex trait loci using
412 promoter interactions.

413 Prioritised gene candidates in ILC3 cells implicate inflammatory 414 processes in CD aetiology

415 We explored the biological functions of the 109 prioritised CD genes in ILC3s based on their
416 public gene set annotations (**Table S9**). Seven biological states or processes were
417 significantly enriched among the gene candidates: IL6-JAK-STAT3 signalling, TNFα signalling
418 via NFκB, IL2-STAT5 signalling, inflammatory response, allograft rejection, IFNγβ response,
419 and TGFβ signalling (Hallmark gene sets; **Fig. S7A**). Molecular functions included cytokine
420 receptor activity and NAD+ metabolic activity (GO Term Molecular Functions, **Fig. S7B).** We
421 saw the strongest enrichment of cell-type signatures for tissue-resident immune cells,
422 including gastric and duodenal immune cells, as well as monocytes, dendritic cells, and
423 basophils in the lung (**Fig. S7C**). We also noted the signature for ILC progenitor cells in fetal
424 lung[89], driven by the genes *IL1R1, ICAM1, IFNGR2, PLCG2, CCR6,* and *RORC* (adjusted p =
425 0.0176). Enriched curated pathways highlighted immune-mediated diseases, including
426 rheumatoid arthritis, neuroinflammation, IBD, and bacterial infection (WikiPathways; **Fig.
427 S7D**). Other relevant pathways included T cell differentiation and signalling of IL-18, a key
428 cytokine for ILC3 function[90] (**Fig. S7C**). Leveraging published IBD patient gene sets[91], we also
429 found enrichment for genes differentially expressed in the rectum in patients with CD (adjusted
430 p-value = 1.38×10^{-4}) and ulcerative colitis (adjusted p-value = 0.0156) (**Table S9**, **Fig. S7E**).
431
432 We then investigated which transcription factors (TFs) might regulate the CD gene candidates
433 in ILC3s using two methods. First, we used a gene-centric approach to identify
434 overrepresented genes predicted to be targeted by a given TF (TF targets from MSigDB). This
435 analysis highlighted the architectural protein HMGA1 and the known inflammatory response
436 regulator NFκB (**Table S9** and **Fig. 5A**). Second, we used a region-centric approach,
437 searching for enrichment of predicted TF binding sites across a range of cell types at the PIRs
438 of CD candidate genes in ILC3s. We found significant enrichment for 97 TFs (**Fig. 5B, Table
439 S10**), many of which were previously implicated in inflammatory response, including
440 IKZF1/Ikaros[92], BATF[93], and NFKB3/RELA[94–96], which are all highly expressed in ILC3s **(Fig.
441 5C)** and have established roles in ILC3 biology. Two examples of potential long-range
442 regulation of CD candidate genes by putative TF binding at PIRs are shown in **Fig. 5D** and **E**.
443 In the first example, the promoter of the *IKZF1* gene contacts two upstream PIRs, each

444 containing a separate credible set of fine-mapped CD susceptibility variants and bearing
445 marks of open and active chromatin (ATAC-seq and H3K27ac peaks) in ILC3s. Based on data
446 from lymphoblastoid cell lines, these PIRs recruit multiple TFs: IKZF1 itself, as well as BATF,
447 NFKB3, ATF2, and the architectural proteins CTCF and SA1 (**Fig. 5D**). In the second example,
448 the promoter of *IL1R1* contacts CD risk variant-containing PIRs that have accessible
449 chromatin in ILC3s and contain CTCF binding signals in lymphoblastoid cell lines (**Fig. 5E**).
450
451 Jointly, these results propose inflammatory signalling genes as causal candidates for CD
452 susceptibility in ILC3s.
453

454 *CLN3* contributes to ILC3 inflammatory capacity

455 We next focused on *CLN3,* a gene implicated in the neurodevelopmental disorder Batten
456 disease. *CLN3* was selectively prioritised as a CD risk gene in ILC3s, but not CD4+ T cells,
457 and has not previously been linked to CD or other immune-mediated diseases. Examination
458 of the SuSIE fine-mapped CD GWAS locus underlying *CLN3*'s prioritisation revealed a
459 credible set of variants overlapping two regions considered by multiCOGS. The first region is
460 an ILC3-specific *CLN3* PIR located 14.2 kb downstream of the canonical *CLN3* TSS (red band
461 in **Fig. 6A**). The second region lies between exons 10 and 11 of the canonical *CLN3* transcript,
462 adjacent to an annotated internal promoter (first dark blue band in **Fig. 6A**). Unexpectedly, we
463 found that both regions lacked chromatin accessibility and enhancer activity signals in ILC3s,
464 as well as in all other cell types included in the Ensembl Regulatory Build database (**Fig. 6A**).
465 Data from lymphoblastoid cell lines[97] showed enrichment for the H3K36me3 mark, which is
466 typically associated with transcriptional elongation[98] and facultative heterochromatin[99] (**Fig.
467 6A**). To seek complementary evidence for a regulatory role of this locus, we queried the
468 OpenTargets database[100] for possible colocalisation between the CD risk signal and known
469 *CLN3* expression quantitative trait loci (eQTLs). CD risk GWAS and *CLN3* expression were
470 likely to share a joint causal genetic signal (posterior probability ≥ 0.8, as determined by
471 coloc[101] and reported in OpenTargets) in whole blood[102,103], monocytes[104,105], thyroid[103], small
472 intestine[103], and cerebellum[103]. Notably, the same CD GWAS signals also colocalised with
473 eQTLs for nearby genes, including *APOBR*, which is located ~2 kb downstream of *CLN3* in a
474 divergent orientation, suggesting a complex regulatory architecture at this locus.

475 To further investigate the role of the *CLN3* locus in ILC3s, we used mouse MNK-3 cells as a
476 tractable model for ILC3 activation and effector function. We found that *Cln3* expression was
477 downregulated upon stimulation of MNK-3 cells with IL-23 and IL-1β, cytokines that are
478 essential for ILC3 effector function[106,107] (**Fig. 6B,** left). Consistent with this observation,
479 analysis of published RNA-seq data from primary mouse ILC3s stimulated with TL1A[108] also
480 showed reduced *Cln3* expression (**Fig. 6B,** left). Notably, the adjacent gene *Apobr* was
481 similarly downregulated under IL-23/IL-1β stimulation (**Fig. 6B**, right), in line with eQTL-based
482 evidence of coordinated regulation of these genes in humans.[102,103,104,105] In contrast, TL1A
483 stimulation did not affect *Apobr* expression (**Fig. 6B**, right).
484
485 To interrogate the transcriptional consequences of stimulation-induced *Cln3* repression, we
486 used CRISPR activation (CRISPRa; dCas9-VP64 + MS2-p65-HSF1) to prevent *Cln3*
487 downregulation in MNK-3 cells during stimulation. CRISPRa targeting produced an
488 approximately threefold increase in *Cln3* expression in stimulated MNK-3 cells (**Fig. 6B**, left).
489 Notably, *Apobr* expression was also increased in both basal and stimulated conditions (**Fig.**

490 **6B**, right), potentially reflecting local effects of CRISPRa targeting, but also mirroring the
491 coordinated regulation observed at this locus (**Fig. 6B**). Bulk RNA-seq analysis revealed
492 widespread transcriptional changes following *Cln3* CRISPRa, with 519 differentially expressed
493 genes in unstimulated cells and 722 in stimulated cells relative to scrambled gRNA controls
494 (DESeq2 adjusted p-value < 0.05; **Fig. 6C** and **S8A; Table S11** and **Data S10** at
495 https://osf.io/aq9fb). These genes were enriched for pathways involved in
496 lymphocyte differentiation, activation, and proliferation, including upregulation of *Cd23r*, *Cd74*,
497 and *Fas*, and downregulation of the inflammatory serine proteases *Gzmb* and *Gzmc* (**Fig. 6D,**
498 **Fig. S8B**). Notably, more than half of the genes differentially expressed in stimulated *Cln3*-
499 CRISPRa cells overlapped with genes altered by IL-23/IL-1β or TL1A stimulation in *Cln3*-
500 unperturbed cells[108] (**Fig. 6E**), suggesting that sustained *Cln3* expression counteracts
501 canonical activation-associated transcriptional programmes. In contrast, CRISPR interference
502 (CRISPRi; dCas9-KRAB)-mediated knockdown of *Cln3* resulted in few transcriptional changes
503 beyond *Cln3* and *Apobr* themselves **(Fig. S8C, D; Table S11; Data S11** at
504 https://osf.io/aq9fb). Notably, these included upregulation of *Nos2,* a gene previously
505 implicated in limiting ILC3-driven intestinal inflammation[107].
506
507 Given the coordinated regulation of *Cln3* and *Apobr* expression upon ILC3 stimulation, the
508 limited transcriptional impact of further *Cln3* knockdown in activated cells, and the pronounced
509 effects of *Cln3* overexpression, we next asked whether the CLN3 protein modulates ILC3
510 effector function at a post-transcriptional level. CLN3 is a lysosomal and endosomal protein
511 with established roles in vesicular trafficking, lysosomal homeostasis, and protein
512 turnover[109,110,111], processes that are central to cytokine storage and secretion. Therefore, we
513 ectopically overexpressed *Cln3* in MNK-3 cells and measured cytokine secretion under basal
514 and inflammatory conditions. Overexpression of the myc-tagged CLN3 construct was
515 confirmed by RT-qPCR and immunoblotting (**Fig. S8E**, **F**). As expected, MNK-3 cells
516 constitutively secreted IL-22 and GM-CSF, with further induction of these cytokines upon
517 stimulation, whereas IL-17 production was restricted to stimulated conditions (**Fig. 6F** and
518 **S8G**). Notably, CLN3 overexpression significantly reduced the secretion of IL-17, IL-22, and
519 GM-CSF by stimulated MNK-3 cells (**Fig. 6F** and **S8G**). Basal IL-22 and GM-CSF secretion
520 were also reduced in the absence of stimulation (**Fig. 6F** and **S8G**). Viable cell numbers were
521 quantified at the end of cytokine secretion assays and showed no difference under basal
522 conditions, with a modest reduction in Cln3-overexpressing cells following stimulation (**Fig.**
523 **S8H**).
524
525 Collectively, these results highlight the Batten disease gene *Cln3* and the broader *Cln3/Apobr*
526 locus as regulators of ILC3 inflammatory output, revealing a previously unrecognised role for
527 this locus in shaping ILC function.

528 MultiCOGS prioritises candidate genes for six autoimmune diseases with
529 potential roles in ILC3 inflammatory function

530 Building on the methodologies and data generated in this study, we extended multiCOGS
531 analysis in ILC3s and CD4+ T cells to five other autoimmune GWAS datasets in addition to
532 CD with available summary statistics that showed enrichment at ILC3 PIRs in the RELI
533 analysis: adult-onset asthma, IBD, ulcerative colitis (UC), primary sclerosing cholangitis (PSC)
534 and celiac disease. Across the six traits and two cell types, we detected a total of 332

535 prioritised disease candidate genes (multiCOGS score > 0.5), of which 251 were prioritised in
536 ILC3 cells (**Fig. 7A**) and 266 in CD4+ T cells (**Table S12**). As expected from their shared
537 aetiology, the three traits relating to inflammatory bowel disease (CD, UC, and IBD) clustered
538 together with respect to gene-level multiCOGS scores, while asthma formed an outgroup (**Fig.**
539 **7A**).
540
541 A total of 66 candidate genes were prioritised in ILC3s only, and 81 in CD4+ T cells only (**Table**
542 **S12)**. Notable ILC3-specific candidate genes included several cytokines and receptors
543 involved in type I immune response, such as *CCR2* (celiac disease), *BCL6* and *IL17A* (both
544 asthma), as well as the IL-18 receptor (*IL18R1*), which we previously prioritised for CD, and
545 here also prioritised for celiac disease and asthma. We also noted family members of
546 butyrophilin (BTN) proteins–immunomodulatory transmembrane proteins involved in
547 recognition of microbial antigens–prioritised in both CD and asthma (*BTN3A1* and *BTN3A2*),
548 specific to ILC3 cells. Finally, we noted that *CLN3* was prioritised for the broader IBD trait
549 (multiCOGS score 0.538, **Table S12**) in addition to CD, again selectively in ILC3s.
550
551 Pathway analysis of the prioritised genes across the analysed traits revealed shared enriched
552 GO terms for inflammatory processes such as cytokine binding and immune receptor activity
553 (**Fig. 7B; Table S13A**). To gain further insight into the role of the prioritised genes in ILC3
554 inflammatory function, we turned to a recent CRISPRi screen for putative regulators of IL-22
555 expression in MNK-3 cells following IL-23/IL-1β stimulation[112] (**Fig. 7C**). Of the multiCOGS
556 gene candidates across all profiled autoimmune diseases, six were significant positive
557 regulators and five were significant negative regulators of IL-22 protein production, as detected
558 by the CRISPRi screen (**Table S13B** and labelled in red in **Fig. 7D** and **7E**). Among the IL-22
559 activators were three candidate genes for IBD-related traits, all with known strong roles in IL-
560 22 activation (*IL23R*, *BATF,* and *RORC*). The remaining three IL-22 activators were all
561 candidate genes for asthma alone: *GNA15*, *SESN1,* and *GATA3*, of which only *GATA3* has
562 been previously reported to directly activate IL-22 in ILC3s[113]. Meanwhile, the five multiCOGS
563 genes putatively downregulating IL-22 production were all associated with IBD-related traits
564 (*PTPN2, NEMF, HCLS1, PPP5C, and KXD1*). Of these, only *PTPN2* has direct evidence for
565 IL-22 repression, through STAT3 dephosphorylation[114]. The other putative IL-22 negative
566 regulators have diverse functions in protein homeostasis (*NEMF*), actin remodelling (*HCLS1*),
567 stress signalling (*PPP5C*), and lysosome localisation (*KXD1*). Overall, multiCOGS genes were
568 significantly enriched among the genes scoring highly for positive IL-22 regulation (GSEA p =
569 0.0284, **Table S13C**; genes driving the association labelled in **Fig. 7D**), implicating the control
570 of ILC3 activation as an important mechanism underpinning the effects of the prioritised genes
571 on autoimmune disease risk.
572
573 In summary, this analysis expands the compendium of prioritised GWAS gene candidates with
574 potential roles in ILC3s to six autoimmune disease traits and demonstrates the potential role
575 of many prioritised genes in ILC3 inflammatory function.

576

577 # Discussion

578 In this study, we present high-resolution promoter interaction profiling in ILC3s, revealing tens
579 of thousands of promoter contacts with enhancers and GWAS variants associated with
580 multiple immune diseases, including those that are unique to ILC3s compared with their
581 phenotypically related counterparts in the adaptive immune system, CD4+ T cells. ILC3s are
582 a relatively rare cell type that cannot be easily expanded *in vivo*, which makes their
583 chromosomal interaction profiling challenging. Indeed, this problem precluded ILC3 profiling
584 by standard Hi-C alongside type 2 ILCs in a recent mouse study[115]. Robust Capture Hi-C
585 profiling typically requires even higher cell numbers. Our efficient PCHi-C protocol[44] and the
586 use of a four-cutter enzyme (*DpnII)* have enabled a higher-resolution analysis of human ILC3s
587 in this study, adding these clinically-relevant cells to the ever-expanding array of cell types
588 with available promoter interactome maps, including the 17 abundant blood cell types that we
589 profiled previously using high-coverage PCHi-C at a six-cutter enzyme (*HindIII)* resolution[30].
590 While emerging technologies provide complementary solutions for the inference of enhancer-
591 promoter relationships, such as through the correlated activities of these elements across cell
592 types or single cells, genetic evidence and high-throughput perturbation screens, 3D
593 genomics-based approaches continue to offer unique advantages by delivering
594 mechanistically-grounded information in high throughput at a reasonable cost and time
595 investment.
596
597 Unlike in our previous studies, here we take advantage of two conceptually different
598 computational analysis strategies for detecting promoter contacts from Capture Hi-C data. The
599 first strategy is based on our established CHiCAGO pipeline to detect 'significant contacts' –
600 i.e., those whose frequency significantly exceeds the expectation at a given distance and
601 technical noise levels. The second strategy is based on the adaptation of the ABC
602 approach[14,41] to Capture Hi-C data (the Activity-by-Captured-Contact method, ABCC), which,
603 in contrast, considers the raw contact frequency rather than its significance. As expected from
604 this conceptual difference, ABCC prioritises shorter-range contacts compared with CHiCAGO,
605 resulting in the largely non-overlapping sets of identified contacts and GWAS-prioritised
606 genes. However, the longer-range contacts detected using CHiCAGO, which were also
607 enriched for active enhancers, drive the majority of our identified disease associations. From
608 the practical point of view, therefore, these two approaches are largely complementary, and
609 their combined use is warranted. Mechanistically, this suggests that at short linear distances,
610 the background frequencies of promoter-enhancer contacts arising from constrained Brownian
611 motion are sufficient for the functional interactions between these regions. In contrast, at
612 longer ranges, additional factors (e.g., cohesin-mediated loops) are likely required to facilitate
613 the statistically unusual contact frequencies and enable functional interactions.
614
615 We find a strong enrichment for CD-associated SNPs within the ILC3 PIRs, consistent with
616 recent findings showing that superenhancers specific to ILC3 or Th17 cells, rather than to ILC1
617 or Th1 cells, preferentially contain CD-associated variants[21]. Using our multiCOGS strategy
618 that integrates GWAS data processed with multivariate statistical fine-mapping with
619 information on enhancer-promoter links from PCHi-C, we prioritise a total of 109 genes in
620 ILC3s, 29 of which are not detected in CD4+ T cells. Notably, the number of multiCOGS-

621 prioritised genes has increased considerably compared with the results obtained with our
622 previously developed COGS pipeline[30,31]. The key improvements of multiCOGS include
623 summary statistics-based imputation and allowing for multiple causal variants per linkage
624 disequilibrium (LD) block. At the molecular level, the increased recall of prioritised genes
625 reflects the fact that the same LD block often contains multiple regulatory elements (including
626 promoter-proximal and distal enhancers). Variants within each of these elements may have
627 largely independent effects from one another[49,67] and from those within protein-coding
628 regions[116]. Furthermore, we identify cases, such as *IKZF1/DDC*, where multiple causal
629 variants in the same LD block intersect the regulatory elements of different candidate genes,
630 leading to their joint prioritisation. These results reinforce the notion that the assumption of a
631 single causal variant per LD block used by many established GWAS analysis methods
632 (particularly those based on summary data) is unnecessarily restrictive and may miss key
633 genetic mechanisms underpinning disease processes.
634
635 While the enrichment of GWAS signals within enhancers was first demonstrated over a
636 decade ago[16], with the first studies leveraging 3D information for enhancer-gene assignment
637 following shortly thereafter[117–119], the majority of GWAS gene prioritisation studies to date still
638 do not consider 3D chromosomal data[120]. Nonetheless, several computational approaches for
639 variant-to-gene assignment integrating fine-mapped GWAS signals with 3D genomics
640 information and other sources of evidence are now becoming available. For example, FUMA
641 SNP2GENE provides the option to identify candidate genes via enhancer-promoter
642 interactions, but does not integrate fine-mapping SNP probabilities[121]. In addition, the L2G
643 (locus-to-gene) pipeline uses a machine learning algorithm that integrates multiple features,
644 including Capture Hi-C[122]. L2G provides an interpretable output that shows the relative
645 contributions of many factors, including QTL colocalisation, genomic distance, VEP scores[123],
646 and enhancer-promoter interactions, towards an overall gene score per credible set. L2G is
647 available on the OpenTargets platform[37], but it is not easily adaptable to new functional data.
648 Finally, H-MAGMA incorporates Hi-C-derived chromatin interactions to refine SNP-to-gene
649 assignment for non-coding GWAS variants, but does not integrate them into a probabilistic
650 framework[124]. MultiCOGS complements these efforts by providing an unsupervised and
651 interpretable Bayesian framework based on cell-type-specific, mechanistically-grounded
652 readouts that can be applied to 3D genomic data in cell types relevant to the disease context.
653
654 Using multiCOGS across six autoimmune traits to prioritise disease risk-linked genes with
655 potential roles in ILC3s, we produce a compendium of 251 genes, including both known and
656 potentially novel candidates. Integration with a CRISPRi screen for genes affecting ILC3
657 inflammatory response provides a first indication of their potential role in ILC3 biology. This
658 includes 11 prioritised genes that were detected as putative IL-22 activators and repressors
659 in the CRISPRi screen[125]. However, further targeted experiments are still required to gain a
660 deeper understanding of the functional role of the prioritised genes in ILC3 biology and their
661 contribution to autoimmune disease risk.
662
663 The *Cln3* gene, prioritised in our analysis for CD risk in ILC3s but not in CD4+ T cells, underlies
664 the majority of cases of the neurodevelopmental disorder Batten disease. While immune
665 features have been reported in Batten disease and other lysosomal disorders[126,127], the
666 function of *Cln3* in the immune system remains poorly understood. Here, we show that *Cln3*
667 expression is downregulated upon cytokine stimulation of mouse ILC3s, and that *Cln3*

overexpression in an ILC3-like mouse cell line impacts stimulation-induced transcriptional programmes and cytokine production. In contrast, CRISPRi knockdown of *Cln3* did not show a pronounced phenotype in our model system, and, consistent with this, was not detected as a significant hit in the CRISPRi screen for regulators of ILC3 inflammatory response.[125] CLN3 is a transmembrane lysosomal protein with established roles in vesicular trafficking and lysosomal homeostasis[128]. Consistent with this biology, our functional data support a role for activation-induced downregulation in promoting the inflammatory capacity of ILC3s. In addition to its trafficking functions[109,110,111], recent studies have demonstrated that CLN3 is required for the catabolism of glycerophospholipids[87,129], which are key structural components of cellular membranes and have emerging regulatory roles in innate immune signalling. Accordingly, *Cln3* knockdown in mouse monocytes was shown to interfere with LPS-induced secretion of the inflammatory cytokine IL-6[,130]. These observations raise the possibility that CLN3 may influence immune effector functions through effects on membrane composition, vesicular dynamics, or both. Together, our findings implicate CLN3 in the regulation of ILC3 inflammatory function and CD risk, raising the possibility that inflammatory processes may contribute to gastrointestinal manifestations observed in CLN gene deficiency.[131]

Notably, the region harbouring the fine-mapped CD susceptibility variants in the *CLN3* locus lacks active chromatin signals in ILC3s, as well as in other cell types represented in the Ensembl Regulatory Build. This suggests that regulatory activity at this locus may be highly context-specific, potentially emerging only under inflammatory conditions or within discrete cellular states. Supporting this notion, H3K36me3 deposition across this region in lymphoblastoid cell lines was recently proposed as a mark of enhancers that are 'poised' for rapid activation[132]. However, CD-associated variants in this locus may also exert regulatory effects through alternative mechanisms. Several fine-mapped variants in the *CLN3* locus are linked to alternative polyadenylation of the *CLN3* transcript's 3'UTR across multiple tissues[133,134], a mechanism that can influence mRNA stability and translational efficiency and is increasingly recognised as a contributor to complex disease risk[134]. In addition, C*LN3* was reported to undergo splicing-dependent transcriptional activation[135], further expanding the range of potential regulatory mechanisms operating at this locus. The regulatory complexity of the *CLN3* locus is further augmented by its detection as an eQTL for multiple neighbouring genes across diverse cell types. In monocytes, this locus is also an eQTL for the known CD gene *IL27,* with an opposite direction of allelic effect and a lower statistical significance relative to *CLN3* itself[105,136]*.* Notably, *IL27* is not appreciably expressed in either mouse or human ILC3s. In addition, *CLN3* shares eQTLs with, and is divergently expressed from, the apolipoprotein B receptor gene *APOBR*. Consistent with this, we show that *Cln3* and *Apobr* are co-regulated upon IL-23/IL-1β stimulation in a mouse ILC3-like cell line. APOBR has a recognised role in lipid uptake in myeloid cells[137], but its function in the lymphoid compartment remains unclear and is likely mechanistically distinct from that of CLN3*.*

Human ILC3s in our study are derived from tonsillectomy material, but their regulatory elements show 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[138,139]. Furthermore, ILC3s from regularly inflamed tonsils have a closer cytokine profile to mucosal-resident ILC3 populations than ILC3s from resting lymph nodes or peripheral blood[140]. Focused studies in relevant physiological contexts

715 and disease models will further establish the role of ILC3s in mediating the effects of genetic
716 variation. These analyses are, however, complicated by the rarity of ILC3s and a lack of robust
717 human cell line models for this cell type, as well as the strong influence of organismal and
718 environmental factors, which are difficult to reproduce in a laboratory setting either *in vitro* or
719 *in vivo,* on autoimmune disease pathogenesis.

720 In conclusion, we present updated methodologies for profiling and detecting promoter-
721 anchored interactions and for leveraging these data to interpret GWAS signals. Using this
722 framework, we provide a comprehensive catalogue of regulatory chromatin contacts and
723 candidate autoimmune risk genes in ILC3s, and take initial steps toward their functional
724 validation. These findings advance our understanding of ILC3 biology and the contributions of
725 this rare cell type to disease, and highlight the utility of our approach for dissecting regulatory
726 architecture in other rare cell types and complex traits.

727 Methods

728 Human ILC3 cell isolation

729 Three children requiring tonsillectomy were recruited to a prospective study at a tertiary
730 academic care centre through the division of Pediatric Otolaryngology-Head and Neck Surgery
731 at Cincinnati Children's Hospital Medical Center with an institutional review board (IRB)
732 approval. Criteria for enrollment in the study included a history of sleep-disordered breathing
733 or recurrent or chronic tonsillitis requiring removal of the tonsillar tissue. Consent was obtained
734 from parents in the perioperative suite on the day of the procedure. Subjects were excluded
735 from the study if the tonsillar tissue was acutely infected or if anatomic abnormalities were
736 present requiring a more detailed pathologic evaluation post the surgical procedure. Samples
737 were labelled with a de-identified barcode and transferred to the research team for further
738 processing.
739
740 Next, tonsils were dissociated into a single-cell suspension as previously described[141,142].
741 Briefly, Human tonsil tissue was processed by mincing with scissors, followed by transfer of
742 up to 4g of tissue to a gentleMACS C tube (Miltenyi Biotec) containing 8 mL of phosphate-
743 buffered saline (PBS) with 0.5 mg/mL collagenase D and 3000 U/mL DNase I, then dissociated
744 on a GentleMACS Octo Dissociator (Miltenyi Biotec) using “program C (Spleen program 2
745 followed by spleen program 3).” Tissue homogenates were incubated in a 37°C water bath for
746 15 minutes, then dissociated again using “program C” and transferred through a 100 µm cell
747 strainer into 20mL RPMI containing 10% human AB serum (Sigma Aldrich). Next, the cell
748 suspension was overlaid on 10mL of Ficoll-Paque PLUS (GE Healthcare) and subjected to
749 density-gradient separation via centrifugation for 20 min at 1800 rpm, 20°C, slow acceleration
750 and no brake. Leukocytes were collected from the interphase layer and then washed with
751 50mL of PBS for 6 minutes at 1600 rpm, 20°C.
752
753 Single cell suspensions of tonsil mononuclear cells were subjected to positive selection with
754 anti-human-CD3, anti-human-CD19 and anti-human-CD14 (Miltenyi Biotec) and transferred
755 through LD columns (Miltenyi Biotec) according to the manufacturer's guidelines (**Fig. S9**).
756 The depleted cell suspension flowthrough was collected into a 15mL conical tube and then
757 centrifuged for 5 minutes at 1200rpm, 20°C. Subsequently, cells were labelled with

758 LIVE/DEAD™ Fixable Near-IR dead cell stain kit (Invitrogen). Next, cells were labeled with
759 sorting antibody cocktail which contained negative linage (Lin-) CD19 Brilliant Violet (BV)421
760 (HIB19), CD14-BV421 (63D3) and CD3-BV421 (OKT3), and the following antibodies: CD45-
761 FITC, (HI30), CD94-PerCP-Cy5.5 (DX22), CD127-PE-Cy7 (A019D5), cKit-BV510 (104D2)
762 and NKp44-Alexa Fluor (AF)647 (P44-8) all purchased from Biolegend (San Diego, CA),
763 CRTH2-PE (301109, R&D). ILC3 cells were sorted based on the expression of CD45+Lin-
764 CD127+CD94-CRTH2-cKit+NKp44+, similarly to Bar-Ephraim et al. Cell sorting was
765 performed using a FACSAria II sorter (BD Biosciences, Mountain View, CA, USA). Post sorting
766 sorted ILC3 cells were washed with PBS for 5 minutes at 1200rpm, 20°C and then incubated
767 in 100 uL of 2% formaldehyde (in PBS) for 10 minutes, followed by the addition of 0.125M
768 glycine. Next, cells were centrifuged at 400g for 5 minutes at 4°C, resuspended with cold PBS
769 and centrifuged again at 400g for 5 minutes at 4°C, supernatant was discarded, and cells were
770 snap-frozen in liquid nitrogen and then stored at -80°C prior to PCHi-C analysis.

771 Human CD4+ T cell isolation

772 Total CD4+ lymphocytes were obtained from PBMCs from venous blood by negative selection
773 using EasySep Human CD4+ T Cell Enrichment kit (Catalog #19052) from STEMCELL
774 Technologies. Purified CD4+ T cells were washed with PBS for 5 minutes at 1200 rpm, 20°C
775 and then incubated in 100 µL of 2% formaldehyde (in PBS) for 10 minutes, followed by the
776 addition of 0.125M glycine. Next, cells were centrifuged at 400g for 5 minutes at 4°C,
777 resuspended with cold PBS and centrifuged again at 400g for 5 minutes at 4°C, supernatant
778 was discarded, and cells were snap-frozen in liquid nitrogen and then stored at -80°C prior to
779 PCHi-C analysis. Two replicates of 1 million and two more replicates of 50,000 cells were used
780 to generate PCHi-C datasets. The samples were obtained from two male donors after written
781 informed consent under studies “A Blueprint of Blood Cells,” REC reference 12/EE/0040, and
782 “Genes and mechanisms in type 1 diabetes in the Cambridge BioResource,” REC reference
783 05/Q0106/20; both approved by the NRES Committee East of England – Cambridgeshire and
784 Hertfordshire.

785 Promoter Capture Hi-C

786 Promoter Capture Hi-C was performed as previously described[44]. Cells were lysed in a lysis
787 buffer (30 minutes on ice), and digested with *DpnII* (NEB) overnight at 37°C while rotating (950
788 rpm). Restriction overhangs were filled in with Klenow (NEB) using biotin-14-dATP (Jena
789 Bioscience), and ligation was performed in the ligation buffer for 4 hours at 16°C (T4 DNA
790 ligase; Life Technologies). After overnight de-crosslinking at 65°C, the ligated DNA was
791 tagmented to produce fragments of 300-700 bp. Ligation products were isolated using MyOne
792 C1 streptavidin beads (Life Technologies), followed by washing with Wash&Binding buffer and
793 nuclease-free water. Isolated Hi-C ligation products on the beads were then used directly for
794 PCR amplification, and the final Hi-C library was purified with AMPure XP beads (Beckman
795 Coulter). Promoter Capture Hi-C was performed using a custom-designed Agilent SureSelect
796 system following the manufacturer’s protocol. The PCHi-C libraries were paired-end
797 sequenced (100 bp) on an Illumina HiSeq 2500 machine at a sequencing depth of ~400 million
798 reads per sample (**Table S1**).

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[143] and quality-controlled using HiCUP[144]. Quality metrics for all generated PCHi-C datasets are reported in **Table S1**. Significant interactions were then detected across the replicates by CHiCAGO[39] as previously described[40] at single *DpnII* fragment resolution and in bins of fragments approximately 5 kb in length, with the baited promoter fragments left solitary (unbinned).

Leaving the baited *DpnII* fragment unbinned meant that nearly every baited fragment was occupied by a single protein-coding gene promoter. In contrast, a third (33%) of baited fragments in the *HindIII*-based Capture Hi-C design (with a median fragment size of 4 kb) contained two or more promoters. Therefore, leaving the baited fragment unbinned significantly improved the resolution and interpretability of analyses such as (multi)COGS.

For CHiCAGO analysis at single-fragment resolution, p-value weights were estimated following our previously described procedure[40] and are listed in **Table S11**; default p-value weights were used for the 5 kb 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 5 kb 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[30] 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[40,145,146], direct comparison across these two datasets is challenging. To partially address this issue, we pooled the reads in the *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[39] with the default *HindIII*-based parameters and integrated them with the significant interactions from the Javierre et al. study[30]. 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[147], obtaining the plot shown in **Fig. S1A.**

Alternative promoter analysis

We used the CHiCAGO results for ILC3 PCHi-C data at 5 kb 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

844 with all baited alternative promoters for the same gene, and partially shared PIRs as those
845 that interacted with a subset of alternative promoters for the same gene. We defined distinct
846 PIRs as those that only interacted with a single promoter fragment (CHiCAGO score ≥5). To
847 increase the stringency with which we called PIRs "distinct", we applied two further criteria.
848 First, if a PIR interacted with another alternative promoter at a lenient CHiCAGO score ≥3, we
849 defined that PIR as shared. Second, if the adjacent fragment to the PIR in question interacted
850 with another alternative promoter at a CHiCAGO score ≥3, we also defined that PIR as shared.
851 We note that, under our classification rules, the PIRs of genes with only two alternative
852 promoters included in the analysis can only be classified as "fully shared" or "distinct".
853 Therefore, the "partially shared" PIR category was only applicable to the subset of genes with
854 more than two baited alternative promoters.

855 Epigenomic data pre-processing

856 For epigenetic data analysis in ILC3s, the SRA accession list was downloaded from the GEO
857 accession GSE77299. The SRA files were converted to FASTQ file,s and sequencing
858 adapters were trimmed from reads using *trim galore*
859 (https://github.com/FelixKrueger/TrimGalore). The reads were filtered by PHRED score ≥30
860 and examined for proper pairing with a mate (when paired-end). The sequencing quality and
861 duplication level were checked using FastQC
862 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequences were mapped to
863 the hg38 reference genome using STAR with modifications for aligning ChIP-seq and ATAC-
864 seq reads. Samtools[148] was used to select reads with a MAPQ score of 255, which is the flag
865 for uniquely mapping reads from STAR[149]. ATAC-seq reads were filtered, retaining properly
866 paired and oriented reads using the samflag=3. PCR duplicates were removed using
867 samtools. We then removed reads that fell within blacklisted regions using Bedtools[150]
868 intersect. The final filtered BAM file was then converted to a BED file using Bedtools bamtobed.
869 This conversion breaks read-pairing and ensures each read contributes to peak identification
870 with MACS2[151]. The ATAC-seq reads in BED format were shifted by +4 bp on the (+) strand,
871 and -5 bp on the (-) strand to account for the Tn5 transposase cut site. Peaks were called
872 using MACS2 using three biological replicates per sample as the treatment group with an input
873 ChIP-seq control sample. The replicate correlation between the ATAC-seq samples was poor,
874 with a <10% overlap between biological replicates. This result was consistent with the high
875 level of duplication and low peak count (8,852) in the worst sample (SRR3129112). Thus, our
876 ATAC-seq results were limited to the sample withthe best quality metrics (SRR3129113). In
877 total, we detected 34,077 H3K27ac peaks and 72,825 ATAC-seq peaks. For epigenetic data
878 analysis in CD4+ T cells, we used BLUEPRINT epigenome datasets from male donors
879 C002Q1, S008H1, and S007G7.

880 Activity-By-Captured-Contact (ABCC)

881 For a given gene-enhancer pair, the ABC score is the normalised product of enhancer Activity
882 (proxied by the levels of chromatin accessibility and relevant histone modifications) and
883 Contact (proxied by 3D contact frequency detected from a chromosome conformation capture
884 assay)[14,41]. In the original implementation of ABC, Activity is estimated as the geometric mean
885 of read counts of DHS/ATAC-seq peaks and Contact by KR-normalised Hi-C contact
886 frequency between the respective element and gene promoter[41]. The resulting product is

887 divided by the sum of all ABC values for a given gene from enhancers within a 5-megabase
888 window around the transcription start site:

889 $$ABC\ Score_{E,G} = \frac{A_E \times C_{E,G}}{\Sigma e\ \boldsymbol{within\ 5\ Mb\ of\ G}\ A_e \times C_{e,G}}$$

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

897 $$N_{exp} = B_{mean}/(s_i * s_j),$$

898 where B_{mean} is the CHiCAGO-estimated Brownian noise level and s_i and $s_{,j}$ and the bait- and
899 other end-specific scaling factors. For promoters that could not be baited in the Capture Hi-C
900 design and those that were filtered out due to QC fail, we estimated the expected normalised
901 read count directly from the interaction distance *d*, using the distance function f(*d*) fitted by
902 CHiCAGO. Due to the strong bias of the distance function *d* towards the very short range
903 interactions (<1.5kbp) and to ensure we do not disregard long-distance interactions, in the
904 imputation procedure we introduced a contact frequency cap for candidate enhancers that are
905 closer than at least one fragment away from the bait equal to the contact frequency prediction
906 of distance function *d* at 1.5 kbp (median fragment length). Please refer to Additional File 1 in
907 the publication presenting the CHiCAGO pipeline[39] for the formal definition of these
908 parameters and their estimation procedures.
909
910 The imputed normalised read counts were used as Contact data in the ABC pipeline, and the
911 public H3K27ac and ATAC-seq data in ILC3s processed as described above were used to
912 compute Activity. To validate the ABCC approach, we took advantage of the high-throughput
913 CRISPRi-FlowFISH data from Fulco et al.[41], which presented the impact of perturbing ~3,500
914 enhancer elements on the expression of 30 genes in K562 cells. Since PCHi-C data for K562
915 cells are not currently available, we used our previously published PCHi-C dataset in the
916 related primary cell type, erythroblasts[30], to generate the ABC scores based on these data
917 and the ATAC-Seq and H3K27ac ChIP-Seq datasets for K562 cells from Fulco et al. In
918 comparison with the original ABC scores from Fulco et al. based on pooling conventional Hi-
919 C data from multiple cell types, our approach showed a higher precision (69.1% vs 58.3%) at
920 the same level of recall (58.3%) of CRISPRi-FlowFISH-validated enhancer-promoter pairs
921 (**Fig. S2**). To select ABCC score cutoff, we optimised the Pearson correlation between per-
922 gene ABCC numerator and gene expression ($R_{ABC\text{-}GE}$), in an approach inspired by Xu et al.[152].
923 We opted to use a single ABCC score cutoff of 0.023 in all analysed cell types, as it was close
924 to the maximum $R_{ABC\text{-}GE}$ in each cell type, as well as to the cutoff of 0.02 that yielded an optimal
925 precision-recall of CRISPRi-FlowFISH-validated enhancer-promoter pairs in K562 cells.

926 Microarray gene expression data analysis

927 The microarray CEL files were downloaded from the GEO accession number GSE78896. The
928 CEL files were then analysed using AltAnalyze (http://www.altanalyze.org/). Probes were
929 filtered for a DABG (detection above background) as previously described[153]. Probes were
930 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[154] 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[155]. 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 differentially 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 5 kb 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 well as the 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.

LOLA enrichment analysis

We performed LOLA v1.18[156] enrichment analysis to assess whether active and/or open regulatory elements of multiCOGS-prioritised genes were enriched for specific transcription factor binding sites and chromatin features compared to all genes tested by multiCOGS.

We defined active/open PIRs as those with overlapping ATAC-seq or H3K27ac ChIP-seq peaks within significant PIRs identified by promoter capture Hi-C interactions (CHiCAGO) or predicted by our ABCC algorithm for multiCOGS-prioritised genes. The background universe comprised all active/open PIRs from the same datasets for all tested genes. Regions were converted to GRanges objects using the GenomicRanges package, and enrichment was tested using the LOLA core pipeline with the LOLA Core RegionDB, using default parameters. Significant enrichments were defined as those with q-value < 0.05.

RELI analysis

RELI[56] (v0.1.1a) 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 R^2 value ≥ 0.8. LD blocks are then intersected with the genetic feature BED files. A null distribution is generated using randomly shuffled LD blocks (n=1,000) and 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.

Promoter-interacting regulatory elements were determined as input for RELI as follows. The PIR sets were the union of PCHi-C interactions (CHiCAGO score ≥ 5, binned to 5 kb or DpnII fragment-level resolution) and ABC enhancers, excluding any trans-chromosomal interactions. Regulatory elements were then defined as the union of peaks of open chromatin and H3K27ac in ILC3 and CD4+ T cells (using ATAC-seq and ChIP-seq data as above). The true intersection between these regulatory elements and PIRs in each cell type was then determined using pybedtools *intersect*. The coordinates for these regions were lifted over from hg38 to hg19 using UCSC liftOver (v. 377), then sorted and merged for use with RELI. RELI was run against all 495 traits with ≥ 10 independent risk loci and of European ancestry in the GWAS Catalog. 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 < 0.05 were defined as significant. For depicting RELI results, we labelled only significant traits with enrichment ≥2.

Standard COGS

To run standard COGS[30,31], 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 to 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[157] 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 irrespective of whether 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). PIRs with CHiCAGO interaction scores ≥5 or ABC scores of ≥0.04 were used as 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[37] (L2G gene prioritisation score > 0.5 for five CD studies[71,158–161]), the IBDDB database of functionally validated targets[80], a functional screen of IBD genes[81],

experimentally validated IBD and CD genes from DisGeNET[82] that had evidence "AlteredExpression", "Biomarker", "Posttranslationalmodification", or "Therapeutic" or CD-containing exonic variants in a recent IBD exome study[83].

Multivariate GWAS fine-mapping

The Sum of Single Effects (SuSIE) model allows for multiple causal variants within a GWAS locus[68,69]. We downloaded summary data for Crohn's disease[71] (GCST004132), Ulcerative colitis[71] (GCST004133), Inflammatory Bowel Disease[71] (GCST004131), Celiac Disease[162] (GCST000612), Adult onset Asthma[163] (GCST007799) and Primary Sclerosing Cholangitis[164] (GCST004030) from the GWAS Catalog. and used LD block data for EUR from ldetect (https://bitbucket.org/nygcresearch/ldetect-data/src/master/), which we liftOvered[165] 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[66]. For blocks with appreciable association signals (minimum $p < 10^{-6}$), we used the susieR package[68,69] 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[30,31], and used the posterior probabilities of association as input for multiCOGS.

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 additionally ran multiCOGS on each category separately by specifying the *feature.names* argument in the *compute_cogs* function.

Assessing the biological function of CD-prioritised genes

The Gene2Func tool in FUMA (v1.5.2) was run using all multiCOGS genes with a score ≥0.5, Ensembl version 102, and GTex v8. As a background, we used all genes with assigned multiCOGS scores in ILC3s, of which 17,984 had a recognised Ensembl Gene ID in FUMA. Multiple testing correction was done via the Benjamini-Hochberg method (FDR) with an adjusted p-value cutoff of 0.05 and a minimum of 2 genes in a set. The MsigDB version was

1060 v7.0. We additionally checked for enrichment of multiCOGS genes in The Inflammatory
1061 Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA)
1062 Framework[91]. We filtered the 496 datasets of differentially expressed (DE) genes (adjusted p-
1063 value < 0.05 and absolute log2 fold change ≥2) that were compared across the same tissues
1064 and selected only sets with a maximum of 2,000 DE genes, to avoid mis-estimation of the
1065 normalised enrichment score, resulting in 24 datasets. Then we ran the *enricher* function in
1066 the R package clusterProfiler[166] (version 4.2.2) for all multiCOGS genes with a score ≥0.5.
1067

1068 Cell culture

1069 Mouse MNK-3 cells[112] and the derived lines were cultured in DMEM with glucose/pyruvate/ L-
1070 glutamine supplemented with 10% fetal bovine serum, 1X penicillin-streptomycin, 10 ng/ml
1071 mouse recombinant IL-2 and IL-7 (R&D Systems), and 50 µM 2-mercaptoethanol. Media for
1072 CRISPRi MNK-3 (MNK-3i) cells contained 10 µg/ml blasticidin S, and media for CRISPRa
1073 MNK-3 (MNK-3a) cells contained 10 µg/ml blasticidin S and 1250 µg/ml hygromycin B. MNK-
1074 3i/a cells with sgRNA additionally received 2 µg/ml puromycin. MNK-3 activation was induced
1075 with 10 ng/ml IL-1β and 10 ng/ml IL-23 (R&D Systems).
1076

1077 CRISPR activation and interference

1078 MNK-3i cells were generated as described[167] from parental MNK-3 cells. In brief, MNK-3 cells
1079 were transduced with lentivirus containing pLenti CMV rtTA3 Blast (Addgene #26429),
1080 selected by blasticidin S, and then infected with TRE3G-dCas9-KRAB-P2A-mCherry
1081 lentivirus. Following incubation with doxycycline, mCherry-positive cells were subcloned, and
1082 Western blot analysis confirmed robust expression of doxycycline-inducible dCas9-KRAB.
1083 MNK-3a cells were lentivirally engineered from MNK-3 to constitutively express the dCas9-
1084 VP64 fusion gene (Addgene #61425) and the MS2-p65-HSF1 transactivator complex
1085 (Addgene #89308), selected by blasticidin S and hygromycin B, and subcloned. All cells were
1086 tested for mycoplasma.

1087 Sequences for Cln3-targeting and scrambled gRNAs were based on published sgRNA
1088 libraries for MNK-3i[168] and MNK-3a[169] are listed in **Table S14** alongside RT-qPCR primer
1089 sequences. sgRNA sequences and their reverse complement were synthesised by Sigma,
1090 annealed, and cloned into lenti sgRNA(MS2)_puro optimised backbone (Addgene #73797) for
1091 MNK-3a or sgOpti (Addgene #85681) for MNK-3i using Esp3I digestion as previously
1092 described[170]. sgRNA plasmid integration was confirmed by Sanger Sequencing (Ohio State
1093 Comprehensive Cancer Center Genomics Core, Columbus, OH, USA). Lentiviral plasmids
1094 pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260) were transfected along with the
1095 sgRNA plasmid into HEK293T cells (Mirus TransIT-293T transfection reagent). Lentivirus
1096 media was harvested and filtered 48-72 hr post-transfection. Puromycin selection began 36
1097 hr after lentiviral guide transduction into MNK-3i/a cells in the presence of polybrene. Bulk
1098 transduced populations were used for experiments and maintained in selection antibiotics.
1099 RT-qPCR confirmed repression (MNK-3i lines after 48 hr doxycycline incubation) or
1100 overexpression (MNK-3a) of target genes relative to *Actb* and respective scramble control
1101 (Trizol RNA isolation; Verso cDNA synthesis).

1102 To induce CRISPRi guide expression, MNK-3i stably expressing Cln3-targeting and
1103 scrambled (Scr) gRNAs were incubated with 2 µg/ml doxycycline for 48 hr. To confirm
1104 stimulation, cells were harvested 21 hr after cytokine stimulation and stained for intracellular
1105 IL-17F and IL-22 (eBioscience IL-22 clone 1H8PWSR and IL-17F clone eBio18F10; BD Life
1106 Sciences Cytofix/Cytoperm kit). Expression of IL-17F and IL-22 was assessed on
1107 FACSymphony (BD Life Sciences) and compared against a respective scrambled control.

1108 RNA-sequencing

1109 RNA was harvested by spin column (Qiagen RNeasy kit) for polyA-selected 2x150bp bulk
1110 RNAseq (Illumina platform, University of Cincinnati Genomics, Epigenomics, and Sequencing
1111 Core, Cincinnati, OH, USA). RNA-seq samples were generated in triplicate.
1112
1113 Raw paired-end RNA-seq reads were quantified using *kallisto* (v0.48.0) against the mouse
1114 reference transcriptome (GENCODE release M32, GRCm39). Transcript indices were first
1115 generated with *kallisto index*, and transcript abundances were quantified for each sample
1116 using *kallisto quant* with 100 bootstrap replicates. Transcript-level abundance estimates were
1117 subsequently summarised to the gene level in R using the *tximport* package (v1.30.0) together
1118 with a transcript-to-gene mapping file. Sample metadata, including experimental condition,
1119 CRISPR status, and replicate information, were compiled into a metadata table. Gene-level
1120 count matrices generated by *tximport* were then used as input for normalisation and differential
1121 expression analysis with *DESeq2* (v1.38.0). Sample metadata, including experimental
1122 condition, CRISPR status, stimulation, and replicate information, were compiled into a
1123 metadata table.

1124 Gene-level count matrices were then used for normalisation and differential expression
1125 analysis with DESeq2 (v1.38.0). A variance-stabilising transformation (rlog) was applied for
1126 visualisation and principal component analysis to identify batch effects. Differential expression
1127 analyses were performed using linear models incorporating relevant covariates. For wild-type
1128 samples, stimulation status was tested while including CRISPR type as a batch covariate. For
1129 CRISPRa and CRISPRi samples, models including interaction terms between CRISPR
1130 treatment and stimulation were used to assess treatment-specific effects. Adjusted p-values
1131 were calculated using the Benjamini-Hochberg method, and genes with adjusted p-values <
1132 0.05 were considered statistically significant.

1133 RNA isolation and quantitative RT–PCR

1134 Total RNA was isolated from snap-frozen cells using QIAshredder columns and the RNeasy
1135 spin-column system (QIAGEN). Complementary DNA (cDNA) was synthesised using the
1136 High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific).

1137 Quantitative PCR was performed using TaqMan chemistry with TaqMan Fast Advanced
1138 Master Mix (Thermo Fisher Scientific) on a QuantStudio 5 Real-Time PCR System (Thermo
1139 Fisher Scientific). Cln3 expression was quantified using the TaqMan Gene Expression Assay
1140 Mm00487021_m1 and normalised to the housekeeping gene Hprt using assay
1141 Mm03024075_m1. Reactions were performed in technical triplicate. Relative gene expression
1142 was calculated using the ΔΔCt method, with MNK-3 cells electroporated with GFP mRNA used
1143 as the reference control condition.

Design and generation of in vitro-transcribed mRNA

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

In vitro–transcribed mRNA was generated with a Cap 1 structure and incorporated N1-methylpseudouridine. Transcripts contained a poly(A) tail and were supplied in RNase-free sodium citrate buffer (pH 6.4) at a concentration of 1 mg ml^{-1}. Control mRNA encoding GFP was generated using the same chemistry.

mRNA electroporation and cytokine stimulation

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

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

ELISA assay

Cytokines in cell culture supernatants were quantified by ELISA using DuoSet kits for mouse IL-17, IL-22, and GM-CSF (R&D Systems) according to the manufacturer’s instructions. When necessary, samples were diluted to fall within the dynamic range of the standard curve. Absorbance was measured at 450 nm with wavelength correction at 560 nm using a GloMax Discover microplate reader (Promega). Cytokine concentrations were determined by interpolation from standard curves using a four-parameter logistic fit.

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

Immunoprecipitation and immunoblotting

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

Myc-tagged proteins were enriched by incubation of clarified lysates with Myc-Trap agarose beads (ChromoTek) for 1 hr at 4 °C with rotation. Beads were washed in buffer containing

0.05% DDM, and bound proteins were recovered for analysis. Input, unbound, and bound fractions were quantified by BCA assay (Thermo Fisher Scientific), denatured in LDS sample buffer with reducing agent, and resolved by SDS–PAGE on 4–12% Bis-Tris gels (Thermo Fisher Scientific). Proteins were transferred to PVDF membranes, stained with Revert 700 Total Protein Stain (LI-COR), and imaged prior to immunoblotting.

Membranes were blocked and probed with antibodies against myc tag (Cell Signaling Technology #2278, 1:1000) or GFP (Invitrogen #A-11122, 1:2000). Fluorescent secondary antibodies were used at 1:10,000 and blots were imaged using the Odyssey DLx Imaging System (LI-COR).

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

The analysis is based on data from Table S5 in Brown et al[125], containing a gene-level analysis of a CRISPRi screen in MNK-3i cells. In the experiments performed by Brown et al., MNK-3i cells were induced with doxycycline to express CRISPRi (dCas9-KRAB) machinery and were transduced with a lentiviral gRNA library targeting 20,003 genes. The cells were then stimulated by IL-23 and IL-1β and sorted into subpopulations expressing high and negative levels of the inflammatory cytokine IL-22 released by activated ILC3s. The quantity of each sgRNA in $IL22^{Neg}$ and $IL22^{High}$ cells was detected through PCR amplification and next-generation sequencing. To focus on sgRNA targeting expressed genes, the genes were filtered to those with an average transcript per million (TPM) of ≥2.5 in RNAseq data from MNK-3i+scramble (sgSCR) cells treated with dox (48 hr) and stimulated with 10 ng/ml IL-1β/23 (21 hr). The “test” command from MAGeCK (version 0.5.9.5) [171] was applied to generate normalised (method = total) gene-level rankings using Robust Rank Aggregation (RRA). The sgRNA enriched in the $IL22^{Neg}$ population pointed towards genes positively regulating IL-22 production, implicating them in ILC3 inflammatory response. In contrast, sgRNA showing enrichment in the $IL\text{-}22^{High}$ population points to ILC3 ‘anti-inflammatory’ genes.

In the present study, we first filtered the genes in **Table S5** from in Brown et al[125] to those that had been profiled in the multiCOGS experiment, based on an identical gene name between the mouse and human data, leading to a total set of 6438 genes. The genes were ranked based on their MAGeCK score for positive or negative regulation of IL-22 production. We then ran GSEA against each of these rankings, for the 142 multiCOGS genes for inflammatory traits, using the “pathway” function in MAGeCK. We considered significant CRISPRi genes to be those with an adjusted p-value < 0.05 in the gene-level RRA analysis.

Data availability

Raw PCHi-C data generated in this study for ILC3s are deposited in the Gene Expression Omnibus (GEO) under the accession number GSE216267. Processed R data files containing CHiCAGO scores at the fragment-level and 5kb-binned resolution can be found in the same repository. PCHi-C data for CD4+ T cells were deposited to the European Genome-Phenome Archive (EGA) under managed access in accordance with the conditions of donor consent, under the accession number EGAS50000001316. Raw RNA-seq reads and counts for the CLN3 CRISPRi/a experiments in MNK-3 cells are deposited in GEO under the accession

number GSE313942. Supplementary Data files, including significant CHiCAGO interactions at fragment-level and 5kb resolution in ILC3 and CD4+ T cells, ABCC pairs in both cell types and DESeq2 objects for the CLN3 CRISPRi/a experiments, were deposited to Open Science Framework (https://osf.io/aq9fb).

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 Enhanced Genomics Ltd. J.M.W. is an employee of Amicus Therapeutics, Inc. and holds equity in the company in the form of stock-based compensation; Amicus had no input into this piece of work. C.W. is also an employee of GSK; GSK had no input into this work.

Contributions

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

Acknowledgements

The authors would like to thank Laurence Game and Ivan Andrew at the LMS Genomics facility for sequencing, Michiel Thiecke and Oliver S. Burren for technical advice and Wing Leung for technical assistance. M.S. is core-funded by the Medical Research Council as an MRC Programme Leader (MC-A652-5QA20). C.W. is funded by the Wellcome Trust (WT220788), the MRC (MC_UU_00002/4), GSK and MSD, and supported by the NIHR Cambridge BRC (BRC-1215-20014). Research activities related to this manuscript that were conducted at Cincinnati Children's Hospital Medical Center were supported by the NIH (U01AI150748 to

E.R.M, M.T.W, L.C.K; R01AI153442 to E.R.M; R21AI156185 to E.R.M; R01HG010730, U01AI130830, R01NS099068, R01GM055479, P01AI150585, R01AI141569 to M.T.W; R01AI024717 to M.T.W and L.C.K; R01AR073228 to M.T.W, L.C.K and S.W; R01DK107502, R01AI148276, U19AI070235, U01HG011172, and P30AR070549 to L.C.K); Cincinnati Children's Research Foundation (ARC Award to E.R.M, M.T.W and L.C.K; Center for Pediatric Genomics grants to E.R.M, M.T.W and L.C.K); and the L.B. Research and Education Foundation (to N.L.). M.T.W. S.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**.** Research activities related to this manuscript that were conducted at Ohio State University's Medical Center were supported by the NIH (R01AI134035 to E.M.O), Pelotonia Scholars Graduate (to R.A.B.) and PostDoctoral (to A.S.) Fellowships. This project was funded in part by a Commercialisation Award from the Babraham Institute to M.S. and P.F and by the VIB core funding to V.M. M.F. was supported by the British Heart Foundation (BHF; FS/18/53/33863), the BHF Cambridge Centre for Research Excellence (RE/18/1/34212) and the NIHR Exeter Biomedical Research Centre. We thank the UK's National Institute for Health and Care Research (NIHR) BioResource volunteers for their participation and gratefully acknowledge NIHR BioResource centres. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. For the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

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. Liu, Z. *et al.* Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nature Genetics* **55**, 796–806 (2023).
14. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
15. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
16. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
17. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **11**, 424 (2020).
18. 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.
19. Robinette, M. L. *et al.* Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat. Immunol.* **16**, 306–317 (2015).
20. 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).
21. Koues, O. I. *et al.* Distinct Gene Regulatory Pathways for Human Innate versus Adaptive Lymphoid Cells. *Cell* **165**, 1134–1146 (2016).
22. Shih, H.-Y. *et al.* Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell* **165**, 1120–1133 (2016).
23. 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).
24. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals

1341 folding principles of the human genome. *Science* **326**, 289–293 (2009).
1342 25. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals
1343 principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
1344 26. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-
1345 resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
1346 27. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their
1347 long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
1348 28. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the
1349 mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).
1350 29. Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W. & Fraser, P.
1351 Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter
1352 Interactions. *J. Vis. Exp.* (2018) doi:10.3791/57320.
1353 30. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-
1354 coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384.e19 (2016).
1355 31. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune
1356 disease candidate genes. *Genome Biol.* **18**, 165 (2017).
1357 32. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range
1358 interactions with related autoimmune risk loci. *Nat. Commun.* **6**, 10069 (2015).
1359 33. Baxter, J. S. *et al.* Capture Hi-C identifies putative target genes at 33 breast cancer risk
1360 loci. *Nat. Commun.* **9**, 1028 (2018).
1361 34. Ray-Jones, H. *et al.* Mapping DNA interaction landscapes in psoriasis susceptibility loci
1362 highlights KLF4 as a target gene in 9q31. *BMC Biol.* **18**, 47 (2020).
1363 35. Shi, C. *et al.* Chromatin Looping Links Target Genes with Genetic Risk Loci for
1364 Dermatological Traits. *J. Invest. Dermatol.* **141**, 1975–1984 (2021).
1365 36. Thiecke, M. J., Yang, E. J., Burren, O. S., Ray-Jones, H. & Spivakov, M. Prioritisation of
1366 Candidate Genes Underpinning COVID-19 Host Genetic Traits Based on High-
1367 Resolution 3D Chromosomal Topology. *Front. Genet.* **12**, 745672 (2021).
1368 37. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification
1369 and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
1370 38. Fang, H. & Knight, J. C. Priority index: database of genetic targets in immune-mediated
1371 disease. *Nucleic Acids Res.* **50**, D1358–D1367 (2022).
1372 39. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C
1373 data. *Genome Biol.* **17**, 127 (2016).
1374 40. Freire-Pritchett, P. *et al.* Detecting chromosomal interactions in Capture Hi-C data with
1375 CHiCAGO and companion tools. *Nat. Protoc.* **16**, 4144–4176 (2021).
1376 41. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from
1377 thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
1378 42. Cotman, S. L. & Staropoli, J. F. The juvenile Batten disease protein, CLN3, and its role
1379 in regulating anterograde and retrograde post-Golgi trafficking. *Clin. Lipidol.* **7**, 79–91
1380 (2012).
1381 43. Tang, C. *et al.* A human model of Batten disease shows role of CLN3 in phagocytosis at
1382 the photoreceptor–RPE interface. *Communications Biology* **4**, 1–18 (2021).
1383 44. Ho, J. S. Y. *et al.* TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal
1384 inflammation. *Cell* **184**, 2618–2632.e17 (2021).
1385 45. Malysheva, V. et al. Novel Method WO 2021/064430. *World Patent* (2021).
1386 46. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold
1387 Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
1388 47. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acid Research* **51**, (2022).
1389 48. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat.*
1390 *Biotechnol.* **30**, 224–226 (2012).
1391 49. Ray-Jones, H. & Spivakov, M. Transcriptional enhancers and their communication with
1392 gene promoters. *Cell. Mol. Life Sci.* **78**, 6453–6485 (2021).
1393 50. Cella, M. *et al.* Subsets of ILC3-ILC1-like cells generate a diversity spectrum of innate
1394 lymphoid cells in human mucosal tissues. *Nat. Immunol.* **20**, 980–991 (2019).

51. Jan-Abu, S. C., Kabil, A. & McNagny, K. M. Parallel origins and functions of T cells and ILCs. *Clin Exp Immunol* **213**, 76–86 (2023).
52. Robinette, M. L. & Colonna, M. Immune modules shared by innate lymphoid cells and T cells. *J. Allergy Clin. Immunol.* **138**, 1243–1251 (2016).
53. von Burg, N. *et al.* Activated group 3 innate lymphoid cells promote T-cell-mediated immune responses. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12835–12840 (2014).
54. Schroeder, J.-H., Howard, J. K. & Lord, G. M. Transcription factor-driven regulation of ILC1 and ILC3. *Trends Immunol.* **43**, 564–579 (2022).
55. Cairns, J., Orchard, W. R., Malysheva, V. & Spivakov, M. Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics* **35**, 4764–4766 (2019).
56. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
57. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).
58. Cerezo, M. *et al.* The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Res* **53**, D998–D1005 (2024).
59. Allez, M. *et al.* CD4+NKG2D+ T cells in Crohn's disease mediate inflammatory and cytotoxic responses through MICA interactions. *Gastroenterology* **132**, 2346–2358 (2007).
60. Pariente, B. *et al.* Activation of the receptor NKG2D leads to production of Th17 cytokines in CD4+ T cells of patients with Crohn's disease. *Gastroenterology* **141**, 217–26, 226.e1–2 (2011).
61. Rabe, H. *et al.* Distinct patterns of naive, activated and memory T and B cells in blood of patients with ulcerative colitis or Crohn's disease. *Clin Exp Immunol* **197**, 111–129 (2019).
62. Ma, C. *et al.* Critical Role of CD6highCD4+ T Cells in Driving Th1/Th17 Cell Immune Responses and Mucosal Inflammation in IBD. *J Crohns Colitis* **13**, 510–524 (2019).
63. Pearson, C. *et al.* ILC3 GM-CSF production and mobilisation orchestrate acute intestinal inflammation. (2016) doi:10.7554/eLife.10066.
64. Tomás-Daza, L. *et al.* Low input capture Hi-C (liCHi-C) identifies promoter-enhancer interactions at high-resolution. *Nature Communications* **14**, 1–16 (2023).
65. Thiecke, M. J. *et al.* Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep.* **32**, 107929 (2020).
66. Rüeger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* **14**, e1007371 (2018).
67. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254 (2022).
68. 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).
69. 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).
70. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).
71. 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).
72. Naser, S. A. *et al.* Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J. Gastroenterol.* **18**, 412–424 (2012).
73. 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).
74. Assadi, G. *et al.* LACC1 polymorphisms in inflammatory bowel disease and juvenile

1449 idiopathic arthritis. *Genes Immun.* **17**, 261–264 (2016).
1450 75. Sun, R., Hedl, M. & Abraham, C. TNFSF15 Promotes Antimicrobial Pathways in Human
1451 Macrophages and These Are Modulated by TNFSF15 Disease-Risk Variants. *Cell Mol*
1452 *Gastroenterol Hepatol* **11**, 249–272 (2021).
1453 76. Lahiri, A., Hedl, M., Yan, J. & Abraham, C. Human LACC1 increases innate receptor-
1454 induced responses and a LACC1 disease-risk variant modulates these outcomes. *Nat.*
1455 *Commun.* **8**, 15614 (2017).
1456 77. Schmitt, H., Neurath, M. F. & Atreya, R. Role of the IL23/IL17 Pathway in Crohn's
1457 Disease. *Front Immunol* **12**, 622934 (2021).
1458 78. Powell, M. D., Read, K. A., Sreekumar, B. K. & Oestreich, K. J. Ikaros zinc finger
1459 transcription factors: Regulators of cytokine signaling pathways and CD4+ T helper cell
1460 differentiation. *Front. Immunol.* **10**, 1299 (2019).
1461 79. Chang, K. *et al.* Multi-omics profiles refine L-dopa decarboxylase (DDC) as a reliable
1462 biomarker for prognosis and immune microenvironment of clear cell renal cell
1463 carcinoma. *Front. Oncol.* **12**, 1079446 (2022).
1464 80. Khan, F., Radovanovic, A., Gojobori, T. & Kaur, M. IBDDB: a manually curated and text-
1465 mining-enhanced database of genes involved in inflammatory bowel disease. *Database*
1466 **2021**, baab022 (2021).
1467 81. Ntunzwenimana, J. C. *et al.* Functional screen of inflammatory bowel disease genes
1468 reveals key epithelial functions. *Genome Med.* **13**, 181 (2021).
1469 82. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human
1470 disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2016).
1471 83. Sazonovs, A. *et al.* Large-scale sequencing identifies multiple genes and rare variants
1472 associated with Crohn's disease susceptibility. *Nat. Genet.* 1–9 (2022).
1473 84. Xiao, Y., Huang, Q., Wu, Z. & Chen, W. Roles of protein ubiquitination in inflammatory
1474 bowel disease. *Immunobiology* **225**, 152026 (2020).
1475 85. Macho-Fernandez, E. *et al.* Lymphotoxin beta receptor signaling limits mucosal damage
1476 through driving IL-23 production by epithelial cells. *Mucosal Immunol.* **8**, 403–413
1477 (2015).
1478 86. Schmidtke, C. *et al.* Lysosomal proteome analysis reveals that CLN3-defective cells
1479 have multiple enzyme deficiencies associated with changes in intracellular trafficking. *J*
1480 *Biol Chem* **294**, 9592–9604 (2019).
1481 87. Laqtom, N. N. *et al.* CLN3 is required for the clearance of glycerophosphodiesters from
1482 lysosomes. *Nature* **609**, 1005–1011 (2022).
1483 88. Mole, S. E. & Cotman, S. L. Genetics of the neuronal ceroid lipofuscinoses (Batten
1484 disease). *Biochim. Biophys. Acta* **1852**, 2237–2241 (2015).
1485 89. He, P. *et al.* A human fetal lung cell atlas uncovers proximal-distal gradients of
1486 differentiation and key regulators of epithelial fates. *Cell* **185**, 4841–4860.e25 (2022).
1487 90. Victor, A. R. *et al.* IL-18 drives ILC3 proliferation and promotes IL-22 production via NF-
1488 κB. *J. Immunol.* **199**, 2333–2342 (2017).
1489 91. Massimino, L. *et al.* The Inflammatory Bowel Disease Transcriptome and
1490 Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nature Computational*
1491 *Science* vol. 1 511–515 Preprint at https://doi.org/10.1038/s43588-021-00114-y (2021).
1492 92. Ueta, M. *et al.* Mucocutaneous inflammation in the Ikaros Family Zinc Finger 1-keratin
1493 5-specific transgenic mice. *Allergy* **73**, 395–404 (2018).
1494 93. Hildner, K., Punkenburg, E., Abendroth, B. & Neurath, M. F. Immunopathogenesis of
1495 IBD: Batf as a Key Driver of Disease Activity. *Dig. Dis.* **34 Suppl 1**, 40–47 (2016).
1496 94. Griffin, M. J. *et al.* Early B-cell factor-1 (EBF1) is a key regulator of metabolic and
1497 inflammatory signaling pathways in mature adipocytes. *J. Biol. Chem.* **288**, 35925–
1498 35939 (2013).
1499 95. Bianchi, E., Rogge, L. & Vecellio, M. *Role of the IL-23/IL-17 Pathway in Chronic*
1500 *Immune-Mediated Inflammatory Diseases: Mechanisms and Targeted Therapies*.
1501 (Frontiers Media SA, 2021).
1502 96. Atreya, I., Atreya, R. & Neurath, M. F. NF-kappaB in inflammatory bowel disease. *J.*

1503 *Intern. Med.* **263**, 591–596 (2008).
1504 97. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–
1505 32 (2016).
1506 98. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin
1507 landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88
1508 (2007).
1509 99. Chantalat, S. *et al.* Histone H3 trimethylation at lysine 36 is associated with constitutive
1510 and facultative heterochromatin. *Genome Res.* **21**, 1426–1437 (2011).
1511 100. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two
1512 years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2018).
1513 101. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic
1514 association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
1515 102. Lepik, K. *et al.* C-reactive protein upregulates the whole blood expression of CD59 - an
1516 integrative analysis. *PLoS Comput. Biol.* **13**, e1005766 (2017).
1517 103. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**,
1518 580–585 (2013).
1519 104. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune
1520 System of Human Populations. *Cell* **167**, 643–656.e17 (2016).
1521 105. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human
1522 Immune Cells. *Cell* **167**, 1398–1414.e24 (2016).
1523 106. Longman, R. S. *et al.* CX_3CR1^+ mononuclear phagocytes support colitis-associated
1524 innate lymphoid cell production of IL-22. *J Exp Med* **211**, 1571–1583 (2014).
1525 107. Zhao, X. *et al.* Elevated nitric oxide during colitis restrains GM-CSF production in ILC3
1526 cells via suppressing an AhR-Cyp4f13-NF-κB axis. *Nature Communications* **16**, 1–18
1527 (2025).
1528 108. Castellanos, J. G. *et al.* Microbiota-Induced TNF-like Ligand 1A Drives Group 3 Innate
1529 Lymphoid Cell-Mediated Barrier Protection and Intestinal T Cell Activation during Colitis.
1530 *Immunity* **49**, 1077–1089.e5 (2018).
1531 109. Lacy, P. & Stow, J. L. Cytokine release from innate immune cells: association with
1532 diverse membrane trafficking pathways. *Blood* **118**, 9–18 (2011).
1533 110. Calcagni', A. *et al.* Loss of the batten disease protein CLN3 leads to mis-trafficking of
1534 M6PR and defective autophagic-lysosomal reformation. *Nature Communications* **14**,
1535 3911 (2023).
1536 111. Cotman, S. L. & Lefrancois, S. CLN3, at the crossroads of endocytic trafficking.
1537 *Neurosci Lett* **762**, 136117 (2021).
1538 112. Allan, D. S. J. *et al.* An in vitro model of innate lymphoid cell function and differentiation.
1539 *Mucosal Immunol* **8**, 340–351 (2015).
1540 113. Zhu, J. GATA3 regulates the development and functions of innate lymphoid cell subsets
1541 at multiple stages. *Front. Immunol.* **8**, 1571 (2017).
1542 114. Zhang, Y., Ding, H., Wang, X. & Ye, S.-D. Modulation of STAT3 phosphorylation by
1543 PTPN2 inhibits naïve pluripotency of embryonic stem cells. *FEBS Lett.* **592**, 2227–2237
1544 (2018).
1545 115. Michieletto, M. F. *et al.* Multiscale 3D genome organization underlies ILC2 ontogenesis
1546 and allergic airway inflammation. *Nat. Immunol.* 1–13 (2022).
1547 116. Castel, S. E. *et al.* Modified penetrance of coding variants by cis-regulatory variation
1548 contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).
1549 117. Dryden, N. H. *et al.* Unbiased analysis of potential targets of breast cancer susceptibility
1550 loci by Capture Hi-C. *Genome Res* **24**, 1854–1868 (2014).
1551 118. He, H. *et al.* Multiple functional variants in long-range enhancer elements contribute to
1552 the risk of SNP rs965513 in thyroid cancer. *Proc Natl Acad Sci U S A* **112**, 6128–6133
1553 (2015).
1554 119. Lawrenson, K. *et al.* Cis-eQTL analysis and functional validation of candidate
1555 susceptibility genes for high-grade serous ovarian cancer. *Nat Commun* **6**, 8234 (2015).
1556 120. Costanzo, M. C. *et al.* Realizing the promise of genome-wide association studies for

effector gene prediction. *Nat. Genet.* **57**, 1578–1587 (2025).
121. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
122. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
123. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
124. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23**, 583–593 (2020).
125. Brown, R. A. *et al.* CRISPRi Screening Identifies SON and MAP4K1 as Regulators of Type III Cytokine Expression in Innate Lymphoid Cells. *bioRxiv* (2025) doi:10.1101/2025.08.15.670561.
126. Castaneda, J. A., Lim, M. J., Cooper, J. D. & Pearce, D. A. Immune system irregularities in lysosomal storage disorders. *Acta Neuropathologica* **115**, 159–174 (2007).
127. Hersrud, S. L., Kovács, A. D. & Pearce, D. A. Antigen presenting cell abnormalities in the Cln3(-/-) mouse model of juvenile neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* **1862**, 1324–1336 (2016).
128. Mirza, M. *et al.* The CLN3 gene and protein: What we know. *Mol Genet Genomic Med* **7**, e859 (2019).
129. Brudvig, J. J. *et al.* Glycerophosphoinositol is Elevated in Blood Samples From pigs, Mice, and CLN3-Affected Individuals. *Biomark. Insights* **17**, 11772719221107765 (2022).
130. 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).
131. Ziółkowska, E. A. *et al.* Gene therapy ameliorates bowel dysmotility and enteric neuron degeneration and extends survival in lysosomal storage disorder mouse models. *Sci Transl Med* **17**, eadj1445 (2025).
132. 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).
133. 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).
134. 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).
135. 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.
136. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
137. Borén, J., Packard, C. J. & Binder, C. J. Apolipoprotein B-containing lipoproteins in atherogenesis. *Nat Rev Cardiol* **22**, 399–413 (2025).
138. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
139. 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).
140. 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).
141. Reighard, S. D. *et al.* Therapeutic Targeting of Follicular T Cells with Chimeric Antigen Receptor-Expressing Natural Killer Cells. *Cell Rep Med* **1**, (2020).

142. Krabbendam, L., Nagasawa, M., Spits, H. & Bal, S. M. Isolation of Human Innate Lymphoid Cells. *Curr. Protoc. Immunol.* **122**, e55 (2018).
143. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
144. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
145. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **18**, 1046–1055 (2021).
146. 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).
147. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018) doi:10.48550/arXiv.1802.03426.
148. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
149. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
150. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
151. Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv* (2018) doi:10.1101/496521.
152. Xu, Z. *et al.* Structural variants drive context-dependent oncogene activation in cancer. *Nature* **612**, 564–572 (2022).
153. Lockstone, H. E. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief. Bioinform.* **12**, 634–644 (2011).
154. 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).
155. 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).
156. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).
157. 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.
158. 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).
159. 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).
160. 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).
161. 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).
162. Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
163. Ferreira, M. A. R. *et al.* Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am. J. Hum. Genet.* **104**, 665–684 (2019).
164. Ji, S.-G. *et al.* Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.* **49**, 269–273 (2017).
165. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–8 (2006).
166. 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).
167. Saini, A. *et al.* Cell type-specific enhancers regulate IL-22 expression in innate and adaptive lymphoid cells. *bioRxiv* (2025) doi:10.1101/2025.04.02.646834.
168. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, (2016).
169. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828–863 (2017).
170. Moyer, T. C. & Holland, A. J. Generation of a conditional analog-sensitive kinase in human cells using CRISPR/Cas9-mediated genome engineering. *Methods Cell Biol* **129**, 19–36 (2015).
171. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).
172. Honap, S. *et al.* JAK inhibitors for inflammatory bowel disease: recent advances. *Frontline Gastroenterol.* **15**, 59–69 (2024).
173. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208–227 (2007).
174. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics* **53**, 1300–1310 (2021).
175. Chicco, D. & Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData mining* **16**, (2023).

Figures

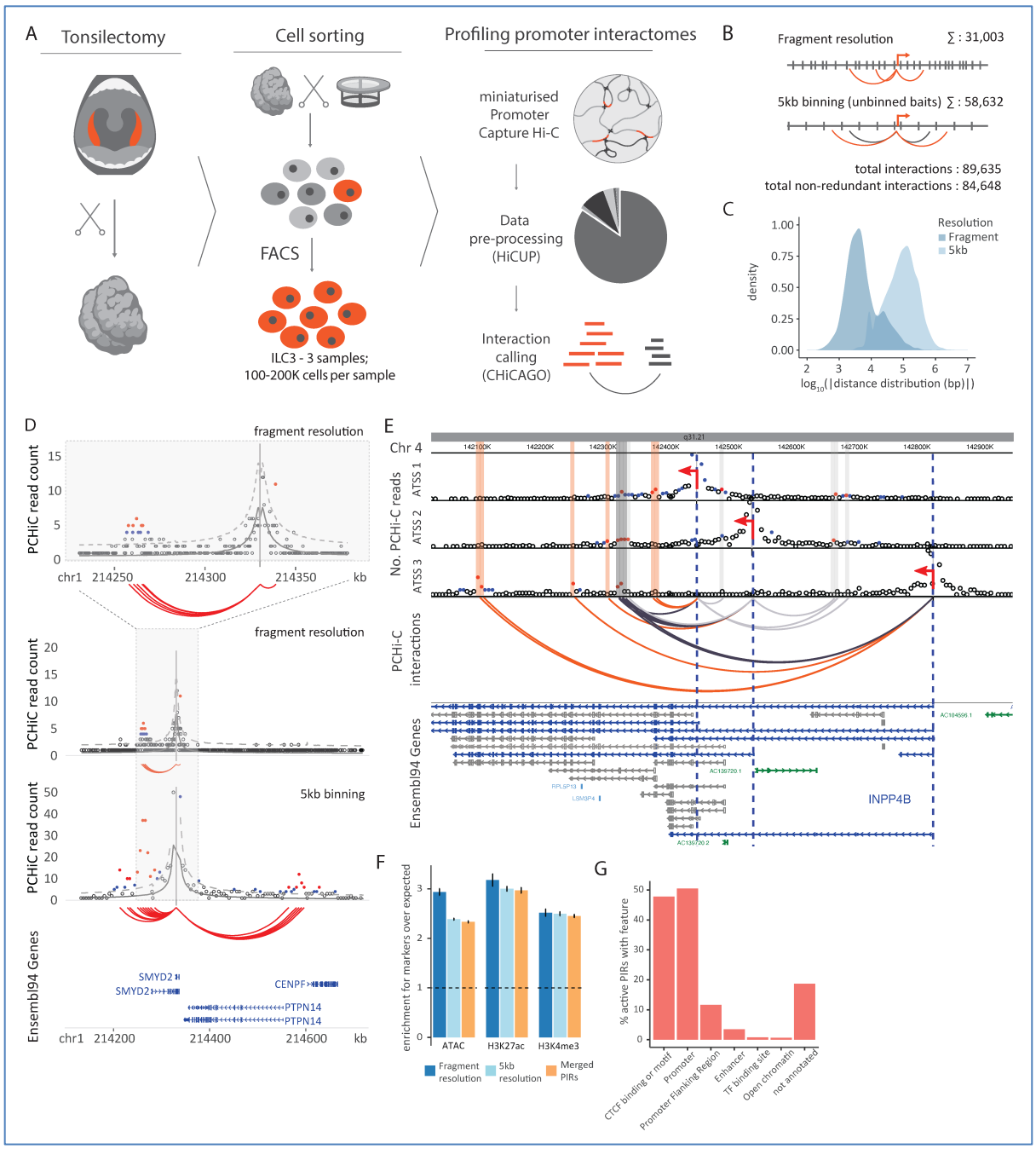

Figure 1. Compendium of promoter interactions in ILC3s. A. Outline of the study. **B.** Chromosomal interaction binning strategy. The analysis is done in two modes: fragment resolution (no binning) and 5kb binning. In the 5kb binning mode, the baited (captured) fragment containing a gene promoter, is left unbinned to enable high-resolution linkage between the promoter and distal enhancers. Interactions uniquely detected in one mode only are shown as red arcs, and those detected in both modes are shown as grey arcs. The numbers of significant interactions are given for each mode individually and merged across both modes (see Methods for details). **C.** Comparison of promoter-PIR distance distributions for PIRs detected at fragment and 5kb resolution. **D.** Example of chromosomal interactions for the *SMYD2* gene at fragment and 5kb resolution. The inset shows a zoomed-in view of the promoter interactions detected at fragment resolution. **E.** Example of multiple degrees of contact sharing between alternative promoters for the *INPP4B* gene. Captured alternative promoters are indicated by red arrows and blue dashed lines. The transcripts driven by these promoters (based on

Ensembl 94) are shown in blue, and other *INPP4B* transcripts are shown in grey. Transcripts for processed pseudogenes are shown in light blue and lincRNAs in green. PIRs are categorised as fully shared between alternative promoters (dark grey arcs), partially shared (light grey arcs) or distinct (red arcs). **F.** Enrichment of PIRs for the markers of active enhancers and promoters (H3K27ac and H3K4me3) and accessible chromatin (ATAC) in hILC3s. The error bars represent 95% confidence intervals, accounting for error propagation. **G.** Characterisation of active and/or open ILC3 PIRs at merged fragments as per Ensembl annotations and CTCF motifs.

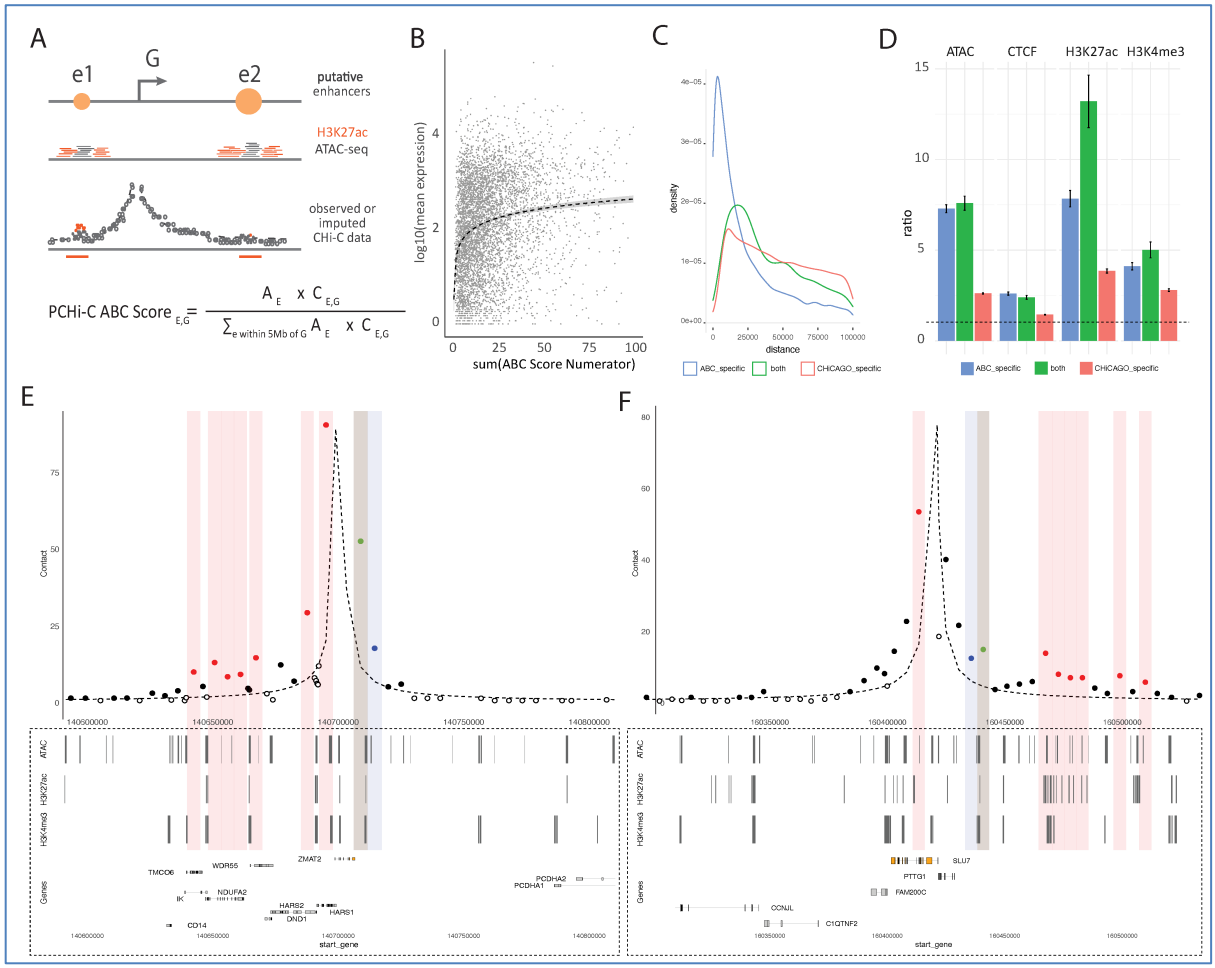

Figure 2. Combining ABCC and CHiCAGO to link distal elements with target genes. **A.** Schematic depicting the adaptation of the Activity-By-Contact (ABC) model for use with PCHi-C data, termed Activity-By-Captured-Contact (ABCC). **B.** Correlation between gene expression and ABC numerator score summed across all predicted enhancers per gene. The dashed line shows a mixed model fit via restricted maximum likelihood, with the shaded area around the line representing the confidence interval. **C.** Interaction distance comparison across CHiCAGO-specific, ABCC-specific and shared interactions. **D.** Enrichment for markers of active/open regulatory elements in CHiCAGO-specific, ABCC-specific, and shared regulatory elements. **E, F.** Representative examples of CHiCAGO- and ABCC-detected contacts (for *SLU7* and *ZMAT2* promoters). The dashed line shows expected counts estimated using the CHiCAGO distance function. PIRs detected with CHiCAGO at 5 kb resolution are shown as red dots and shading, with ABCC as blue dots and shading and by both approaches as green points and shading. Black filled dots represent imputed counts considered by ABCC, corresponding to the maximum value between observed and expected counts. Unfilled dots represent observed counts falling below expected values.

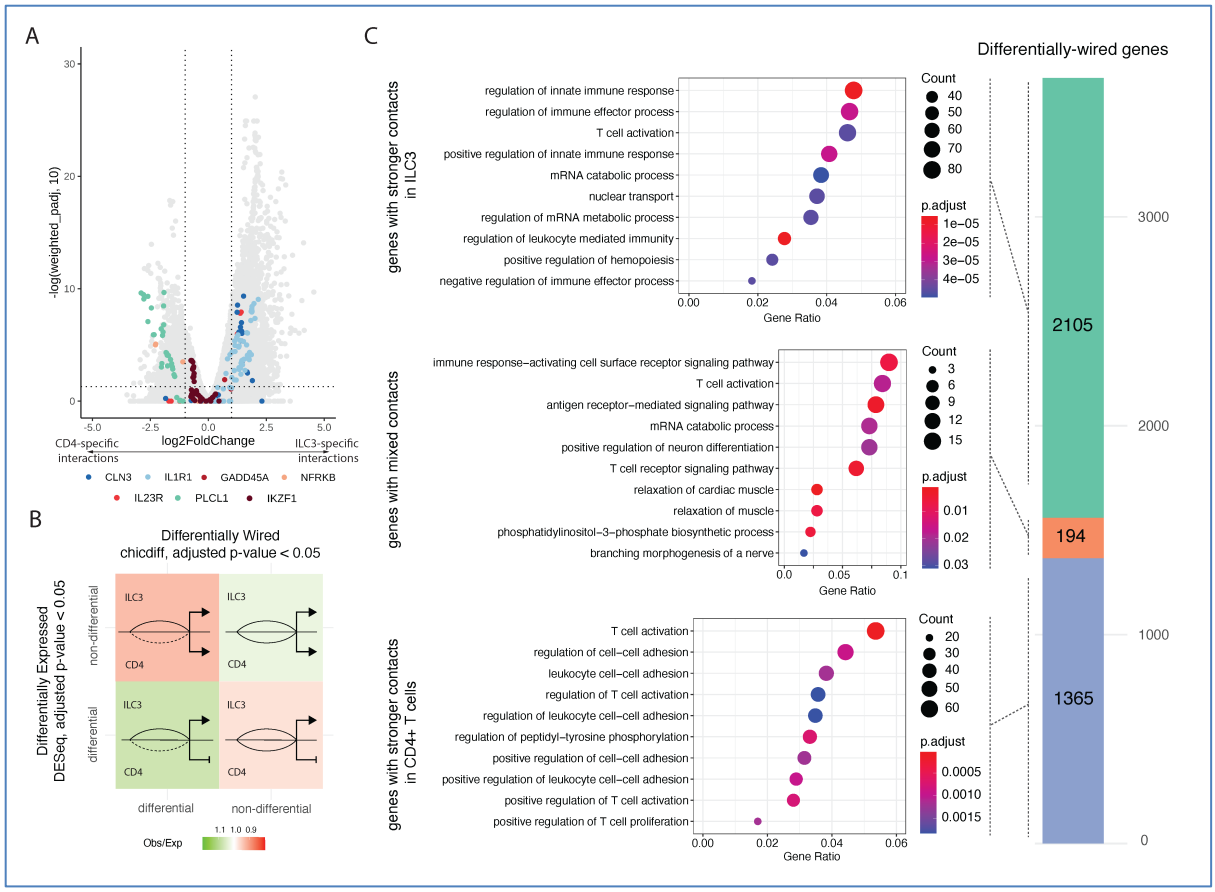

Figure 3. Differential enhancer-promoter interactions between ILC3s and CD4+ T cells. **A.** Volcano plot of differential interactions between ILC3s and CD4+ T cells detected by Chicdiff, highlighting those of selected immune-related genes (*CLN3*, *IL1R1*, *GADD45A*, *NFKB*, *IL23R*, *PLCL1*, *IKZF1*). **B.** Relationship between differential expression (DESeq2, adjusted p < 0.05) and differential wiring of promoter contacts (Chicdiff, adjusted p < 0.05). **C.** Gene Ontology enrichment analysis of genes with stronger contacts in ILC3s (top), CD4+ T cells (bottom) or a mixture of contacts that are stronger in either cell type (middle), showing biological processes related to immune cell activation, adhesion, and differentiation. Bubble size reflects the number of genes; colour indicates adjusted p-values. The bar plot shows the overlap between differentially wired genes (as evaluated by Chicdiff) in ILC3s and CD4+ T cells.

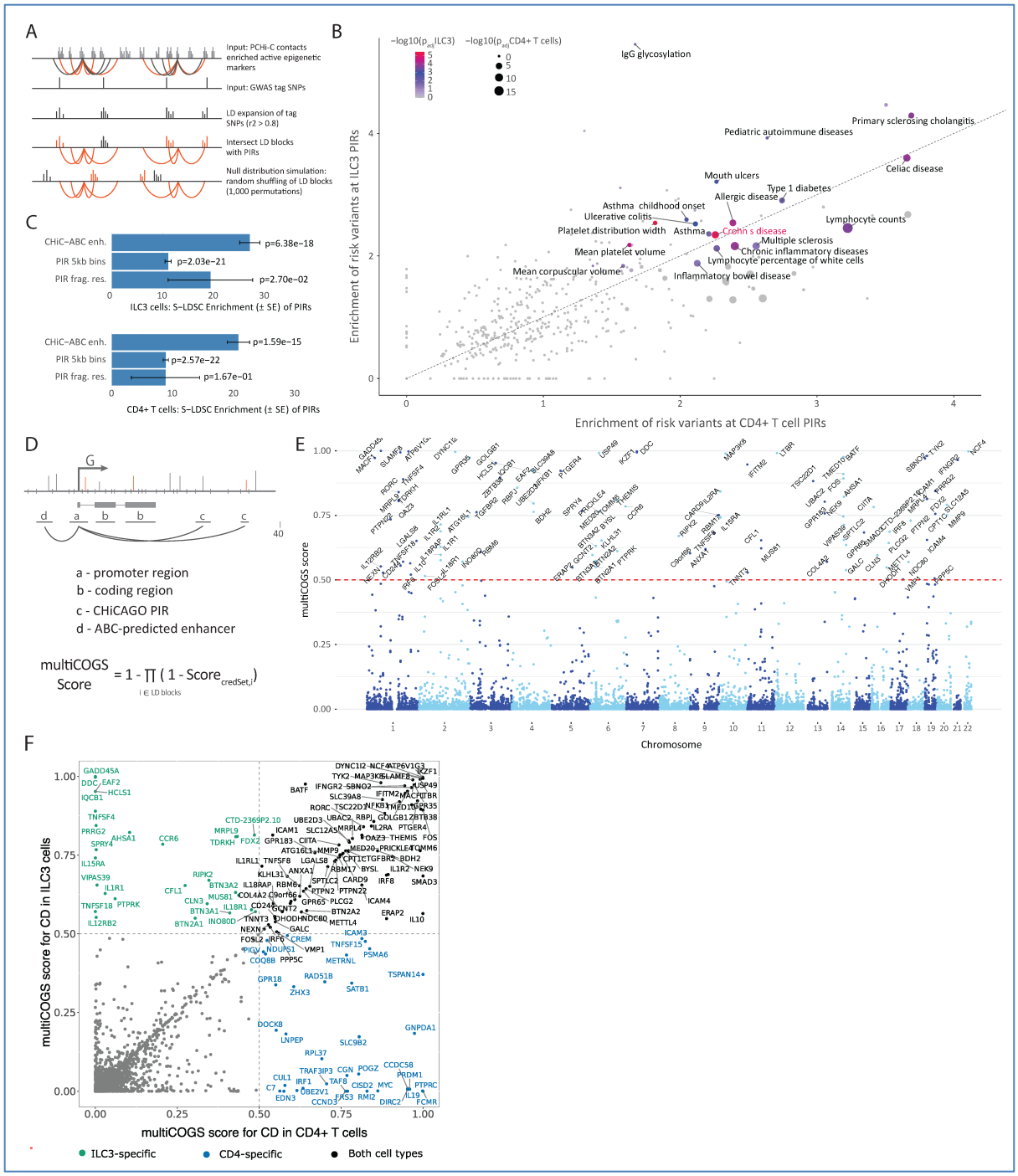

Figure 4. Statistical integration of PCHi-C results in ILC3s and CD4+ T cells with GWAS enables gene prioritisation for Crohn's disease (CD). A. Schematic of the RELI algorithm used for estimating the enrichment of genetic risk loci within PIRs. **B.** RELI enrichment of risk variants in ILC3 vs CD4+ T cell PIRS across 495 diseases and traits. Traits with $\log_{10}$(BH corrected p-value in ILC3s) < 0.05 are labelled. **C.** Stratified LD score regression analysis for enrichment of CD risk heritability at PIRs of ILC3s and CD4+ T cells. **D.** Schematic of the multiCOGS algorithm. **E.** Manhattan plot of multiCOGS gene prioritisation scores for CD risk based on GWAS integration with promoter interactions in ILC3s. Genes with multiCOGS scores above 0.5 are labelled. **F.** Comparison of multiCOGS scores for CD obtained with promoter interactions detected in ILC3s and CD4+ T cells. Prioritised genes are labelled in green (multiCOGS scores > 0.5 in ILC3s only), blue (multiCOGS scores > 0.5 in CD4+ T cells only) and black (multiCOGS scores > 0.5 in both cell types). All other genes are shown as grey dots.

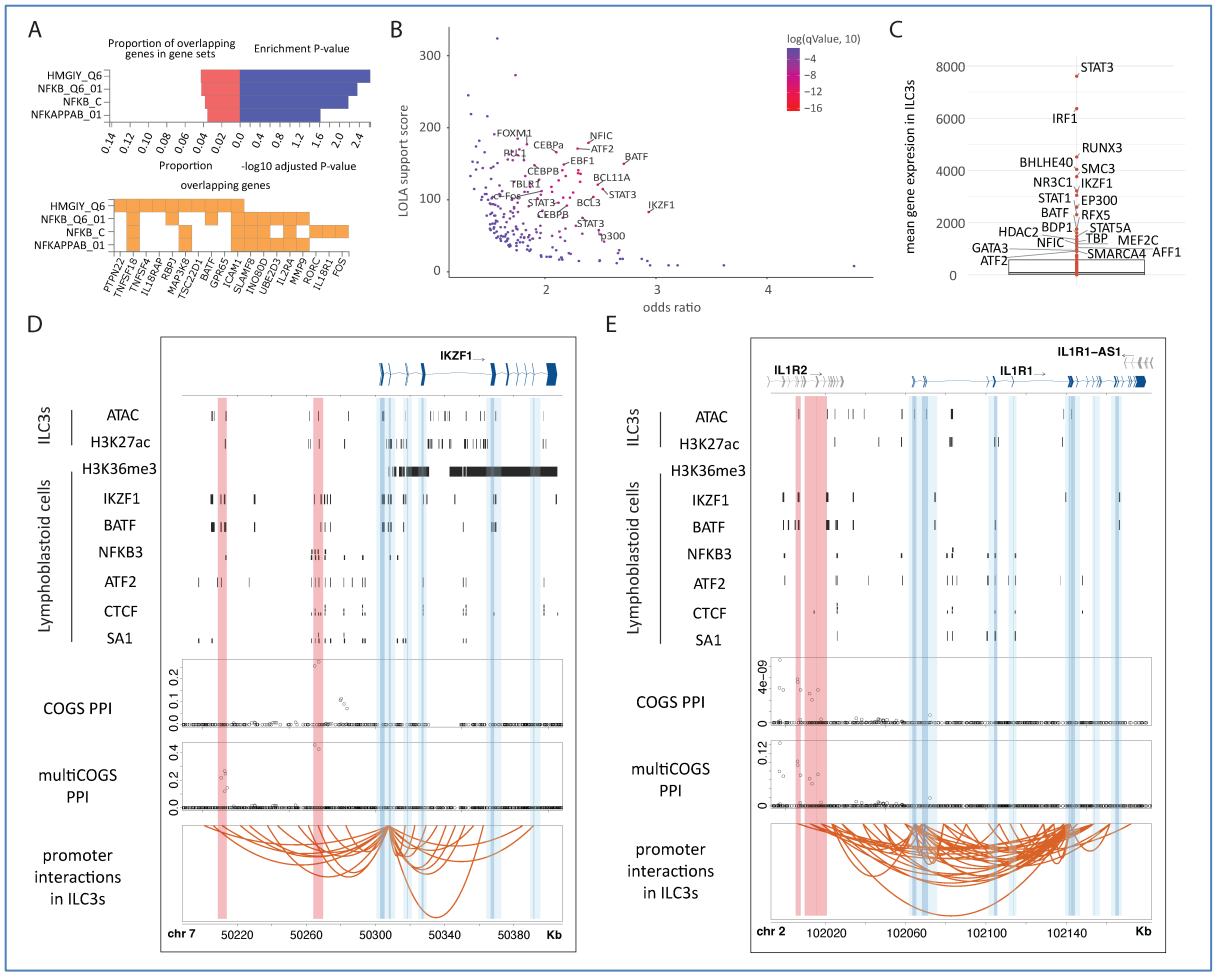

Figure 5. Characterisation of genes associated with CD risk prioritised by multiCOGS in ILC3s and their putative TF regulators. A. Significant sets of multiCOGS-prioritised genes predicted to bind specific TFs in their promoter regions, according to the MSigDB TF targets database, detected using the GENE2FUNC pipeline in FUMA[121]. TF sets are labelled (rows), with the proportion of all multiCOGS genes per set and the associated p-values shown on the top panel, and the gene names on the bottom panel. **B.** Enrichment analysis for TF binding sites at active PIRs for genes prioritised by multiCOGS vs active PIRs of all genes submitted to multiCOGS analysis. **C.** Expression of TFs enriched at the PIRs of prioritised genes. Outliers are removed for clarity. **D and E.** Examples of genes prioritised by multiCOGS for CD (*IKZF1*, and *IL1R1*), showing patterns of TF binding in lymphoblastoid cell lines, and posterior probability profiles of classic COGS and multiCOGS. Vertical dark blue and light blue bands, respectively, highlight annotated gene promoters and promoter-proximal regions (+/- 5 restriction fragments) considered in (multi)COGS analysis in addition to PIRs. Vertical red bands highlight PIRs harbouring CD risk-associated SNPs with high posterior probability of inclusion. Orange arcs correspond to significant interactions (CHiCAGO score > 5) at 5kb resolution for *IKZF1* (E) and *IL1R1* (F), respectively.

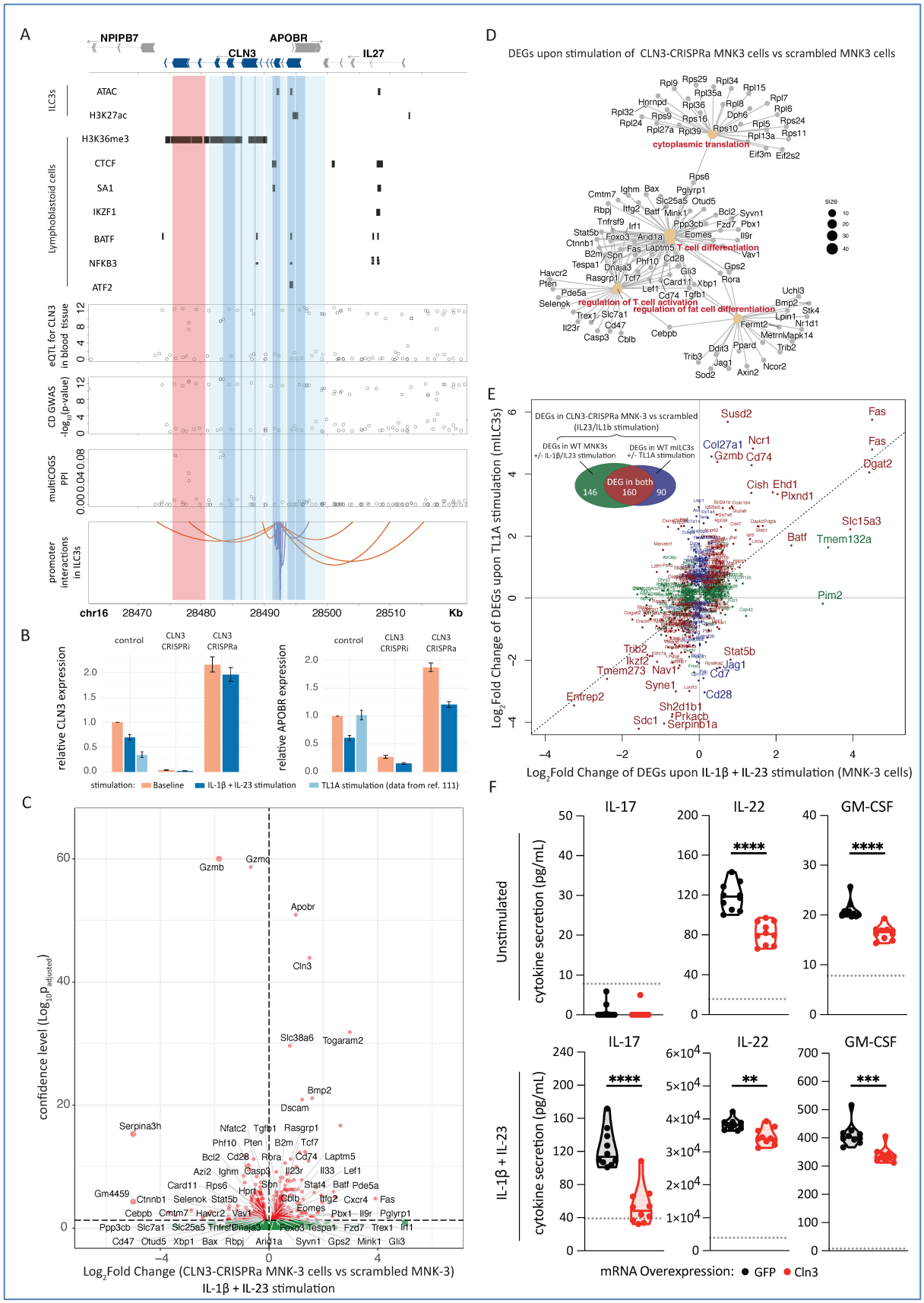

Figure 6. Evidence for the role of CLN3 in ILC3 inflammatory function. A. Interaction profile of the human *CLN3* promoter alongside the tracks of TF binding, blood eQTLs, CD GWAS and SuSiE posterior probabilities of inclusion. Dark blue and light blue bands, respectively, highlight the locations

of annotated *CLN3* promoters and promoter-proximal regions (+/- 5 restriction fragments) considered by multiCOGS in addition to PIRs. Red band highlights the ILC3-specific PIR containing CD-associated SNPs with high posterior probability of inclusion. Orange and purple arcs, respectively, depict significant interactions (CHiCAGO score > 5) in ILC3s at 5kb and single-fragment resolution. **B.** Up- and down-regulation of *Cln3* and *Apobr* upon TL1A stimulation in mouse primary ILC3s (RNA-seq data from Ref.[108]) and upon IL-23/IL-1β stimulation in CLN3-targeted CRISPRi and CRISPRa MNK-3 cells (RNA-seq data from this study). **C.** Differential expression of genes in IL-23/IL-1β-stimulated *Cln3*-CRISPRa MNK-3 cells relative to scrambled gRNA controls. Red dots - differentially expressed genes (stimulated *Cln3*-CRISPRa DEGs, DESeq2 adjusted p-value < 0.05), with other genes shown as green dots. **D.** Network-style representation of GO term enrichment analysis of stimulated *Cln3*-CRISPRa DEGs. **E.** Changes in the expression of stimulated *Cln3*-CRISPRa DEGs (dots) upon either IL-23/IL-1β or TL1A stimulation of unperturbed MNK-3 cells (data from Ref.[108]). **F.** Evidence that *Cln3* overexpression decreases inflammatory cytokine secretion. MNK-3 cells were electroporated with GFP mRNA (black) or Cln3-myc mRNA (red), then cultured either unstimulated (top row) or stimulated with IL-1β and IL-23 (bottom row) for 24 hr. Cytokine concentrations (IL-17, IL-22, GM-CSF) in culture supernatants were quantified by ELISA. Each point represents an individual biological replicate (n=10 per condition). The data shown are from one representative experiment of three independent experiments performed. Dotted line indicates the lower limit of quantification for each assay. Statistical significance was assessed using an unpaired Welch's t-test. p<0.01 (**), p<0.001 (***), p<0.0001 (****).

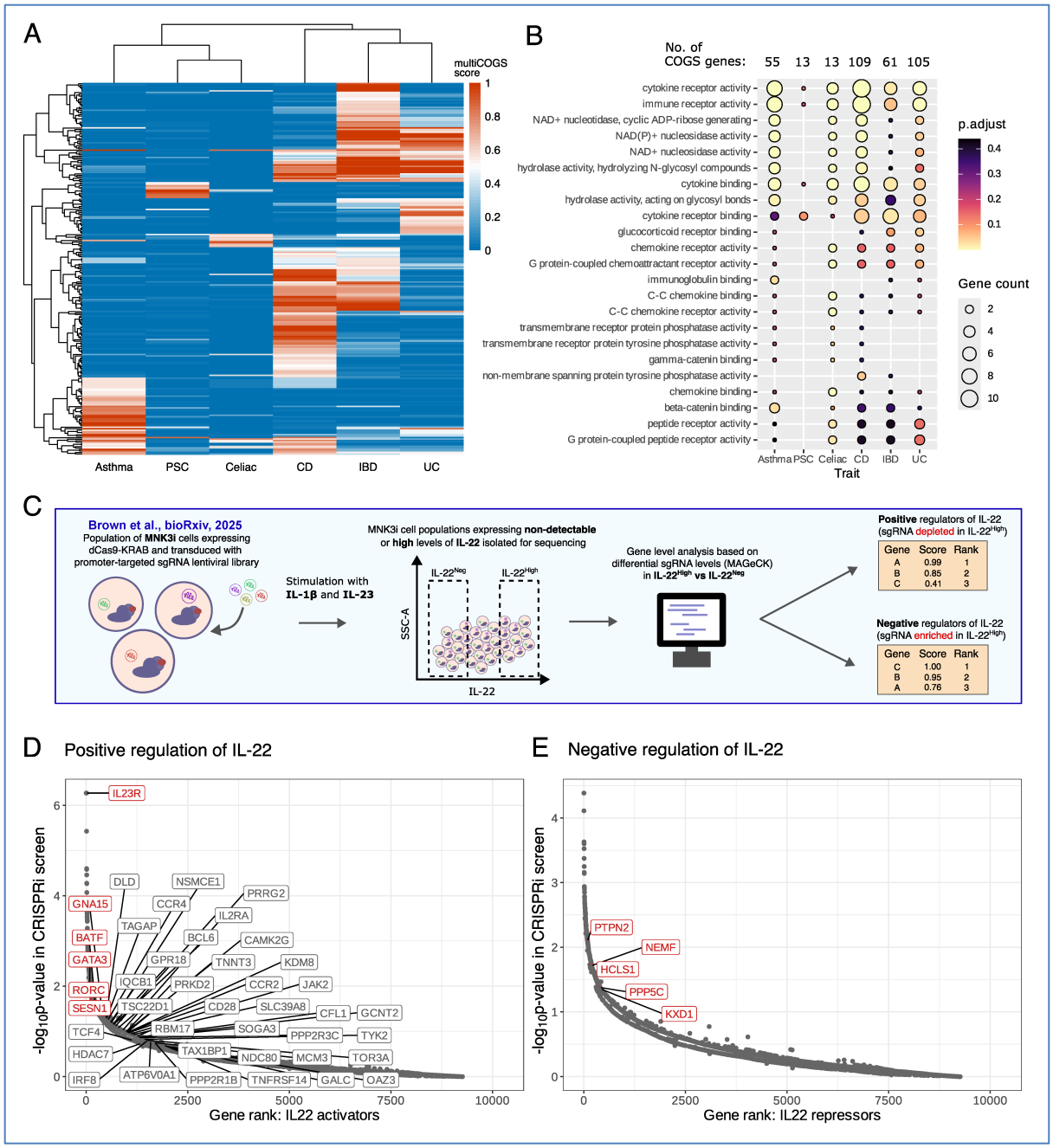

Figure 7. A compendium of prioritised genes in ILC3s for six autoimmune diseases. **A.** MultiCOGS results across asthma, primary sclerosing cholangitis (PSC), Celiac Disease, Crohn's Disease (CD), Inflammatory Bowel Disease (IBD) and Ulcerative Colitis (UC) in ILC3 cells. Rows represent each gene that scored at least 0.5 in one of the traits. Colours show the multiCOGS score in each trait. Clustering on genes (rows) and traits per cell type (columns) is based on Euclidean distance. **B**. Significant hallmark pathways identified in at least one of the traits in ILC3 cells by GO term analysis. **C.** Schematic of the MNK-3 CRISPRi screen for detecting genes involved in the regulation of IL-22 signalling[125]. **D.** multiCOGS genes for all six traits visualised among the CRISPRi results, which are ranked by evidence of positive IL-22 regulation in the MNK-3i cells. The multiCOGS genes with $p < 0.05$ in the screen are labelled in red. MultiCOGS genes driving GSEA signal ("leading edge") are labelled in grey. **E.** Similar to D, but for genes ranked by score for negative IL-22 regulation in the MNK-3i screen. Red labels indicate multiCOGS genes significant in the screen at $p < 0.05$. Since GSEA for multiCOGS genes among IL-22 repressors was not significant, the leading edge genes are not labelled. CD: Crohn's Disease, IBD: Inflammatory Bowel Disease, GSEA: Gene Set Enrichment Analysis, PSC: primary sclerosing cholangitis, UC: Ulcerative Colitis.

Supplementary Figures

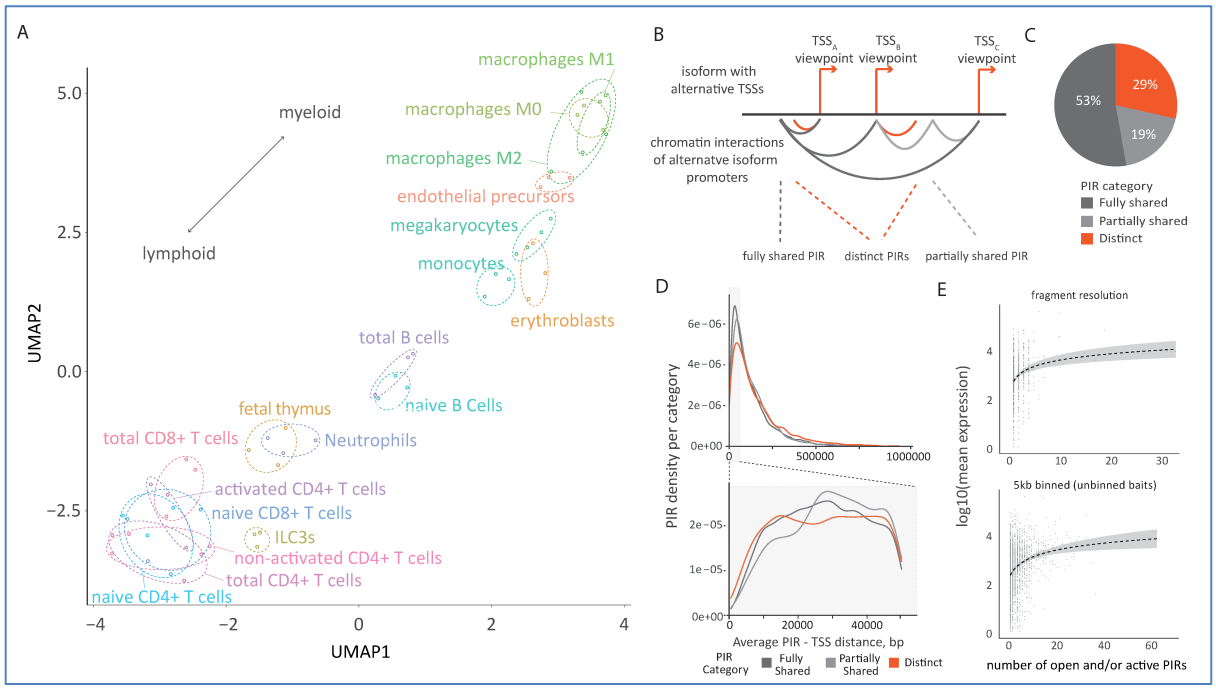

Figure S1. Compendium of promoter-enhancer interactions in ILC3s. A. UMAP of CHiCAGO scores detected for PCHi-C in ILC3s versus public data in 17 primary human blood cell types[30]. **B.** Scheme representing the classification of PIRs detected at alternative transcription start sites (ATSS) of the same gene: 'fully shared' (shared across all captured ATSSs), partially shared and distinct (unique to a single ATSS). **C.** Pie chart showing the degree of enhancer sharing across alternative transcription start sites (ATSS) for short-range contacts. **D.** Distance distribution of ATSS-specific and shared PIRs at 5kb binned (baits unbinned) resolution. Top panel - interactions up to 1Mb (Kruskal-Wallis test $p < 2.22e-16$; pairwise Wilcoxon test $p = 8.68e-6$ [partially shared vs fully shared], $p = 4.46e-8$ [partially shared vs distinct] and $p < 2.22e-16$ [fully shared vs distinct]; bottom panel - interactions up to 50kb (Kruskal-Wallis test $p = 7.65e-5$; pairwise Wilcoxon test $p = 9.8e-5$ [partially shared vs fully shared], $p = 6e-4$ [partially shared vs distinct] and $p = 1$ [fully shared vs distinct]). **E.** Correlation between gene expression and number of regulatory elements identified in CHiCAGO PIRs at fragment and 5kb (solitary baits) resolution.

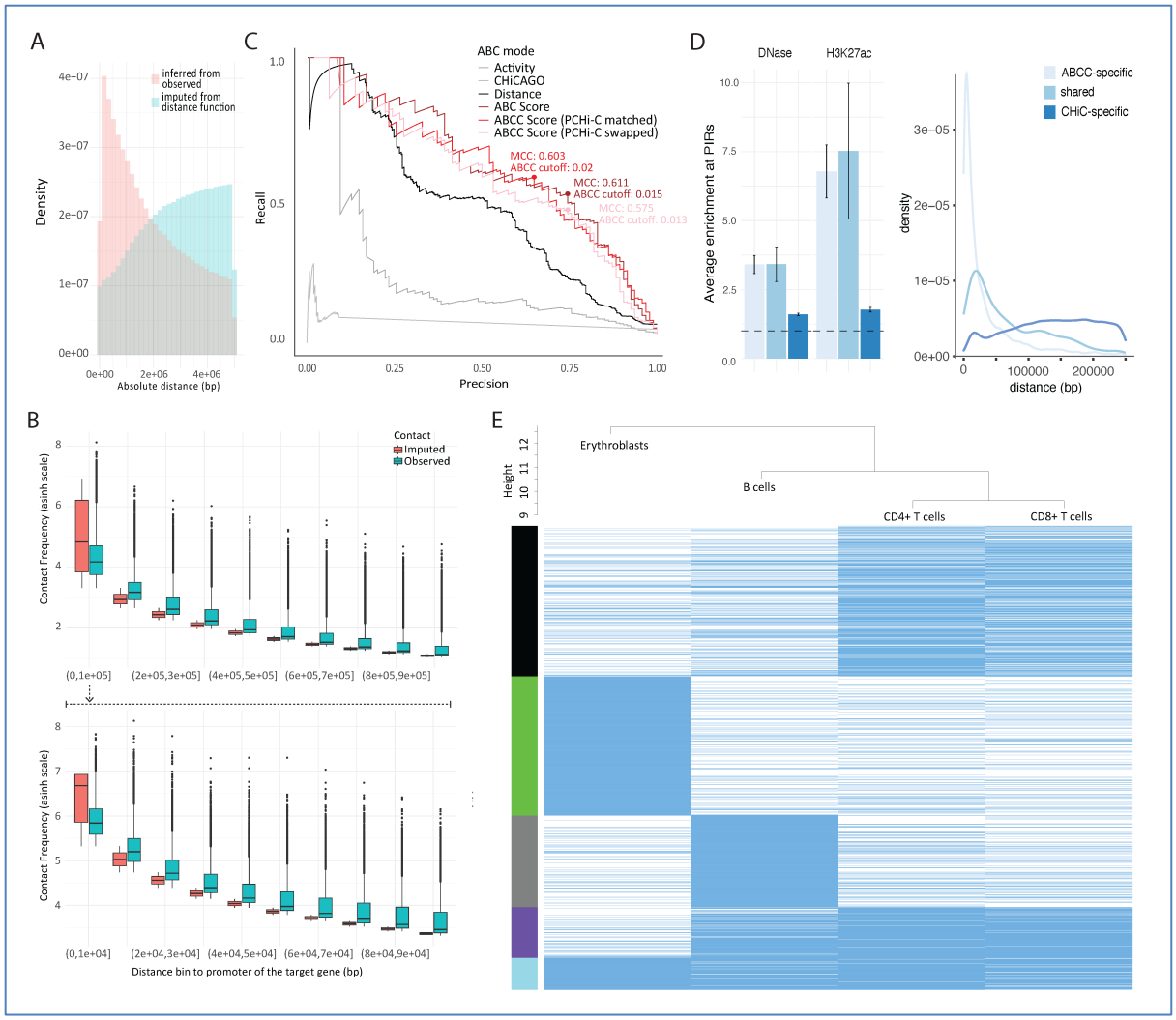

Figure S2. Benchmarking the ABCC approach with public data. A. Density distribution of promoter interactions inferred from observed PCHi-C contact frequencies (pink) and those imputed using the CHiCAGO distance function (cyan) across genomic interaction distances. **B.** Contact frequency distributions stratified by distance. Observed PCHi-C contacts are shown in green, imputed contacts (using expected frequencies estimated using the CHiCAGO distance function) are shown in blue. Similarly to standard ABC, frequency capping is introduced for short-range imputed contacts (<5kb). **C.** Precision–recall curves benchmarking the predictive performance of different scoring approaches for enhancer–promoter interactions in erythroblasts. Curves compare the scoring across: CHiCAGO-detected contacts, Activity alone, Distance alone, the conventional ABC score, and PCHi-C-based ABCC score in two modes: "matched" - using PCHi-C cell-type specific profile for erythroid cells and "swapped", in which a PCHi-C dataset with a similar read coverage from a different cell type, CD4+ T cells, is used instead. MCC: Matthews correlation coefficient, an alternative to the AUC metric that is more informative under class imbalance and more sensitive to performance at a fixed decision threshold[175]. **D.** Enrichment of epigenetic markers at PIRs: DNase - chromatin accessibility and H3K27ac - active enhancers (left panel) and distance distribution of ABCC-specific, PCHiC-specific and shared enhancer-promoter links (right panel) in K562 cells for 0.023 ABCC threshold. **E.** Hierarchical clustering heatmap of enhancer–promoter interactions predicted with ABCC across cell types (erythroblasts, B cells, CD4+ T helper cells, CD8+ T cells).

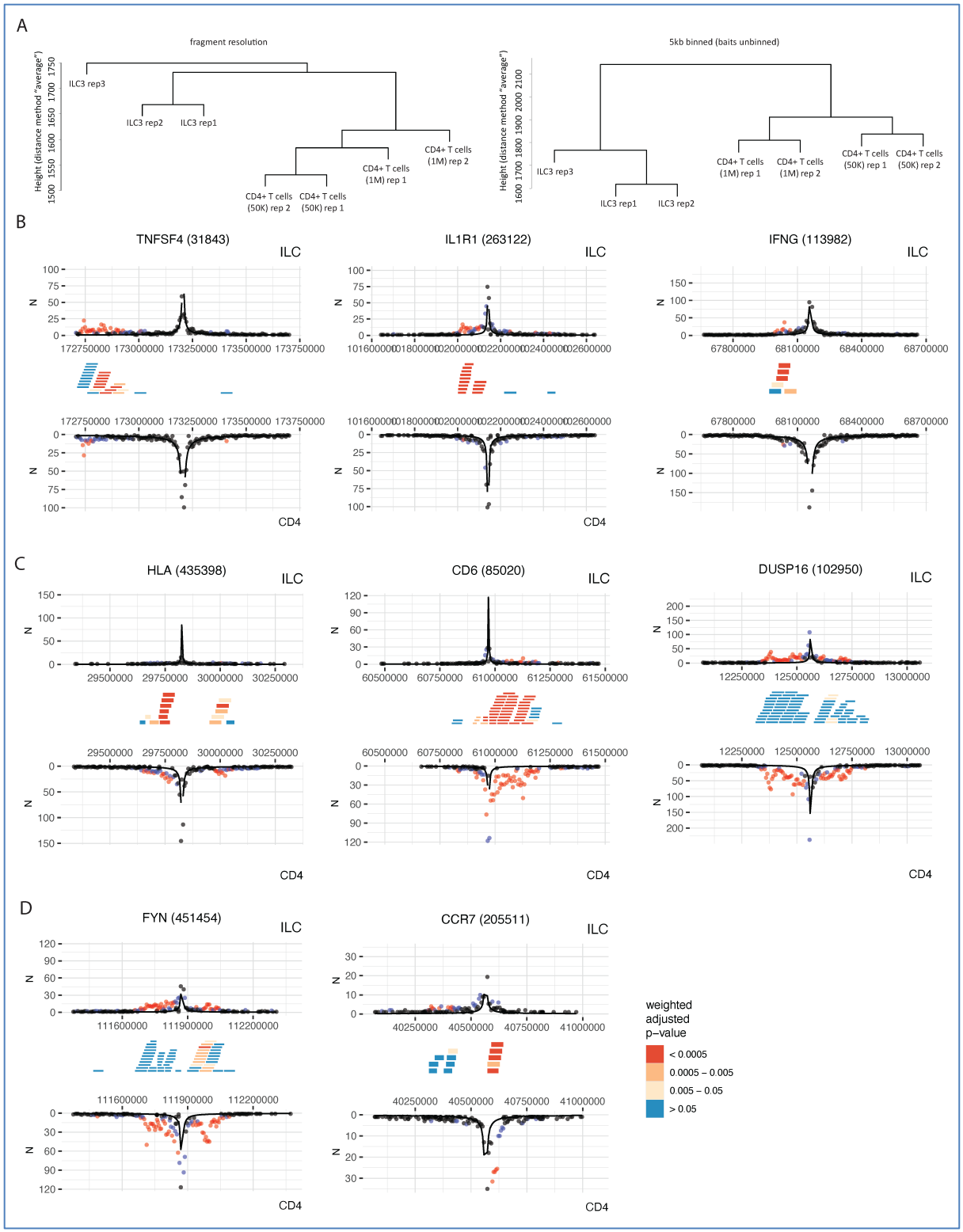

Figure S3. Genes with differential contacts in ILC3s and CD4+ T cells. A. Hierarchical clustering of ILC3s and CD4+ T cells PCHi-C datasets. **B-D.** Examples of captured promoters with differential wiring between ILC3s and CD4+ T cells: promoters with stronger **(B)** and weaker **(C)** contacts in ILC3s compared with CD4+ T cells, as well as with both types of contacts **(D).**

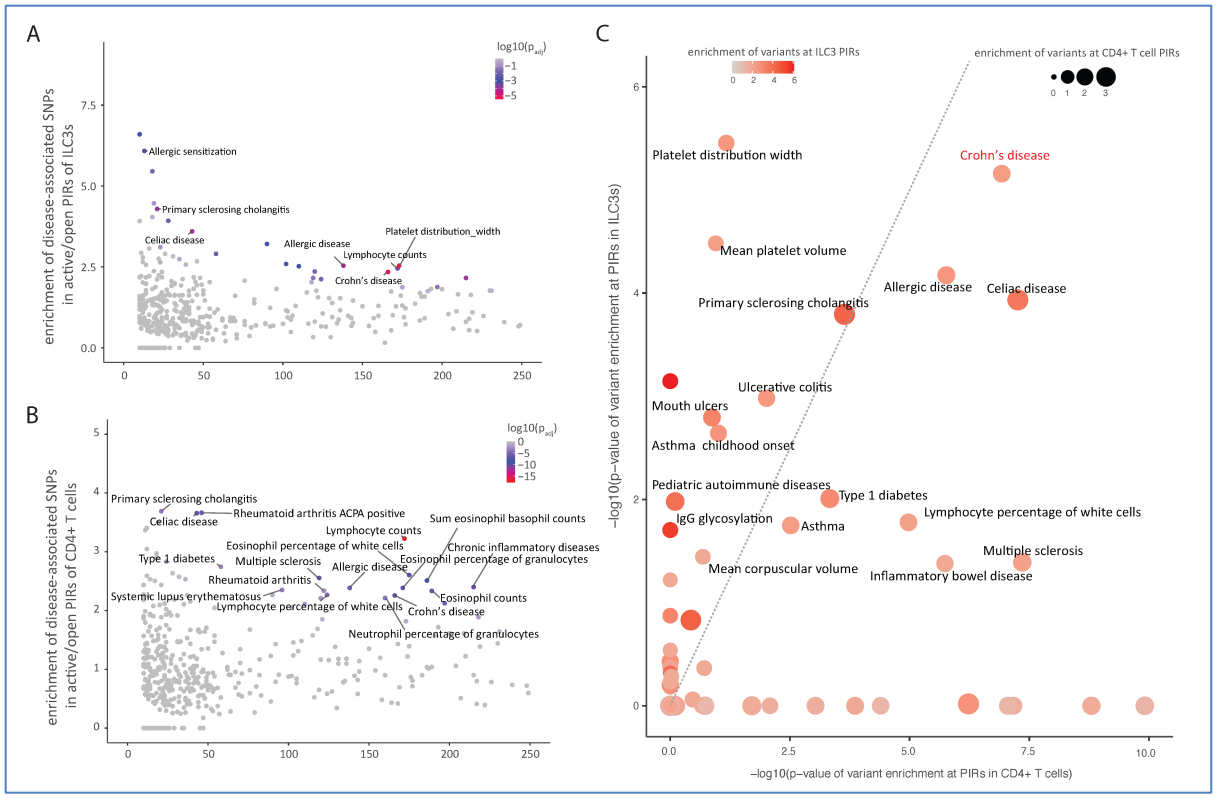

Figure S4. Supplementary information for the RELI analysis of risk loci enriched in ILC3 and CD4+ T-cell PIRs. A-B. RELI enrichment of risk variants in ILC3s **(A)** and CD4s PIRs (**B**) across 495 diseases and traits. Traits with $\log_{10}$(BH corrected p-value in ILC3s) < 0.001, number of loci per trait > 10, and enrichment > 2.2 are labelled. **C.** Adjusted p-value of RELI enrichment of risk variants ILC3s vs CD4s PIRs across 495 diseases and traits. Traits with $\log_{10}$(BH corrected p-value in ILC3s) < 0.05 are labelled.

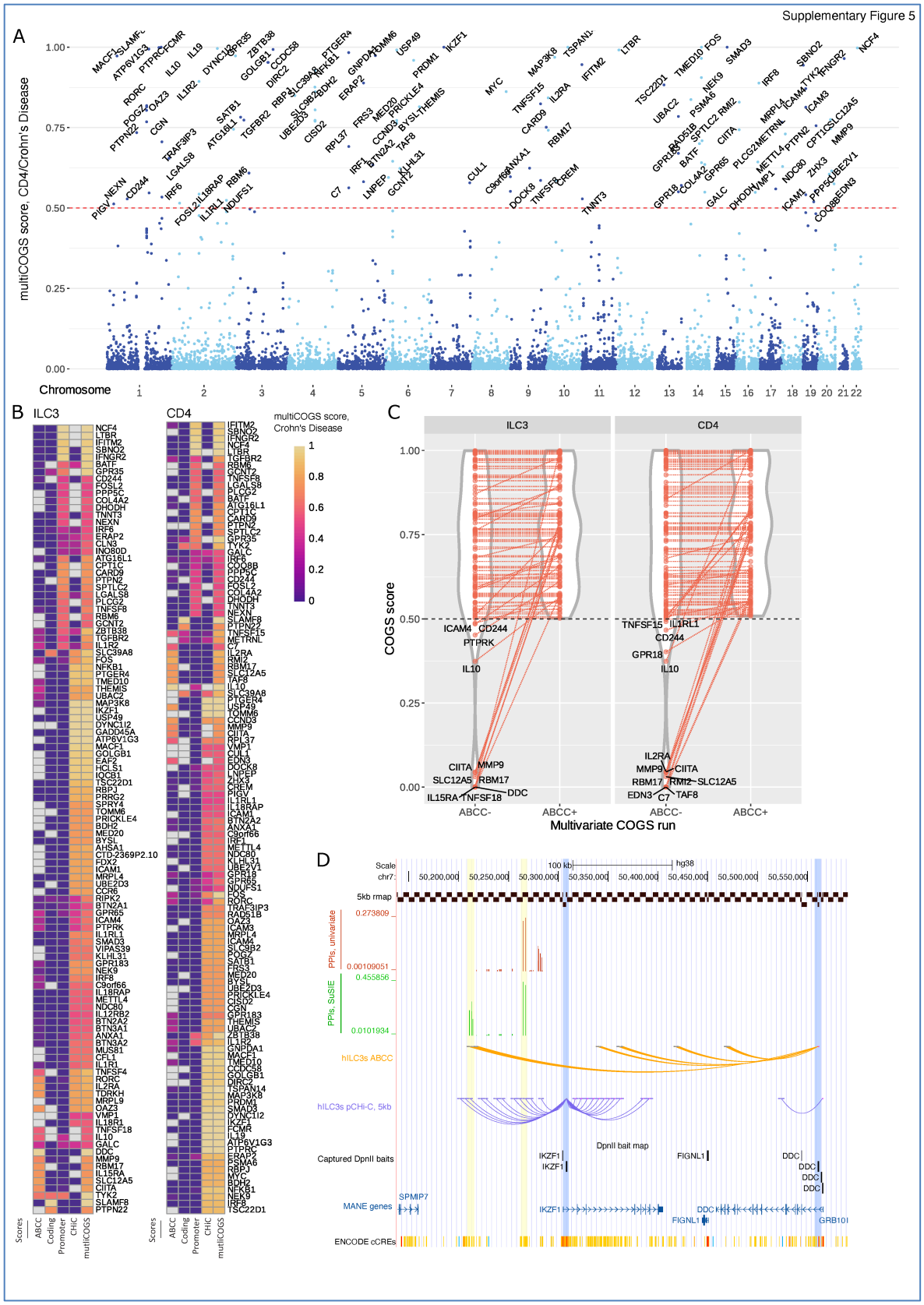

Figure S5. MultiCOGS prioritises gene sets in Crohn's Disease. **A.** Manhattan plot showing multiCOGS for CD risk based on promoter contacts in CD4+ T cells. **B.** Heatmaps of region contributions to multiCOGS scores in ILC3s and CD4s in CD. **C.** Illustration of genes that were only prioritised for CD with the addition of ABCC, in ILC3s and CD4s. In each graph, the multiCOGS score

with and without ABCC is plotted for all genes that were prioritised in the full multiCOGS run (score > 0.5 with ABCC). **D.** Illustration of multiCOGS prioritisation of *IKZF1* and *DDC* in ILC3s in the 7p locus. In this locus, multivariate fine mapping identifies two credible sets of variants (yellow bars), whereas univariate fine mapping only detects one. PCHi-C interactions connect these likely causal variants to the *IKZF1* promoter (first blue bar). However, ABCC interactions also connect one of the credible sets to the DDC promoter (second blue bar). Thus, multiCOGS prioritises both genes, whereas classic COGS prioritises only *IKZF1*.

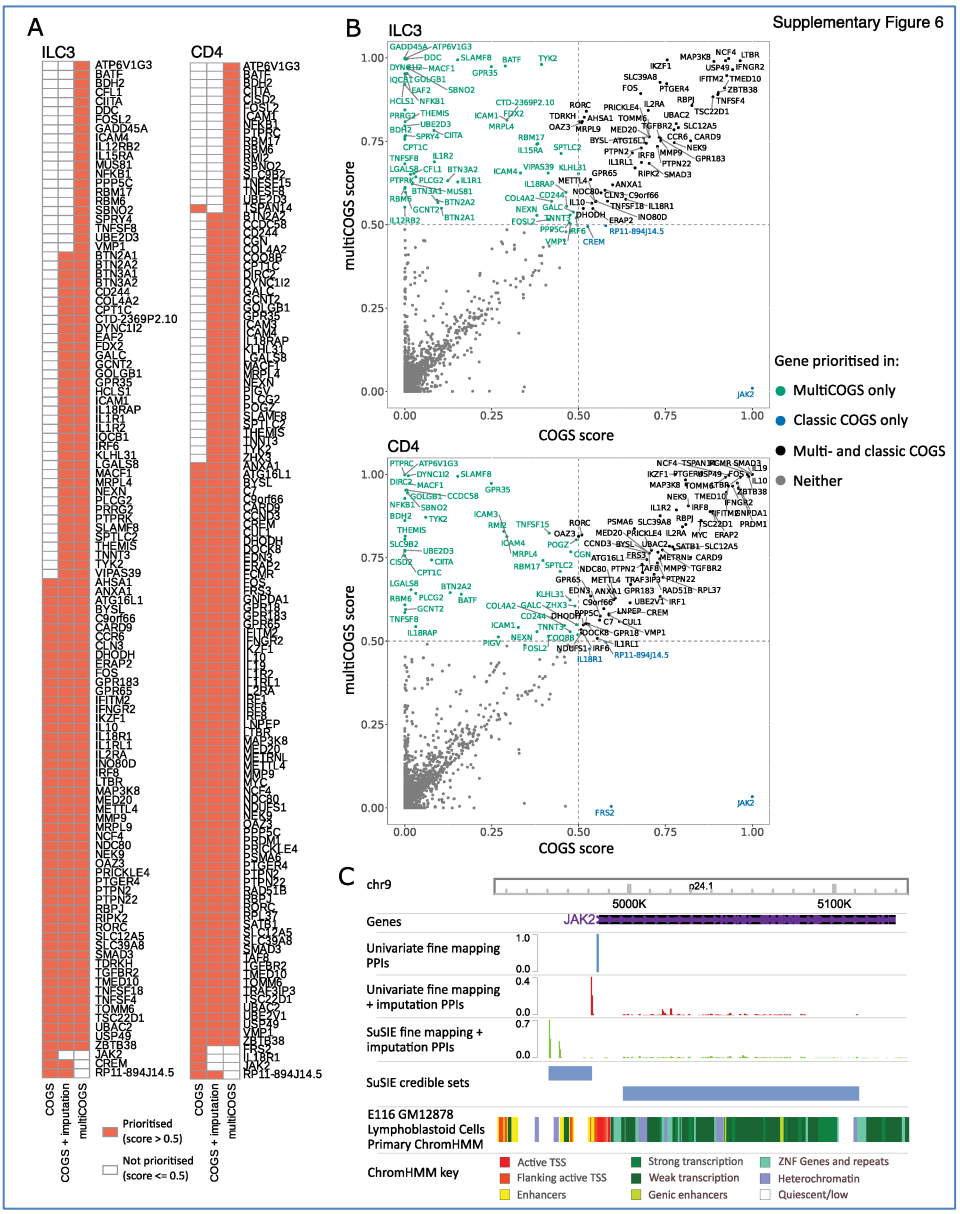

Figure S6. Comparison of gene prioritisation for Crohn's Disease in classic COGS versus multiCOGS. A. Comparison of prioritised gene sets between classic COGS, classic COGS plus imputation, and multiCOGS (i.e. imputation plus multivariate fine mapping, processed via the multiCOGS algorithm) for CD. Shown for ILC3s and CD4+ T cells. **B.** Comparison of COGS scores and multiCOGS scores for genes in ILC3 cells (top) and CD4+ T cells (bottom) for CD. Green labels indicate genes prioritised in multiCOGS only, blue in classic COGS only, and black in both. **C.** Plot of the *JAK2* locus, showing the shift of the most likely causal variant from the promoter of *JAK2* to a region around 20kb upstream of the promoter upon multivariate fine mapping, leading to a lower multiCOGS vs classic COGS score. No chromosomal interactions were observed between this region and the *JAK2* promoter in ILC3 or CD4+ T cells.

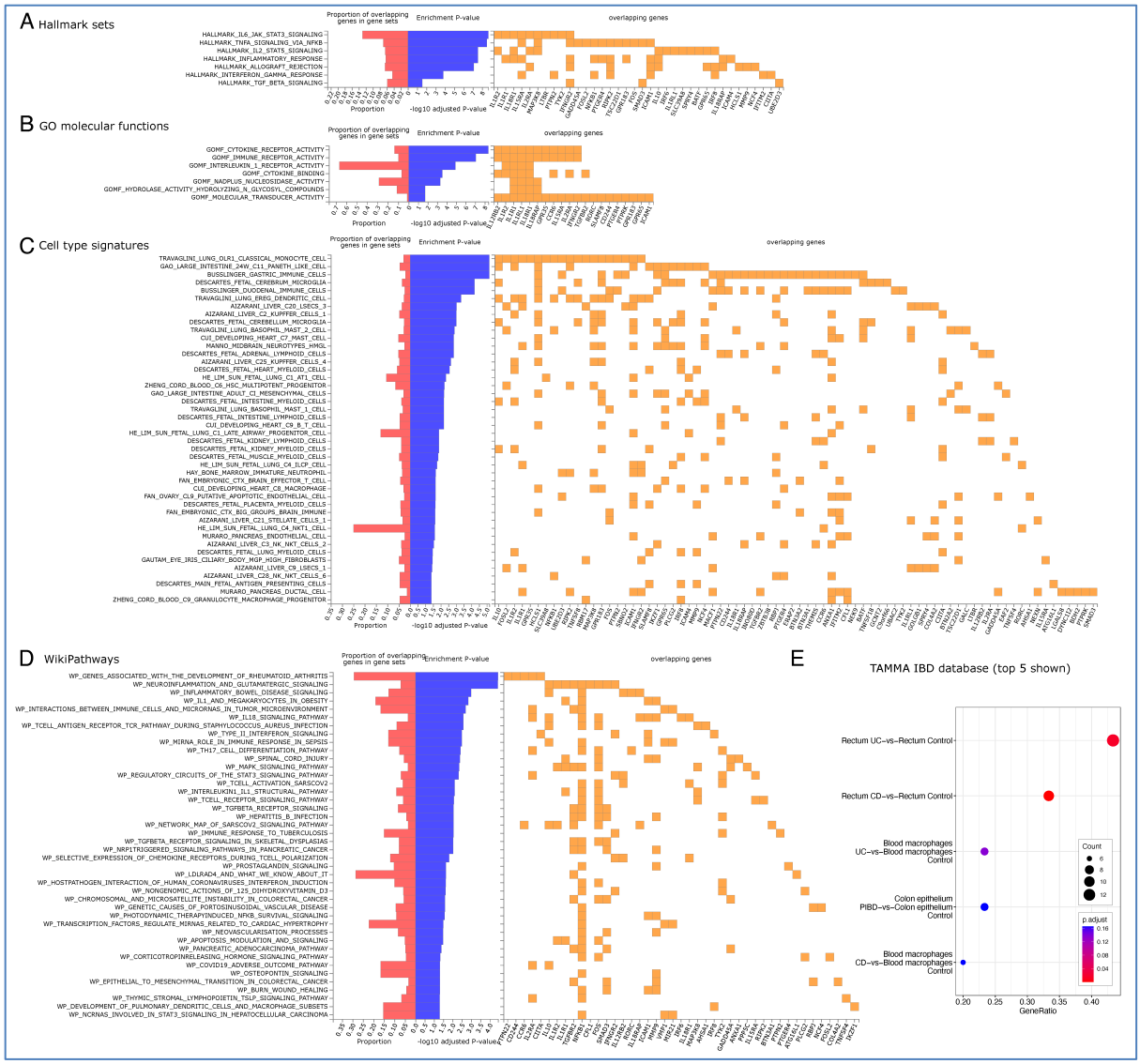

Figure S7. Biological annotation of multiCOGS CD genes in ILC3s. A-D. Enriched gene sets among multiCOGS genes detected using the GENE2FUNC pipeline in FUMA[121], for the following databases: (**A**) MSigDB hallmark sets, (**B**) GO molecular functions, (**C**) MSigDB cell type signatures, (**D**) MSigDB WikiPathways. **E**. Enrichment analysis for differentially expressed gene sets among multiCOGS genes in the TAMMA IBD database.

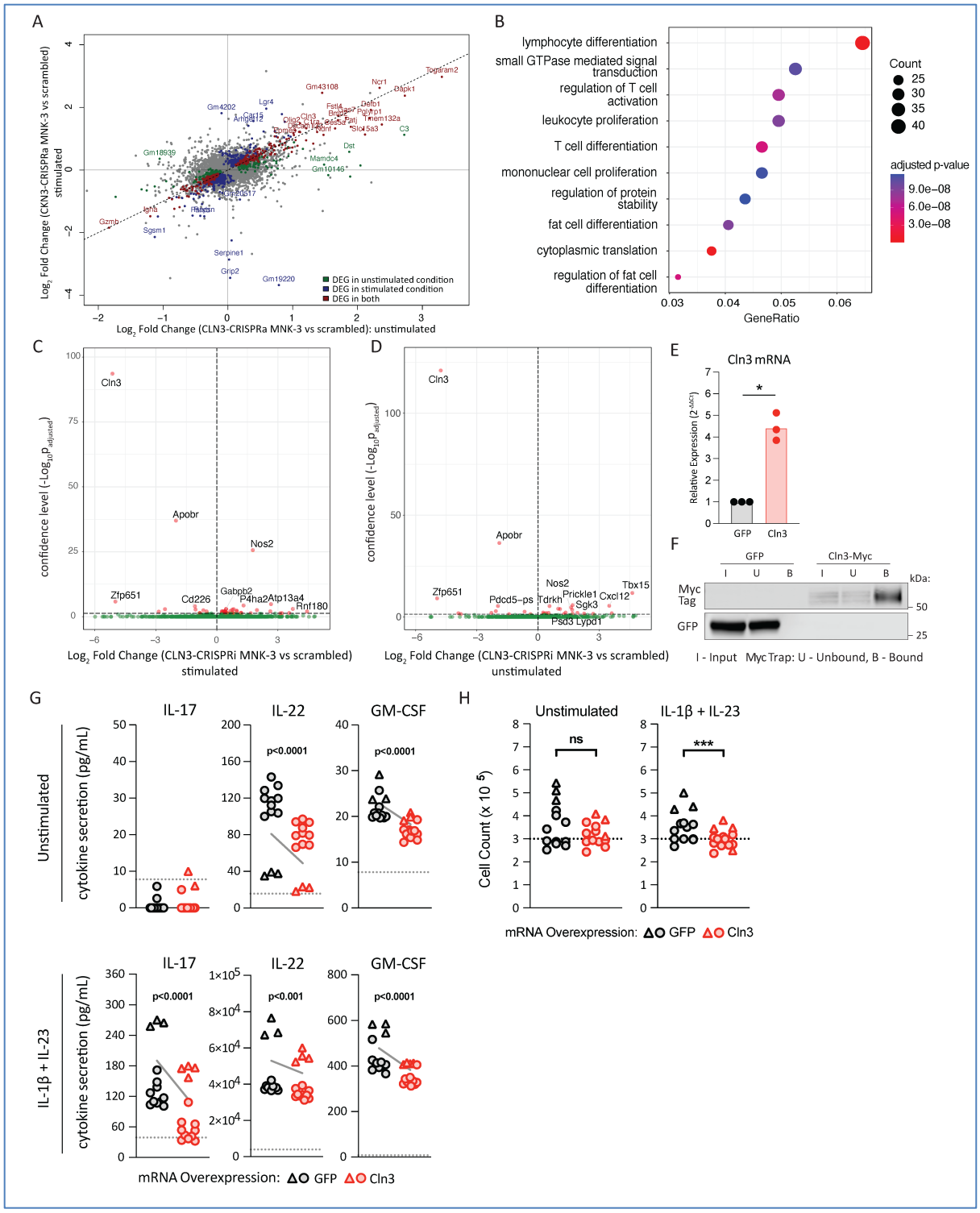

Figure S8. Additional information on the role of *Cln3* in ILC3 inflammatory function.
A. Comparison of differentially expressed genes in IL23/IL-1βl-stimulated vs unstimulated *Cln3*-CRISPRa cells (relative to scrambled gRNA controls). **B.** GO term enrichment analysis for genes differentially expressed upon CLN3-CRISPRa stimulation. **C.** Differential expression of genes in IL23/IL-1βl-stimulated *Cln3*-CRISPRi MNK-3 cells vs scrambled gRNA controls. Red - differentially expressed genes (DESeq2 adjusted p-value < 0.05), green - all other genes. **D.** Differential expression of genes in unstimulated *Cln3*-CRISPRi MNK-3 cells vs scrambled gRNA controls. Red - differentially expressed genes (DESeq2 adjusted p-value < 0.05), green - all other genes. **E.** *Cln3* expression in MNK-3 cells electroporated with Cln3-myc mRNA or GFP mRNA. Transcript abundance was quantified by qPCR,

normalised to *Hprt*, and expressed relative to the GFP mRNA control. Each point represents an independent experiment. Statistical significance was assessed using a paired Welch's t-test, $p<0.05$ (*). **F.** Verification of CLN3-myc protein expression and Myc tag-dependent pulldown. MNK-3 cells were lysed, subjected to immunoprecipitation using Myc-Trap agarose, and resolved by reducing SDS-PAGE. "I" = input lysate; "U" = unbound fraction; "B" = bead-bound fraction. Immunoblotting with anti-myc tag antibody detected a ~65–80 kDa Cln3-myc species selectively enriched in the bound fraction. **G.** Cytokine secretion upon *Cln3* overexpression across independent experiments. MNK-3 cells were electroporated with GFP mRNA (black) or Cln3-myc mRNA (red) and cultured for 24 hr under unstimulated (top row) or IL-1β + IL-23–stimulated (bottom row) conditions. Cytokine concentrations (IL-17, IL-22, GM-CSF) in culture supernatants were quantified by ELISA. Each symbol represents a biological replicate from two independent experiments (triangles vs circles). The dotted horizontal line indicates the lower limit of quantification for each assay. Statistical significance was assessed using a linear mixed-effects model with experiment as a random effect and transfection as a fixed effect (n=13–14 per condition). Solid grey lines indicate group means, with dotted grey bands indicating 95% confidence intervals of the fixed-effect. **H.** Cell numbers upon *Cln3* overexpression with and without inflammatory stimulation. MNK-3 cells were electroporated with GFP mRNA (black) or Cln3-myc mRNA (red) and cultured for 24 hr in unstimulated (left) or IL-1β + IL-23–stimulated (right) media. Each symbol represents a biological replicate from two independent experiments (triangles vs circles). Lines connect the experiment-specific means. Viable cell numbers were quantified by trypan blue exclusion. The dotted line indicates the number of cells seeded at 0 hr. Statistical significance was assessed using a linear mixed-effects model with experiment as a random effect and transfection as a fixed effect (n=13–14 per condition). Not significant (ns), $p<0.001$ (***).

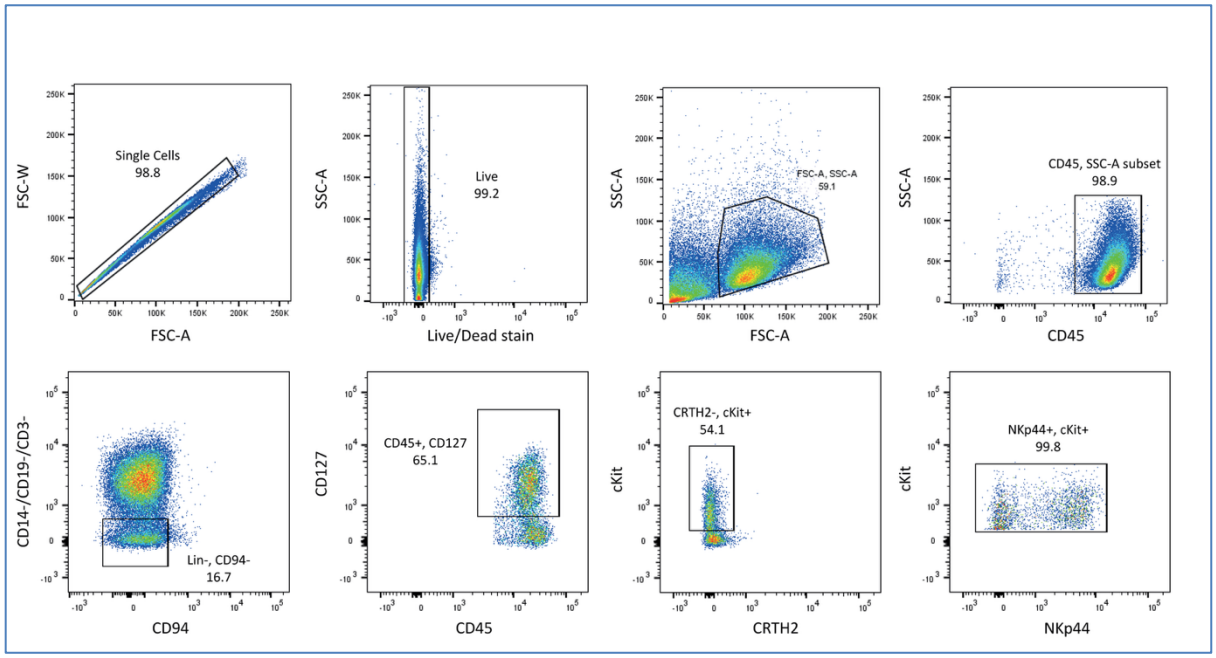

Figure S9. Flow cytometry gating strategy for isolation of human ILC3s from tonsils.

List of supplementary tables

Table S1. PCHi-C quality metrics.
Table S2. Pathway enrichment for genes with ILC3-specific PIRs
Table S3. Pathway enrichment for genes with CD4-specific PIRs
Table S4. Pathway enrichment for genes with differential PIRs between ILC3 and CD4+ T cells
Table S5. Pathway enrichment for genes with non-differential PIRs between ILC3 and CD4+ T cells
Table S6. RELI results
Table S7. Candidate genes prioritised by multiCOGS in ILC3 cells and CD4+ T cells for Crohn's Disease.
Table S8. Prior evidence for candidate genes prioritised by multiCOGS in ILC3 cells and CD4+ T cells for Crohn's Disease.
Table S9. Pathway enrichment for multiCOGS-prioritised CD candidate genes in ILC3s
Table S10. LOLA results for TF enrichment within the PIRs of CD genes in ILC3s
Table S11. Differentially expressed genes upon CRISPR perturbations targeted to the *Cln3* promoter in MNK-3 cells
Table S12. MultiCOGS results in ILC3 and CD4+ T cells across 6 autoimmune traits
Table S13. The biological functions of multiCOGS-prioritised genes across 6 autoimmune traits in ILC3s
Table S14. sgRNA and primer sequences for Cln3 CRISPR targeting

Supplementary Note 1.

MultiCOGS resulted in loss of five candidate genes in one or both cell types, compared with classic COGS (*JAK2*, *CREM*, *FRS2*, *IL18R1* and *RP11-894J14.5*; see **Fig. S6B**). Of these, we were intrigued by the loss of *JAK2* in both cell types, because it is a well-noted candidate gene in IBD, with JAK inhibitors already used to treat ulcerative colitis and CD[172]. The COGS score for *JAK2* was substantially lower across both cell types when genetic imputation and multivariate fine mapping were employed (classic COGS score ~1 in both cell types, multiCOGS score ~0.01 in ILC3s and ~0.03 in CD4s). Upon examining the locus, we discovered that fine mapping with the univariate methodology (Wakefield synthesis[173]) identified the most likely causal variant as rs1887428 (PPI = 0.999) at the *JAK2* promoter, but summary statistic imputation combined with multivariate fine mapping (SuSIE[69]) prioritised the variant rs1327500 (PPI = 0.663), in a region ~20 kb upstream of *JAK2*, without detectable promoter contacts in ILC3 cells or CD4+ T cells (**Fig. S6C**). However, considering that both rs1887428 and rs1327500 are eQTLs for *JAK2* in blood cells, according to eQTLGen[174], *JAK2* remains a strong candidate in this locus by genetic association.