Genome-wide enhancer maps link risk variants to disease genes

Nasser, J

2021-05-13

Nasser , J , Bergman , DT , Fulco , CP , Guckelberger , P , Doughty , BR , Patwardhan , TA , Jones , TR , Nguyen , TH , Ulirsch , JC , Lekschas , F , Mualim , K , Natri , HM , Weeks , EM , Munson , G , Kane , M , Kang , HY , Cui , A , Ray , JP , Eisenhaure , TM , Collins , RL , Dey , K , Pfister , H , Price , AL , Epstein , CB , Kundaje , A , Xavier , RJ , Daly , MJ , Huang , HL , Finucane , HK , Hacohen , N , Lander , ES & Engreitz , JM 2021 , ' Genome-wide enhancer maps link risk variants to disease genes ' , Nature , vol. 593 , no. 7858 , pp. 238-+ . https://doi.org/10.1038/s41586-021-03446-x

http://hdl.handle.net/10138/345133
https://doi.org/10.1038/s41586-021-03446-x

publishedVersion

Article

Genome-wide enhancer maps link risk variants to disease genes

https://doi.org/10.1038/s41586-021-03446-x

Received: 20 August 2020

Accepted: 11 March 2021

Published online: 7 April 2021

Check for updates

Joseph Nasser[1,26], Drew T. Bergman[1,26], Charles P. Fulco[1,24,26], Philine Guckelberger[1,2,26], Benjamin R. Doughty[1,3,26], Tejal A. Patwardhan[1,4], Thouis R. Jones[1], Tung H. Nguyen[1], Jacob C. Ulirsch[1,5], Fritz Lekschas[6], Kristy Mualim[3], Heini M. Natri[3], Elle M. Weeks[1], Glen Munson[1], Michael Kane[1], Helen Y. Kang[3,7], Ang Cui[1,8], John P. Ray[1,25], Thomas M. Eisenhaure[1], Ryan L. Collins[1,9,10], Kushal Dey[11], Hanspeter Pfister[6], Alkes L. Price[1,11,12], Charles B. Epstein[1], Anshul Kundaje[3,13], Ramnik J. Xavier[1,14,15,16], Mark J. Daly[1,17,18,19], Hailiang Huang[1,17,18], Hilary K. Finucane[1,17,18], Nir Hacohen[1,18,20], Eric S. Lander[1,21,22,23,27] ✉ & Jesse M. Engreitz[1,3,7,27] ✉

Genome-wide association studies (GWAS) have identified thousands of noncoding loci that are associated with human diseases and complex traits, each of which could reveal insights into the mechanisms of disease[1]. Many of the underlying causal variants may affect enhancers[2,3], but we lack accurate maps of enhancers and their target genes to interpret such variants. We recently developed the activity-by-contact (ABC) model to predict which enhancers regulate which genes and validated the model using CRISPR perturbations in several cell types[4]. Here we apply this ABC model to create enhancer–gene maps in 131 human cell types and tissues, and use these maps to interpret the functions of GWAS variants. Across 72 diseases and complex traits, ABC links 5,036 GWAS signals to 2,249 unique genes, including a class of 577 genes that appear to influence multiple phenotypes through variants in enhancers that act in different cell types. In inflammatory bowel disease (IBD), causal variants are enriched in predicted enhancers by more than 20-fold in particular cell types such as dendritic cells, and ABC achieves higher precision than other regulatory methods at connecting noncoding variants to target genes. These variant-to-function maps reveal an enhancer that contains an IBD risk variant and that regulates the expression of *PPIF* to alter the membrane potential of mitochondria in macrophages. Our study reveals principles of genome regulation, identifies genes that affect IBD and provides a resource and generalizable strategy to connect risk variants of common diseases to their molecular and cellular functions.

Each GWAS association could provide insights into a biological mechanism that underlies the risk of disease in humans[1,5]. However, identifying these mechanisms has proved to be challenging. GWAS associations often include dozens of variants in linkage disequilibrium with one another that tag a single causal variant. Most causal variants do not directly alter protein-coding sequences but instead occur in noncoding gene regulatory elements such as enhancers[2,3], which can influence gene expression over long distances[6,7]. Furthermore, common diseases appear to involve contributions from multiple cell types, and many enhancers appear to act in specific cell types or states[8]. As such, connecting a GWAS association to function requires distinguishing among many possible variants, target genes and cell types[1,5].

Recent developments have set the stage for addressing these challenges. To distinguish among multiple possible variants at a locus, recent studies have applied statistical fine-mapping to prioritize likely causal variants for thousands of GWAS signals[9–11], including identifying 93 noncoding credible sets for IBD[9]. To link noncoding variants to their target genes and cell types, we recently developed the ABC model to identify enhancers in a particular cell type and predict their target genes based on data about chromatin state and three-dimensional contacts[4]. Together, these advances suggest an approach to connect GWAS signals to their target genes and cell types.

Here, we build ABC enhancer–gene maps for 131 biological samples (biosamples) and apply these maps to analyse fine-mapped genetic variants associated with 72 diseases and complex traits (Extended Data Fig. 1). These ABC maps link 5,036 GWAS signals to predicted genes and cell types, with improved accuracy compared to existing approaches. These predictions nominate previously undescribed regulatory mechanisms for IBD and identify genes that influence multiple diseases through effects in different cell types, including at the IBD risk locus at chromosome 10q22.3. Together, our study demonstrates a generalizable strategy to build regulatory maps of the genome to connect genetic associations to molecular mechanisms of disease.

A list of affiliations appears at the end of the paper.

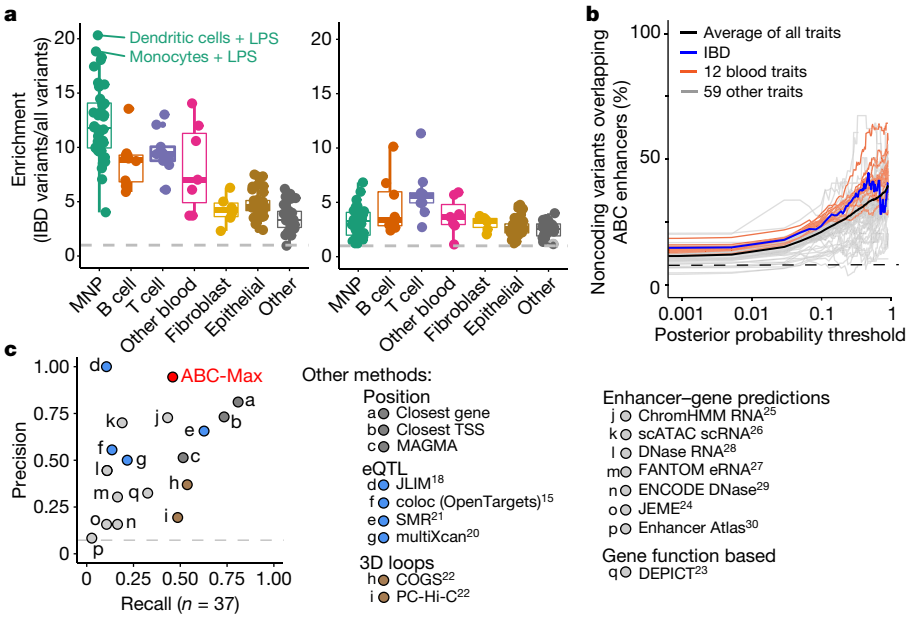


Fig. 1 | ABC maps connect fine-mapped variants to enhancers, genes and cell types. a, Enrichment of fine-mapped IBD variants (PIP ≥ 10%) in ABC enhancers (left) and all other accessible regions (right) in each of the 131 biosamples. MNP, mononuclear phagocytes. Box plots show the median (middle line) and interquartile range (boxes) and whiskers show observations less than or equal to quartiles ± 1.5× the interquartile range. **b**, Fraction of noncoding variants above a given PIP threshold that overlap an ABC enhancer in any biosample. Black line, weighted average across 72 traits. Traces are shown for PIP thresholds above which there are at least five variants. Dashed line, fraction of all common noncoding variants that overlap ABC enhancers. **c**, Precision–recall plot of connections between noncoding IBD credible sets and known IBD-associated genes[14], considering the 37 credible sets with 1 known gene within 1 Mb (Methods). Precision, fraction of identified genes corresponding to known genes; recall, fraction of the 37 known genes identified. For methods for which quantitative scores were available (for example, colocalization probability (Methods)), the plot shows the performance of choosing the gene with the best score per locus (Extended Data Fig. 6b). Data for eQTLs, 3D loops, and other enhancer–gene predictions were obtained from previous studies[15,18,20–30].

ABC enhancer–gene maps in 131 biosamples

We used the ABC model to construct genome-wide maps of enhancer–gene connections across 131 human biosamples, including 74 distinct primary cell types, tissues and cell lines from the ENCODE Project[8] and other sources (Extended Data Fig. 1 and Supplementary Tables 1, 2). For each biosample, we calculated ABC scores for each gene and chromatin accessible element within 5 megabases (Mb) by multiplying the estimates of enhancer activity and three-dimensional contact frequencies between enhancers and promoters. Candidate element–gene pairs that exceeded a chosen threshold were defined as 'enhancer–gene connections' and elements predicted to regulate at least one gene were defined as 'ABC enhancers' (Methods).

Across 131 biosamples, we identified 6,316,021 enhancer–gene connections for 23,219 expressed genes and 269,539 unique enhancers. In a given biosample, ABC identified an average of 48,441 enhancer–gene connections for 17,605 unique enhancers, comprising 2.9 Mb of enhancer sequence (around 12% of chromatin-accessible regions, which is 0.11% of the mappable genome) (Extended Data Fig. 2 and Supplementary Table 2). On average, each ABC enhancer was predicted to regulate 2.7 genes, each gene was predicted to be regulated by 2.8 ABC enhancers (Extended Data Fig. 2) and only 19% of enhancer–gene connections were shared between pairs of biosamples (Extended Data Fig. 3).

We validated these predictions by comparing them to a compendium of CRISPR perturbations that included 5,755 tested element–gene pairs in 11 cell types and states (including previous data[4,12] as well as additional CRISPR experiments that we performed here (Supplementary Tables 3, 4)). ABC performed well at classifying regulatory connections (area under the precision–recall curve = 0.64) and outperformed other methods, similar to our previous observations using a subset of the CRISPR data[4] (Extended Data Fig. 4 and Supplementary Table 5).

Enrichment of GWAS variants in enhancers

To assess the use of these maps in connecting disease variants to functions, we first quantified the enrichment of GWAS variants in ABC enhancers (Supplementary Table 6). Leveraging our previous fine-mapping analyses[9], we examined 24,922 fine-mapped variants with posterior inclusion probability (PIP) ≥ 10% for 72 diseases and traits, focusing on credible sets that did not contain any coding or splice site variants (Methods and Extended Data Fig. 5a).

Fine-mapped GWAS variants showed notable enrichments (up to 48-fold) in ABC enhancers in cell types relevant to each trait (Fig. 1a). These enrichments were stronger in ABC enhancers than in previously defined enhancer regions (Fig. 1a and Extended Data Fig. 5b–d), and in some cases showed evidence of allele-specific acetylation of histone 3 lysine 27 (H3K27ac) signals (Methods).

For example, fine-mapped variants for IBD were significantly enriched in ABC enhancers in 65 biosamples (Fisher's exact test with Bonferroni correction $P < 0.001$; 'enriched biosamples'), including 56 of the 66 biosamples that correspond to immune cell types, immune cell lines, or gut tissue (Fig. 1a and Supplementary Table 6). The most-enriched biosample showed a 21-fold enrichment and corresponded to activated dendritic cells, which are known to have an important role in the initiation of inflammation in IBD[13,14].

Across all signals for these 72 traits, ABC enhancers contained 40% of the 2,520 noncoding variants with PIP ≥ 95%, compared to 7.5% of all common noncoding variants (Fig. 1b and Extended Data Fig. 5e, f). For IBD and 12 blood cell traits, which have better coverage of relevant cell types in our dataset, ABC enhancers contained 46% of 732 noncoding variants with PIP ≥ 95% (Fig. 1c). Notably, this analysis probably underestimates the proportion of causal variants that reside in ABC enhancers because we still lack the appropriate data for many relevant cell types. We anticipate that most of the causal noncoding GWAS variants

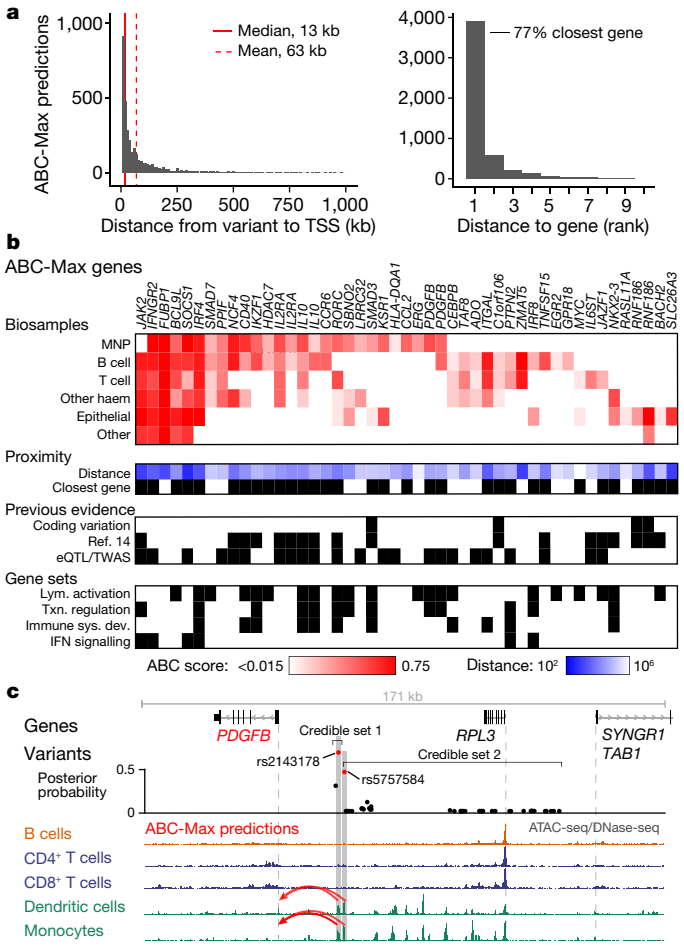


Fig. 2 | Connecting variants to target genes. a, Histogram on the left, distances from the predicted variant to the TSS of the ABC-Max target gene. Histogram on the right, distance rank of the gene in the locus. Data include predictions for all 72 traits. **b**, ABC-Max predictions for 47 noncoding IBD credible sets linking to 43 unique genes (4 genes are linked to 2 sets each). Immune sys. dev., immune system development; Lym, lymphocytes; Txn, transcription. Heat map, ABC scores in six biosample categories (maximum value within each category). Red scale, ABC score; blue scale, $\log_{10}$-transformed genomic distance from variant to gene TSS. Black boxes indicate that the gene is the closest to the lead single-nucleotide polymorphism (SNP), was implicated in IBD risk based on coding variation or experimental evidence about gene function[14], was identified by previous eQTL colocalization or transcriptome-wide association study (TWAS) analyses, or is in an enriched gene set (Methods). **c**, ABC-Max predictions and chromatin state at the *PDGFB* locus. Red colour denotes variants, enhancer–gene connections and target genes identified by ABC-Max. Grey bars, variants in two credible sets overlap ABC enhancers. Vertical dotted lines, TSSs.

will reside in ABC enhancers when ABC maps are expanded to include hundreds of additional cell types (Extended Data Fig. 5e).

Evaluating gene predictions

We next used ABC to connect noncoding GWAS signals to target genes. For each trait, we intersected fine-mapped variants (PIP ≥ 10%) with ABC enhancers in enriched biosamples, and assigned each credible set to the target gene with the highest ABC score (ABC-Max) (Supplementary Note 1).

For example, the IBD risk locus at chromosome 1q32.1 had been previously fine-mapped to identify two independent credible sets[9] (Extended Data Fig. 1b). Both credible sets include noncoding variants

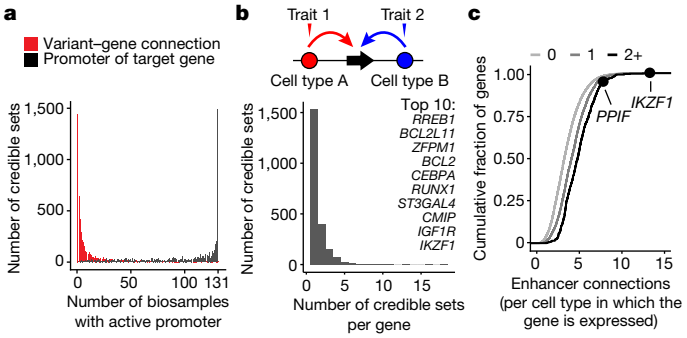


Fig. 3 | Cell-type specificity of ABC predictions. a, Histogram of the number of biosamples in which (red) a variant–gene connection is predicted by ABC-Max (that is, an ABC enhancer regulates the target gene in a given biosample) and (grey) the promoter of the targeted gene is active (Methods). **b**, Histogram of the number of GWAS signals per gene (unique credible sets with no overlapping variants with PIP ≥ 10%, Methods). Model at the top depicts a gene linked to different traits via different variants. Circles. enhancers; black arrows, gene; coloured arrows, ABC predictions; triangles, variants. **c**, Number of predicted enhancer–gene connections (per biosample in which the promoter of a gene is active), for genes linked by ABC-Max to zero traits, one trait by one or more variants, or two or more traits via different variants. Labels, two genes described in text.

with PIP ≥ 10% that overlap ABC enhancers in monocytes stimulated with bacterial lipopolysaccharide (LPS), the biosample with the second highest enrichment for IBD (Fig. 1a). For both credible sets, ABC-Max predicted that these enhancers regulate multiple genes in the locus, but the gene with the highest ABC score was *IL10*, a key anti-inflammatory cytokine that is known to be important for IBD[14] (Extended Data Fig. 6a).

To evaluate ABC-Max and other previous predictions, we examined a set of genes previously linked to IBD based on coding variants or evidence from experimental models[14] (Supplementary Tables 8, 9). We analysed the 37 noncoding credible sets within 1 Mb of one of these genes, and tested how often ABC-Max or other methods prioritized the known gene above all other genes in the locus (median genes per locus, 15; range, 4–67). We visualized performance using a precision–recall plot, where recall is the fraction of credible sets for which the known gene is identified (sensitivity) and precision is the fraction of predicted genes corresponding to known genes (positive predictive value) (Fig. 1c).

As a baseline, we tested the heuristic of assigning each GWAS credible set to the closest gene—a method that is widely used to annotate GWAS loci[15,16] and has been shown to assign approximately 70% of metabolite GWAS loci to genes with plausible biochemical functions[17]. Connecting the lead variant to the closest gene correctly identified the known IBD-associated gene for 30 out of 37 credible sets (81% precision and 81% recall) (Fig. 1c). A similar approach—which selects the closest transcription start site (TSS)—identified the known IBD-associated gene in 27 out of 37 cases (73% precision and 73% recall) (Fig. 1c, Supplementary Note 2).

We next evaluated other approaches to connect regulatory variants to disease-associated genes, including predictions based on signals from expression quantitative trait loci (eQTLs)[18–21], three-dimensional contacts[22], gene set enrichment[23] or other enhancer–gene maps[24–31] (Methods). Most of these approaches achieved lower precision and recall than closest gene (Fig. 1c).

Finally, we evaluated ABC-Max. Of the 37 credible sets, 18 included a variant that overlapped an ABC enhancer in an enriched biosample, and ABC-Max identified the known gene in 17 out of 18 cases (94% precision and 49% recall) (Fig. 1c). Thus, ABC-Max identifies a high-confidence set of genes at these IBD GWAS loci, with higher precision than other enhancer maps. Although ABC-Max had lower recall than the closest gene, the

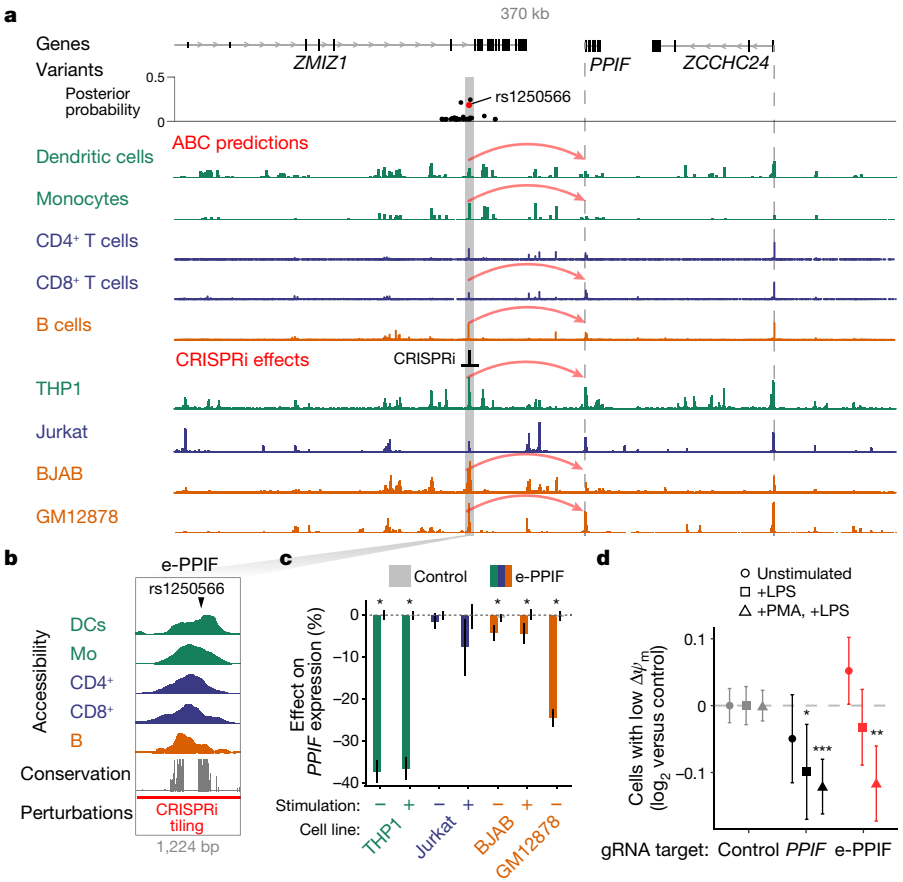


Fig. 4 | An enhancer regulates *PPIF* expression and mitochondrial function. **a**, An IBD risk variant (rs1250566) overlaps with an enhancer that is predicted to regulate *PPIF*. Signal tracks represent chromatin accessibility (from ATAC-seq (assay for transposase-accessible chromatin using sequencing) or DNase-seq (DNase I hypersensitive site sequencing)). Grey bar, enhancer containing rs1250566; dashed lines, TSSs; red arrows for dendritic cells, monocytes, $CD4^+$ and $CD8^+$ T cells and B cells, ABC-Max predictions; red arrows for THP1, Jurkat, BJAB and GM12878 cells, CRISPRi leads to a significant decrease in *PPIF* expression. **b**, The 1,224-bp region at the *PPIF* enhancer (e-PPIF). Accessibility, DNase-seq or ATAC-seq of primary immune cells. DCs, dendritic cells; Mo, monocytes. Conservation, phastCons 100-mammal alignment. Red bar, region targeted with CRISPRi gRNAs. **c**, Effects of CRISPRi at e-PPIF on the expression of *PPIF* in immune cell lines in resting and stimulated conditions. Data are mean ± 95% confidence intervals of the mean. Two-sided Student's *t*-test with Benjamini-Hochberg correction (*$P < 0.05$) for 164 CRISPRi gRNAs targeting e-PPIF compared with 814 negative control gRNAs. Adjusted *P* values from left to right: 4.68×10^{-101}, 4.86×10^{-112}, 0.019, 0.044 and 1.48×10^{-71}. **d**, Effects of CRISPRi gRNAs (targeting e-PPIF, *PPIF* promoter or negative controls) on $\Delta\psi_m$, quantified as the frequency of THP1 cells carrying those gRNAs that had a low compared with a high MitoTracker Red signal (Extended Data Fig. 10f–h). The $\log_2$-transformed fraction of cells (treatment over unstimulated control) is shown. We tested THP1 cells in unstimulated conditions, stimulated with LPS, and differentiated with phorbol 12-myristate 13-acetate (PMA) and stimulated with LPS (Methods). Data are mean ± 95% confidence intervals for the mean of 40, 9 and 5 gRNAs for control, *PPIF* and e-PPIF, respectively. Two-sided Wilcoxon rank-sum test; *$P = 0.0163$, **$P = 0.00426$, ***$P = 0.000356$ versus the control.

fraction of loci with a prediction will probably increase upon expanding the ABC maps to include additional relevant cell types in the gut.

Because this curated gene set may have certain biases, we conducted additional analyses to benchmark ABC-Max for IBD and other traits (Supplementary Note 2). We found that ABC-Max selected genes at IBD loci that showed stronger gene set enrichments compared to other approaches (Extended Data Fig. 6b), often selected the gene with the closest TSS (Extended Data Fig. 6c) and strongly enriched for identifying high-confidence genes for an independent set of 10 quantitative blood traits (17-fold enrichment) (Extended Data Fig. 6d). Together, these analyses demonstrate that ABC maps can accurately connect fine-mapped variants to target genes for IBD and other complex traits.

We made several observations that help to explain the good performance of ABC-Max (Supplementary Note 2). Notably, assigning each credible set to the gene with the strongest ABC score ('ABC-Max'; precision = 94% for known IBD-associated genes) performed far better than assigning each credible set to all genes linked to an IBD variant ('ABC-All'; precision = 17%) (Extended Data Fig. 6e). This was because individual variants often overlapped ABC enhancers that were predicted to regulate multiple genes (median, 3; range, 1–17), with the known gene having the highest ABC score (Extended Data Fig. 6a). Choosing the gene with the highest score was also important for optimal performance of other prediction methods, such as those based on eQTLs (Extended Data Fig. 6e). This complexity appears to be a fundamental feature of mammalian gene regulation: *cis*-eQTL studies indicate that noncoding variants often regulate multiple genes[32], and CRISPR experiments have identified individual enhancers that regulate up to eight genes in *cis*[4,33]. Our observations are consistent with a model where—although variants often affect the expression of multiple genes—only a subset of these effects are likely to be relevant to disease[34] (Supplementary Note 1).

Regulatory mechanisms at GWAS loci

Having demonstrated that ABC identifies cell types and genes relevant to specific phenotypes, we next applied ABC-Max to GWAS signals for 72 diseases and traits. ABC-Max made a prediction for 5,036 noncoding credible sets, nominating a total of 4,976 fine-mapped variants that overlapped enhancers linked to 2,249 unique genes (Supplementary

Table 10). The distance from the noncoding variant in the ABC enhancer to the TSS of the predicted target gene ranged from less than 1 kilobase (kb) to 1.1 Mb (median, 13 kb), and 1,139 out of 5,036 predictions (23%) involved a gene that was not the closest (Fig. 2a).

These predictions provide a resource for identifying genes, pathways and regulatory properties of GWAS loci. For example, ABC-Max made predictions for 47 noncoding IBD credible sets, nominating 43 unique genes (4 genes were linked to 2 independent signals in the same locus) (Fig. 2b and Supplementary Tables 10, 11). Many of these genes have previously been reported to have functions in immunity and inflammation and the predicted genes were enriched for genes in the interferon-γ pathway (6 genes; 12-fold enrichment), lymphocyte activation (11 genes; 7-fold enrichment) and regulation of transcription from the promoter of RNA polymerase II (21 genes; 5-fold enrichment) (Fig. 2b). ABC-Max also identified genes that were not the closest or previously annotated gene, such as at the IBD locus at chromosome 22q13, which has been annotated as corresponding to *TAB1* (also known as *MAP3K7IP1*)[35,36]. Here, ABC-Max linked variants in two independent credible sets to platelet-derived growth factor-β (*PDGFB*) in mononuclear phagocytes (for example, monocytes, macrophages and dendritic cells), supporting a causal role for PDGF signalling in IBD[37] (Fig. 2c). We also identified intergenic IBD risk variants linked to *LRRC32* and *RASL11A* (Extended Data Fig. 7 and Supplementary Note 3), and variants located in the introns of *ANKRD55* and *ZMIZ1* were linked to different nearby genes (see below).

Cell-type-specific links to disease

Identifying the cell type in which a gene influences disease can provide additional insights into disease aetiology. We characterized the cell-type specificity of ABC predictions, and found that ABC enhancers containing fine-mapped variants were active in a median of only 4 biosamples, compared to 120 biosamples for the promoters of their target genes (Fig. 3a).

For IBD, the cell-type specificity of ABC-Max predictions identified cases for which a variant was predicted to act only in specific cell lineages or stimulated immune cell states (Extended Data Fig. 8a, b) and enabled the grouping of genes by cell type to improve the detection of enriched gene sets (Extended Data Fig. 8c). At one IBD locus (chromosome 5q11.2), we identified a single fine-mapped IBD risk variant (rs7731626, PIP = 28%) that overlapped an ABC enhancer and was linked to *IL6ST* only in T cell subsets and fetal thymus tissue, even though *IL6ST* is expressed in most cell types. Using CRISPR interference (CRISPRi), we confirmed that this predicted enhancer regulates *IL6ST* in a T cell line but not in three other B cell or monocytic cell lines (Extended Data Fig. 8d).

Such cell-type-specific effects appeared to lead to cases in which a single gene could affect multiple traits. For example, *IKZF1* encodes a transcription factor that is involved in several stages of haematopoietic differentiation, and this gene was linked by ABC to IBD and 11 other traits through different variants in 18 credible sets, including variants associated with erythrocyte, monocyte or neutrophil counts that overlapped ABC enhancers in erythroblasts, monocytes or $CD34^+$ haematopoietic progenitors, respectively (Extended Data Fig. 9a).

In total, we identified 577 genes that were each linked by ABC-Max to different traits through different variants (Fig. 3b and Supplementary Table 12), and for which the predicted variants overlapped ABC enhancers in different sets of biosamples. These 577 genes appeared to have complex enhancer landscapes: they had (1) more predicted ABC enhancer connections (median of 466 across all cell types versus 261 for other genes); (2) more ABC enhancer connections per cell type in which the gene was expressed (median of 4.8 versus 3.3); and (3) more surrounding noncoding sequence (median of 301 kb versus 128 kb distance to the closest neighbouring TSSs, independent of ABC predictions) (Fig. 3c and Extended Data Fig. 9b, c). These observations suggest that genes with complex enhancer landscapes are more likely to influence multiple traits, which may reflect constraints on their precise cell-type-specific transcriptional control[38].

From association to function at 10q22.3

To explore how ABC maps could accelerate experimental studies to characterize individual GWAS loci, we examined the IBD risk locus at chromosome 10q22.3, for which ABC prioritized an unexpected gene. A single high-probability variant (rs1250566, PIP = 19%), which was located in an intron of *ZMIZ1*, overlapped with an ABC enhancer in several immune cell types, including mononuclear phagocytes (Fig. 4a, b). Although this locus has previously been annotated as corresponding to *ZMIZ1*[15,35,39], ABC-Max linked this variant to a different nearby gene, *PPIF*. *PPIF* has a higher ABC score than *ZMIZ1* because the variant is in more frequent three-dimensional contact with the promoter of *PPIF* than with the promoter of *ZMIZ1* (by a factor of 2.3).

To obtain evidence that variation in the predicted *PPIF* enhancer could affect the risk of IBD, we used CRISPRi-FlowFISH (CRISPRi combined with fluorescence in situ hybridization and flow sorting)[4] to perturb each of the 163 accessible elements in a 712-kb region around *PPIF* in four human immune cell lines—THP1, BJAB, GM12878 and Jurkat cells—with and without stimulation with the appropriate immune ligands. We identified 14 enhancers that regulated *PPIF* expression in at least one of these conditions (Extended Data Fig. 10a, b and Supplementary Table 4). Only one of these 14 enhancers contained a fine-mapped IBD variant (the enhancer that was initially predicted by ABC-Max), and this enhancer had a particularly strong effect on *PPIF* expression (up to 43% effect in THP1 cells in unstimulated and LPS-stimulated conditions, two-sided Student's *t*-test, $P < 10^{-111}$) (Fig. 4c and Extended Data Fig. 10b–e).

PPIF encodes cyclophilin D, a ubiquitously expressed protein that regulates metabolism, reactive oxygen species signalling and cell death by controlling the mitochondrial permeability transition and mitochondrial membrane potential ($\Delta\psi_m$)[40]. Accordingly, we tested whether the *PPIF* enhancer containing the IBD variant could tune $\Delta\psi_m$ in THP1 cells. We infected cells with a pool of CRISPRi guide RNAs (gRNAs) that target the *PPIF* enhancer and promoter, stained cells with MitoTracker Red (a fluorescent dye for which the signal increases with $\Delta\psi_m$), sorted cells into three bins based on their level of fluorescence, and sequenced the gRNAs in each bin to infer their effects on $\Delta\psi_m$ (Extended Data Fig. 10f). CRISPRi targeting of the *PPIF* enhancer or promoter indeed increased $\Delta\psi_m$ in THP1 cells in LPS-stimulated, but not unstimulated, conditions (Fig. 4d and Extended Data Fig. 10g, h), consistent with the expected direction of effect of *PPIF*. These experiments indicate that this enhancer can tune the metabolic state of mitochondria in cells that respond to inflammatory stimuli. Notably, changes in $\Delta\psi_m$ have been previously linked to inflammatory responses in macrophages[41], suggesting a path by which tuning *PPIF* expression could affect IBD risk.

Notably, *PPIF* has an extremely complex enhancer landscape (Fig. 3c) (top 0.3% of genes with the most ABC enhancer connections) and the *PPIF* locus also contains GWAS signals for 39 other diseases and traits in addition to IBD (Extended Data Fig. 10a). By comparing these variants to our CRISPRi data, we found a distinct enhancer that regulated *PPIF* only in GM12878 lymphoblastoid cells and contained a variant associated with lymphocyte count and multiple sclerosis (Extended Data Fig. 10b–d). Together, these observations suggest that cell-type-specific transcriptional regulation of *PPIF* may influence risk for multiple complex diseases and traits (Supplementary Note 4).

Discussion

This research created genome-wide maps of more than six million enhancer–gene connections that illuminate the functions of disease variants. These maps revealed new genes and pathways for IBD, nominated 577 genes that control different traits through effects in different cell types and identified a role for an enhancer of PPIF in tuning mitochondrial function in macrophages. We have also prospectively applied ABC maps to identify a variant that regulates *TET2* in haematopoietic progenitors to influence risk for clonal haematopoiesis[42]. By

markedly narrowing the search space of possible variants, cell types and target genes at any given GWAS locus, ABC maps should accelerate variant-to-function studies for many diseases. To facilitate such studies, these maps are available at https://www.engreitzlab.org/abc/.

Our study has several limitations that highlight areas for future work (Supplementary Note 5). First, ABC does not perfectly predict the effects of distal enhancers and does not capture other types of regulatory elements. Second, ABC-Max assumes a single causal gene per variant, although enhancers that contain disease variants often appear to have highly pleiotropic effects. Third, most of these ABC maps involve analysis of data from a single individual and therefore miss enhancers that are present only in certain genotypes or environments. Fourth, assessing the performance of gene predictions requires good sets of gold-standard genes, which remain limited and may contain biases (for example, towards the closest gene or towards genes that tolerate coding variation). Expanding the model beyond the current assumptions and performing more systematic experimental studies will be required to address these limitations.

In summary, our approach highlights a path to creating a comprehensive map of enhancer regulation in the human genome. By refining computational models such as ABC and collecting the necessary epigenomic data, it should be possible to create an accurate map of enhancers and their target genes in *cis* across thousands of cell types and states in the human body. These maps could then be used to link noncoding variants to disease-associated genes and cell types. Such a project is becoming feasible, and will be an essential resource for understanding gene regulation and the genetic basis of human diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03446-x.

1. Claussnitzer, M. et al. A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).
2. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
3. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
4. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet*. **51**, 1664–1669 (2019).
5. Westra, H.-J. & Franke, L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta* **1842**, 1896–1902 (2014).
6. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet*. **21**, 292–310 (2020).
7. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer–promoter interaction specificity. *Trends Cell Biol*. **24**, 695–702 (2014).
8. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
9. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
10. Maller, J. B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet*. **44**, 1294–1301 (2012).
11. Ulirsch, J. C. et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet*. **51**, 683–693 (2019).
12. Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
13. Rescigno, M. & Di Sabatino, A. Dendritic cells in intestinal homeostasis and disease. *J. Clin. Invest*. **119**, 2441–2450 (2009).
14. Graham, D. B. & Xavier, R. J. Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* **578**, 527–539 (2020).
15. Mountjoy, E. et al. Open Targets Genetics: an open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Preprint at https://doi.org/10.1101/2020.09.16.299271 (2020).
16. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. **47**, D1005–D1012 (2019).
17. Stacey, D. et al. ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res*. **47**, e3 (2019).
18. Chun, S. et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet*. **49**, 600–605 (2017).
19. Carvalho-Silva, D. et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res*. **47**, D1056–D1065 (2019).
20. Barbeira, A. N. et al. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet*. **15**, e1007889 (2019).
21. Hauberg, M. E. et al. Large-scale identification of common trait and disease variants affecting gene expression. *Am. J. Hum. Genet*. **100**, 885–894 (2017).
22. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
23. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun*. **6**, 5890 (2015).
24. Cao, Q. et al. Reconstruction of enhancer–target networks in 935 samples of human primary cells, tissues and cell lines. *Nat. Genet*. **49**, 1428–1436 (2017).
25. Liu, Y., Sarkar, A., Kheradpour, P., Ernst, J. & Kellis, M. Evidence of reduced recombination rate in human regulatory domains. *Genome Biol*. **18**, 193 (2017).
26. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol*. **37**, 1458–1465 (2019).
27. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
28. Sheffield, N. C. et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. **23**, 777–788 (2013).
29. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
30. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res*. **48**, D58–D64 (2020).
31. Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet*. **48**, 488–496 (2016).
32. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
33. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
34. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet*. **51**, 592–599 (2019).
35. 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).
36. Jostins, L. et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
37. Linares, P. M. & Gisbert, J. P. Role of growth factors in the development of lymphangiogenesis driven by inflammatory bowel disease: a review. *Inflamm. Bowel Dis*. **17**, 1814–1821 (2011).
38. Wang, X. & Goldstein, D. B. Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. *Am. J. Hum. Genet*. **106**, 215–233 (2020).
39. Imielinski, M. et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat. Genet*. **41**, 1335–1340 (2009).
40. Elrod, J. W. & Molkentin, J. D. Physiologic functions of cyclophilin D and the mitochondrial permeability transition pore. *Circ. J*. **77**, 1111–1122 (2013).
41. Ip, W. K. E., Hoshi, N., Shouval, D. S., Snapper, S. & Medzhitov, R. Anti-inflammatory effect of IL-10 mediated by metabolic reprogramming of macrophages. *Science* **356**, 513–519 (2017).
42. Bick, A. G. et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany. [3]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [4]Department of Statistics, Harvard University, Cambridge, MA, USA. [5]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA. [6]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [7]BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford University School of Medicine, Stanford, CA, USA. [8]Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA, USA. [9]Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA. [10]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [11]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [12]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [13]Department of Computer Science, Stanford University, Stanford, CA, USA. [14]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [15]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA. [16]Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [17]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [18]Department of Medicine, Harvard Medical School, Boston, MA, USA. [19]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [20]Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA. [21]Department of Biology, MIT, Cambridge, MA, USA. [22]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [23]Office of Science and Technology Policy, Executive Office of the President, White House, Washington, DC, USA. [24]Present address: Bristol Myers Squibb, Cambridge, MA, USA. [25]Present address: Systems Immunology, Benaroya Research Institute at Virginia Mason, Seattle, WA, USA. [26]These authors contributed equally: Joseph Nasser, Drew T. Bergman, Charles P. Fulco, Philine Guckelberger, Benjamin R. Doughty. [27]These authors jointly supervised this work: Eric S. Lander, Jesse M. Engreitz. ✉e-mail: lander@broadinstitute.org; engreitz@stanford.edu

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Immune cell lines

We generated epigenomic data to build the ABC model and/or performed CRISPRi experiments in the following human immune cell lines: THP1 (monocytic-like cell line, acute monocytic leukaemia), BJAB (B-cell-like cell line, EBV-negative inguinal Burkitt's lymphoma), GM12878 (EBV-immortalized lymphoblastoid cell line), U937 (monocytic-like cell line, histiocytic lymphoma) and Jurkat (T-cell-like cell line, T cell leukaemia).

Cell culture. We maintained cells at a density between 100,000 and 1,000,000 cells per ml (250,000–1,000,000 per ml for GM12878) in RPMI-1640 (Thermo Fisher Scientific) with 10% heat-inactivated FBS (15% for GM12878; HIFBS, Thermo Fisher Scientific), 2 mM L-glutamine and 100 units per ml streptomycin and 100 mg ml^{-1} penicillin by diluting cells 1:8 in fresh medium every 3 days. Cell lines were regularly tested for mycoplasma, and authenticated through comparison of epigenomic data to published datasets.

Stimulation conditions for ABC maps and CRISPRi experiments. We stimulated BJAB cells with 4 μg ml^{-1} anti-CD40 (Invitrogen, 140409-82) and 10 μg ml^{-1} anti-IgM (Sigma-I0759) for 4 h. We stimulated Jurkat cells with 5 μg ml^{-1} anti-CD3 (Biolegend-317315) and 100 ng ml^{-1} phorbol 12-myristate 13-acetate (PMA, Sigma-P1585) for 4 h. We stimulated THP1 cells with 1 μg ml^{-1} bacterial LPS from *Escherichia coli* K12 (Invivogen, tlrl-peklps) for 4 h. We stimulated U937 cells with 200 ng ml^{-1} LPS for 4 h.

Stimulation conditions for ABC maps across an extended time course in THP1 cells. For THP1 cells, we generated epigenomic data examining a longer time course, by stimulating with PMA (100 ng ml^{-1}) for 12 h, then removing PMA and adding LPS (1 μg ml^{-1}) and profiling at 0, 1, 2, 6, 12, 24, 48, 72, 96 and 120 h after addition of LPS. Because THP1 cells adhere when stimulated with PMA (changing into a more macrophage-like state), we collected the cells by taking out the medium, washing twice, adding TrypLE for 5 min at 37 °C, then supplementing with 100 μL of medium, removing cells from the round-bottom plate and pelleting. These data were used to generate the ABC predictions included in the 131 biosamples.

Epigenomic profiling of immune cell lines

To build ABC maps in human immune cell lines, we generated data using ATAC-seq and chromatin immunoprecipitation sequencing (ChIP–seq) of H3K27ac in BJAB, Jurkat, THP1 and U937 cells, with and without stimulation with the ligands described above.

ATAC-seq. We applied ATAC-seq as previously described[43], with modifications. In brief, we washed 50,000 cells once with 50 μl of cold 1× PBS and added 50 μl of Nuclei Isolation EZ Lysis buffer (Sigma, NUC101-1KT) to resuspend gently, immediately centrifuging at 500*g* for 10 min at 4 °C. The lysis buffer was decanted away from the nuclei pellet. Afterwards, we resuspended the nuclei in 100 μl of Nuclei Isolation EZ Lysis buffer again and centrifuged at 500*g* for 5 min at 4 °C and re-decanted the lysis buffer, which we found to decrease mitochondrial reads although at the cost of library complexity. We then resuspended the nuclear pellet in 50 μl of transposition reaction mix (25 μl buffer TD, 2.5 μl TDE1 (Illumina 15028212); 7.5 μl water, 15 μl PBS, to increase salinity, which we found to increase the signal-to-noise ratio) and incubated the mix at 37 °C for 30 min in a PCR block. Immediately after the transposition reaction, we split the 50 μl reaction volume into two and we added 25 μl of guanidine hydrochloride (buffer PB, Qiagen, 28606) to each as a chaotropic agent to stop the reaction and dissociate the proteins and transposase from the DNA. Keeping one of the reactions as a backup, we proceeded with one by adding 1.8× SPRI beads (Agencourt A63881), waiting 5 min for the DNA to associate to the beads, and then washing the beads twice using 80% ethanol. We then eluted the DNA from the beads using 10 μl of water and added to it 25 μl NEBNext HiFi 2× PCR MasterMix (NEB M0541), with 2.5 μl of each of the dual-indexed Illumina Nextera primers (25 μM). We amplified the PCR reaction for 15 cycles, as previously described[33]. We purified amplified libraries and removed adapters using two clean-ups with 1.8× volume SPRI (Agencourt, A63881). We sequenced these libraries on an Illumina HiSeq 2500. We filtered, aligned and processed the data to generate BAM files as previously described[33].

H3K27ac ChIP–seq. We generated and analysed ChIP–seq data from 5 million cells in each cell line and stimulation state, following previously described protocols[44]. Before collecting the cells for ChIP–seq, the cells (1 million cells per ml) were replenished by a 1:2 (v/v) split in fresh medium and allowed to grow for 4 h. Then, 10 million cells were collected from each cell type at 500,000 cells per ml and washed twice in cold PBS. Cells were resuspended in warm PBS with 1% formaldehyde (28906, Thermo Scientific) and incubated at 37 °C for 10 min. Crosslinking was quenched by adding glycine to a concentration of 250 mM and incubating for 5 min at 37 °C. Cells were placed on ice for 5 min, then washed twice in ice-cold PBS and snap-frozen in liquid nitrogen and stored. Later, crosslinked cells were lysed in 1 ml cell lysis buffer (20 mM Tris pH 8.0, 85 mM KCl, 0.5% NP-40) and incubated on ice for 10 min. The nuclear pellet was isolated by spinning the cell lysis mix at 5,600*g* at 4 °C for 3.5 min and discarding the supernatant. Nuclear pellets were lysed by adding 1 ml nuclear lysis buffer (10 mM Tris-HCl pH 7.5, 1% NP-40 alternative (CAS 9016-45-9), 0.5% sodium deoxycholate, 0.1% SDS) with protease inhibitors on ice for 10 min. The chromatin-containing nuclear lysate was sonicated 3× using a Branson sonifier (on 0.7 s, off 1.3 s, time 2 min, watts 10–12), with 1 min rest between sonifications. Sonicated chromatin was spun down at maximum speed. Then, 300 μl of the clarified supernatant was diluted 1:1 with ChIP dilution buffer (16.7 mM Tris-HCl pH 8.1, 1.1% Triton X-100, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS). To immunoprecipitate H3K27ac, 3 μl of H3K27ac monoclonal antibody (39685, Active Motif) was added to each sample and rotated overnight at 4 °C. The following morning, 50 μl of a 1:1 mix of protein-A (10008D, Invitrogen) and protein-G Dynabeads magnetic beads (10004D, Life Technologies) were washed with blocking buffer (PBS, 0.5% Tween-20, 0.5% BSA with protease inhibitors), resuspended in 100 μl blocking buffer and added to each sample. The samples were rotated end-over-end for 1 h at 4 °C to capture antibody complexes, then washed as follows: once with 200 μl Low-Salt RIPA buffer (0.1% SDS, 1% Triton X-100, 1 mM EDTA, 20 mM Tris-HCl pH 8.1, 140 mM NaCl, 0.1% sodium deoxycholate), once with 200 μl High-Salt RIPA buffer (0.1% SDS, 1% Triton X-100, 1 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl, 0.1% sodium deoxycholate), twice with 200 μl LiCl buffer (250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) and twice with 200 μl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Chromatin was then eluted from the beads with 60 μl ChIP elution buffer (10 mM Tris-HCl pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.1% SDS). Crosslinking was reversed by adding 8 μl of reverse crosslinking enzyme mix (250 mM Tris-HCl pH 6.5, 62.5 mM EDTA pH 8.0, 1.25 M NaCl, 5 mg ml^{-1} proteinase K (25530-049, Invitrogen), 62.5 μg ml^{-1} RNase A (111199150001, Roche)) to each immunoprecipitated sample, as well as to 10 μl of the sheared chromatin input for each sample brought to volume of 60 μl ChIP elution buffer. Reverse crosslinking reactions were incubated 2 h at 65 °C and cleaned using Agencourt Ampure XP SPRI beads (A63880, Beckman Coulter) with a 2× bead:sample ratio. Sequencing libraries were prepared with the KAPA Library Preparation kit (KK8202, KAPA Biosystems). ChIP libraries were sequenced using

single-end sequencing on an Illumina Hiseq 2500 machine (read 1, 76 cycles; index 1, 8 cycles), to a depth of more than 30 million reads per ChIP sample.

Curation of published epigenomic data

Supplementary Table 2 lists the data sources for each ABC biosample, and Supplementary Table 1 describes the epigenomic datasets generated for this study.

ENCODE. We downloaded BAM files for DNase-seq and H3K27ac ChIP–seq experiments from the ENCODE Portal on 17 July 2017[8]. We selected the hg19-aligned BAM files that were marked as 'released' by the ENCODE Portal and were not flagged as 'unfiltered', 'extremely low spot score', 'extremely low read depth', 'NOT COMPLIANT' or 'insufficient read depth'.

Roadmap. We downloaded BAM files for DNase-seq and H3K27ac ChIP–seq from the Roadmap Epigenomics Project (http://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/) on 12 July 2017[45].

Other studies. We downloaded FASTQ files for DNase-seq, ATAC-seq and ChIP–seq data from 13 other studies (Supplementary Table 2), and processed them using our custom pipelines as described below.

Merging cell types. We created a list of cell types across all sources for which we had at least one chromatin accessibility experiment (DNase-seq or ATAC-seq) and one H3K27ac ChIP–seq experiment. In cases in which the same cell types were included in data from the Roadmap Epigenome Project and also from the ENCODE Portal, we used the processed data from Roadmap. In some cases, we combined data from multiple sources (for example, ENCODE data and our own datasets) to expand the number of cell types considered. As a result of this merging, some 'cell types' in our dataset represent data from a single donor and experimental sample, whereas others involve a mixture of multiple donors and/or experimental samples.

Processing of ATAC-seq and ChIP–seq data

We aligned reads using BWA (v.0.7.17)[46], removed PCR duplicates using the MarkDuplicates function from Picard (v.1.731, http://picard.sourceforge.net) and filtered to uniquely aligning reads using samtools (MAPQ ≥ 30, https://github.com/samtools/samtools)[47]. The resulting BAM files were used as inputs into the ABC model.

ABC model predictions

We used the ABC model (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction) to predict enhancer–gene connections in each cell type, based on measurements of chromatin accessibility (ATAC-seq or DNase-seq), histone modifications (H3K27ac ChIP–seq), and chromatin conformation (Hi-C) as previously described[4]. In a given cell type, the ABC model reports an ABC score for each element–gene pair, where the element is within 5 Mb of the TSS of the gene. We previously found that the exact window used does not significantly affect performance; here, we used 5 Mb to maintain consistency with our previous study[4].

In brief, for each cell type, we first called peaks in the chromatin accessibility dataset using MACS2 with a lenient *P*-value cut-off of 0.1.

Second, we counted chromatin accessibility reads in each peak and retained the top 150,000 peaks with the most read counts. We then resized each of these peaks to be 500 bp centred on the peak summit. To this list we added 500 bp regions centred on all gene TSS's and removed any peaks overlapping blacklisted regions (version 1 from https://sites.google.com/site/anshulkundaje/projects/blacklists)[8,48]. Any resulting overlapping peaks were merged. We call the resulting set of regions candidate elements.

Third, we calculated element activity by first counting reads in each candidate element in chromatin accessibility and H3K27ac ChIP–seq experiments, and then taking the geometric mean of the two assays. Chromatin accessibility and H3K27ac ChIP–seq signals in each candidate element were quantile-normalized to the distribution observed in K562 cells.

Fourth, we calculated element–promoter contact using the average Hi-C signal across 10 human Hi-C datasets as described below.

Finally, we computed the ABC score for each element–gene pair as the product of activity and contact, normalized by the product of activity and contact for all other elements within 5 Mb of that gene.

Average Hi-C. To generate a genome-wide averaged Hi-C dataset, we downloaded Knight–Ruiz normalized Hi-C matrices for 10 human cell types[4]. For each cell type, we first transformed the Hi-C matrix for each chromosome to be doubly stochastic. We then we replaced the entries on the diagonal of the Hi-C matrix with the maximum of its four neighbouring bins. Next, we replaced all entries of the Hi-C matrix with a value of NaN or corresponding to Knight–Ruiz normalization factors <0.25 with the expected contact under the power-law distribution in the cell type. We subsequently scaled the Hi-C signal for each cell type using the power-law distribution in that cell type as previously described. Finally, we computed the 'average' Hi-C matrix as the arithmetic mean of the 10 cell-type-specific Hi-C matrices. This Hi-C matrix (5-kb resolution) is available at ftp://ftp.broadinstitute.org/outgoing/lincRNA/average_hic/average_hic.v2.191020.tar.gz.

The averaged Hi-C contacts correlate well with cell-type-specific Hi-C contacts (for example, $R^2 = 0.91$ for K562 cells) (Supplementary Fig. 1). We have previously shown that the ABC score is able to make accurate cell-type-specific enhancer–gene predictions using this averaged Hi-C dataset and outperforms other approaches that use loops or distance instead of quantitative contact frequency[4]. We also find here that using averaged Hi-C data performs similarly to using cell-type-specific promoter capture Hi-C (PC-HiC) data (Extended Data Fig. 4e).

PC-Hi-C. In some evaluations of the performance of the ABC model to CRISPR data (Extended Data Fig. 4e–h), we used ABC predictions for which the contact component of the ABC score is derived from the raw counts in PC-HiC experiments. The PC-HiC data was processed as follows. First we downloaded PC-HiC raw count data from the BLUEPRINT consortium[22]. Second, contacts from restriction fragments that overlap baited promoter regions were linearly adjusted based on the total number of detected contacts for the baited region(s). Third, we re-binned the data from restriction fragment sites to 5-kb resolution. Fourth, to fill in missing values for very short-range contacts, we imputed contact data between the baited restriction fragment and itself using the power-law distribution.

The contact for an enhancer–gene pair is assigned as the counts observed in the PC-HiC experiment corresponding to the baited fragment overlapping the gene promoter and the 5-kb bin overlapping the element.

Estimating promoter activity. In each cell type, we assign enhancers only to genes for which the promoters are 'active' (that is, genes for which the gene is expressed and that promoter drives its expression). We defined active promoters as those in the top 60% of activity (geometric mean of chromatin accessibility and H3K27ac ChIP–seq counts). We used the following set of TSSs (one per gene symbol) for ABC predictions, as previously described[4]: https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction/blob/v0.2.1/reference/RefSeqCurated.170308.bed.CollapsedGeneBounds.bed. We note that this approach does not account for cases in which genes have multiple TSSs either in the same cell type or in different cell types.

For computing global statistics of ABC enhancer–gene connections (Extended Data Fig. 2), we considered all distal element–gene connections ('distal elements' here refers to chromatin-accessible regions that are not promoters of protein-coding genes) with an ABC score ≥ 0.015 and within a distance of 2 Mb.

Processing ABC predictions for variant overlaps

To intersect ABC predictions with variants, we took the predictions from the ABC model and applied the following additional processing steps. First, we considered all distal element–gene connections with an ABC score ≥ 0.015 (Extended Data Fig. 4; lower threshold than our previous study[4] to increase recall and identify gain-of-function variants that increase enhancer activity) and all distal or proximal promoter–gene connections with an ABC score ≥ 0.1 (based on our previous experimental data[4]). Second, we shrunk the approximately 500-bp regions by 150-bp on either side, resulting in an approximately 200-bp region centred on the summit of the accessibility peak. This is because, although the larger region is important for counting reads in H3K27ac ChIP–seq, which occur on flanking nucleosomes, DNA sequences important for enhancer function—such as transcription factor footprints—are most often found in the central nucleosome-free region[49]. In practice, this adjustment does not substantially affect the enrichment of fine-mapped IBD variants (Extended Data Fig. 5d). Third, we included enhancer–gene connections spanning up to 2 Mb, which is greater than the maximum distance of the longest-range enhancer–gene connection that we identified in CRISPR experiments to date (around 1.8 Mb).

CRISPRi-FlowFISH

We applied CRISPRi-FlowFISH to very sensitively test the effects of distal elements on gene expression[4]. In brief, CRISPRi-FlowFISH involves targeting putative enhancers with many independent gRNAs (median = 45) in a pooled screen using CRISPR interference (CRISPRi), which alters chromatin state through the recruitment of catalytically dead Cas9 fused to a KRAB effector domain. After infecting a population of cells with a gRNA lentiviral library, we estimate the expression of a gene of interest. Specifically, we first use fluorescence in situ hybridization (FISH, Affymetrix PrimeFlow assay) to quantitatively label single cells according to their expression of an RNA of interest. Second, we sort labelled cells with fluorescence-activated cell sorting (FACS) into six bins based on RNA expression. Third, we use high-throughput sequencing to determine the frequency of gRNAs from each bin. And finally, we compare the relative abundance of gRNAs in each bin to compute the effects of gRNAs on RNA expression. CRISPRi-FlowFISH provides around 300-bp resolution to identify regulatory elements, has the power to detect effects of as low as 10% on gene expression and provides effect size estimates that match those observed in genetic deletion experiments[4].

Here we applied CRISPRi-FlowFISH to comprehensively test all putative enhancers in an approximately 700-kb region around *PPIF*, and to validate additional selected enhancers (for 12 additional genes) that contained variants that are associated with IBD or other immune diseases or traits. For CRISPRi-FlowFISH experiments for *PPIF*, we designed gRNAs tiling across all accessible regions (here, defined as the union of the MACS2 narrow peaks and 250-bp regions on either side of the MACS2 summit) in the range chr10:80695001–81407220 in any of the following cell lines (with or without stimulation as described above): THP1, BJAB, Jurkat, GM12878, K562, Karpas-422 or U937. For CRISPRi-FlowFISH experiments for other genes, we included gRNAs targeting the promoter of the predicted gene and selected enhancer(s) nearby. We excluded gRNAs with low specificity scores or low-complexity sequences as previously described[4]. We generated cell lines expressing KRAB-dCas9-IRES-BFP under the control of a doxycycline-inducible promoter (Addgene, 85449) and the reverse tetracycline transactivator (rtTA) and a neomycin resistance gene under the control of an EF1α promoter (ClonTech), as previously described[12]. For each, we sorted polyclonal populations with high BFP expression after addition of doxycycline. For GM12878 cells, we used an alternative lentiviral construct to express the rtTA with a hygromycin-resistance gene, as GM12878 appeared to be resistant to selection with neomycin (also known as G418).

We performed CRISPRi-FlowFISH using ThermoFisher PrimeFlow (ThermoFisher 88-18005-210) as previously described, using the probe sets listed in Supplementary Table 13. To ensure robust data, we only included probe sets with twofold signal over unstained cells, and required an uncorrected knockdown at the TSS of >20%. We analysed these data as previously described[4]. In brief, we counted gRNAs in each bin using Bowtie[50] to map reads to a custom index, normalized gRNA counts in each bin by library size, then used a maximum-likelihood estimation approach to compute the effect size for each gRNA. We used the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (implemented in the R stats4 package) to estimate the most likely log-normal distribution that would have produced the observed guide counts, and the effect size for each gRNA is the mean of its log-normal fit divided by the average of the means from all negative-control gRNAs. As previously described, we scaled the effect size of each gRNA in a screen linearly so that the strongest 20-guide window at the TSS of the target gene has an 85% effect, in order to account for non-specific probe binding in the RNA FISH assay (this is based on our observation that promoter CRISPRi typically shows 80–90% knockdown by qPCR)[4]. We averaged the effect sizes of each gRNA across replicates and computed the effect size of an element as the average of all gRNAs targeting that element. We assessed significance using a two-sided *t*-test comparing the mean effect size of all gRNAs in a candidate element to all negative-control guides. We computed the false-discovery rate (FDR) for elements using the Benjamini–Hochberg procedure and used an FDR threshold of 0.05 to call significant regulatory effects.

Comparison of ABC predictions to genetic perturbations. We evaluated the ability of the ABC score and other enhancer–gene prediction methods to predict the results of genetic perturbations using a precision–recall framework. For this analysis the true-positive data are the experimentally measured element–gene pairs that are statistically significant and for which the perturbation of the element resulted in a decrease in gene expression. For these comparisons, (1) we only considered experimentally tested elements in which the element is not within 500 bp of an annotated gene TSS; (2) for perturbations using CRISPRi we excluded pairs in which the element resides within the gene body of the assayed gene; (3) we excluded non-significant pairs for which the power to detect a 25% change in gene expression was less than 80%; and (4) we only included pairs for which the gene is protein-coding (although the ABC model can make predictions for non-coding genes, many of the other predictions methods that we compare to do not make predictions for such genes).

For each experimentally measured element–gene–cell-type tuple, we intersected this tuple with the tuple in the predictions database corresponding to the same cell type, same gene and overlapping element. In cases in which the genomic bounds of an experimentally tested element overlap multiple predicted elements, we aggregated the prediction scores using an aggregation metric appropriate to each individual predictor (for ABC we used 'sum', for correlation- or confidence-based predictors we used 'max'). Similarly, if the predictor did not make a prediction for a particular tuple, it received an arbitrary quantitative score less than the least confident score for the predictor (for ABC we used 0, for other predictors we used 0, −1, 1 as appropriate). Supplementary Table 5 lists the experimental data merged with the predictions.

In the cases in which an enhancer–gene prediction method did not make cell-type-specific predictions, we evaluated the predictions against experimental data in all cell types (Extended Data Fig. 4c). We calculated the area under the precision–recall curve for predictors, or, if the predictor was defined at only one point, we multiplied the precision by the recall.

Similarity of ABC predictions among replicates and biosamples

We evaluated the reproducibility of ABC predictions derived from replicate epigenetic experiments. For each biosample for which

independent biological replicate experiments for both ATAC-seq (or DNase-seq) or H3K27ac ChIP–seq were available, we generated ABC predictions for replicates 1 and 2 separately. To facilitate the reproducibility analysis, when computing the ABC scores for replicate 2, we used the candidate enhancer regions from replicate 1. (Using different sets of candidate regions can confound computing reproducibility. For example, the procedure to define candidate regions (peak calling, extending and merging) could call two separate approximately 500-bp regions in one replicate, but merge them into an around 1-kb region in the second replicate due to minor differences in the peak summits between replicates. In such a case the ABC score of the approximately 1-kb region would be equal to the sum of the ABC scores of the 500-bp regions.)

We then evaluated the quantitative reproducibility of the predictions (Extended Data Fig. 3c) and the number of predictions shared between replicates (Extended Data Fig. 3d). We observed that on average 85% of enhancer–gene predictions in one replicate are shared in the other replicate (at an ABC score threshold of 0.015). The fraction of shared connections between biological replicates increased as the ABC score cut-off increased: 95% of connections called in replicate 1 at a higher confidence threshold of 0.02 were also called in replicate 2 (at the default threshold of 0.015).

We also evaluated the extent to which the reproducibility of ABC predictions depends on the reproducibility of the underlying epigenetic data. For each biosample, we computed the correlation between the ATAC-Seq (or Dnase-Seq) or H3K27ac ChIP–seq signals in the candidate regions for that biosample. As expected, we observed that the fraction of shared ABC predictions between replicates increased as the correlation of the underlying epigenetic data increased (Extended Data Fig. 3e).

We used a similar calculation to compare ABC predictions across cell types and biosamples. For each pair of biosamples, we computed the fraction of predicted enhancer–gene connections shared between the pair. For this analysis, we used the shrunken ABC elements (around 200 bp, see 'Processing ABC predictions for variant overlaps') and considered two connections to be shared if the elements overlapped by at least 1 bp and predicted to regulate the same gene.

Genetic data and fine-mapping

We downloaded summary statistics for IBD, Crohn's disease and ulcerative colitis (European ancestry only)[51] from https://www.ibdgenetics.org/downloads.html. We obtained fine-mapping posterior probabilities and credible sets from a previously published study[9] and analysed the top two conditionally independent credible sets for each locus. We also analysed variants from IBD GWAS loci that were not fine-mapped in this study[51,52]; for each such locus, we analysed all variants from the 1000 Genomes Project in linkage disequilibrium with the lead variant ($r^2 > 0.2$) and weighted each variant evenly (probability = 1/number of variants in linkage disequilibrium). We observed similar results for cell-type enrichments with or without including these non-fine-mapped sets. Throughout this text, analyses of 'IBD' signals are defined as signals associated with Crohn's disease, ulcerative colitis or both.

We obtained fine-mapping results and summary statistics for 71 other traits based on an unpublished analysis (J.C.U., M. Kanai and H.K.F., unpublished data) that analysed data from the UK Biobank (application 31063; fine-mapping data are available at https://www.finucanelab.org/data). In this analysis, up to 361,194 individuals of white British ancestry with available phenotypes and variants with INFO > 0.8, minor allele frequency > 0.01%, and Hardy–Weinberg equilibrium $P > 1 \times 10^{-10}$ were included in the GWAS. Covariates for the top 20 principal components, sex, age, age^2, sex × age, sex × age^2 and dilution factor, where applicable, were controlled for in the association studies. Quantitative traits were inverse rank transformed and associations were estimated using BOLT-LMM[53] for quantitative traits and SAIGE[54] for binary traits. In-sample dosage linkage disequilibrium was computed using LDStore[55], and phenotypic variance was computed empirically. Fine-mapping was performed using the sum of single effects (SuSiE) method[56], allowing for up to ten causal variants in each region. Prior variance and residual variance were estimated using the default options, and single effects (potential 95% credible sets) were pruned using the standard purity filter such that no pair of variants in a credible set could have $r^2 > 0.25$. Regions were defined for each trait as ±1.5 Mb around the most significantly associated variant (with this window chosen based on the linkage disequilibrium structure in the human population), and overlapping regions were merged. Variants in the MHC region (chr. 6: 25–36 Mb) were excluded as were 95% credible sets containing variants with fewer than 100 minor allele counts. Coding (missense and predicted loss of function) variants were annotated using the variant effect predictor v.85[57]. For analysis with ABC, we excluded neuropsychiatric traits (for which we expect existing enhancer–gene maps will not include the appropriate cell types), traits with no entirely noncoding GWAS signals, and analysed only the variants that SuSiE assigned to belong to 95% credible sets (cs_id ! = -1).

For all traits, except where specified, we considered only the 'noncoding credible sets'—that is, those that did not contain any variant in a coding sequence or within 10 bp of a splice site annotated in the RefGene database (downloaded from UCSC Genome Browser on 24 June 2017)[58]. We note that predictions for all credible sets, both coding and noncoding, are reported in Supplementary Table 10 to facilitate future analyses.

Defining enriched biosamples for each trait

For a given trait, we intersected variants with PIP ≥ 10% in noncoding credible sets with ABC enhancers (or other genomic annotations). For each biosample, we calculated a *P* value using a binomial test comparing the fraction at which PIP ≥ 10% variants overlapped ABC enhancers with the fraction at which all common variants overlap ABC enhancers in that cell type. We calculated the latter using common variants in the 1000 Genomes Projects as described in the 'Stratified linkage disequilibrium score regression' section. For each trait, we defined a biosample as significantly enriched for that trait if the Bonferroni-corrected binomial *P* value was <0.001.

Comparison of enrichment of fine-mapped variants in enhancer regions

We compared the enrichment of fine-mapped variants in ABC enhancers and other enhancer definitions (Extended Data Fig. 5c). We analysed each of the previous studies from Fig. 1c reporting cell-type-specific enhancer–gene predictions as well as ChromHMM enhancers in blood cells downloaded from the BLUEPRINT Project[59,60].

Stratified linkage disequilibrium score regression

We compared cell-type enrichments observed for fine-mapped variants to those observed with stratified linkage disequilibrium score regression (S-LDSC), which considers not only variants in genome-wide significant GWAS loci but also in sub-significant loci. To do so, we used S-LDSC to assess the enrichment of disease or trait heritability in ABC enhancers, considering all variants across the genome[61]. We analysed the ABC enhancer regions as defined above, and ran linkage disequilibrium score regression using the baselineLD_v1.1 model using the 1000G_EUR_Phase3_baseline file (downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/; defined as variants in the 1000 Genomes Project with minor allele count >5 in 379 European samples). For comparison, we also analysed heritability enrichment in all other accessible regions for each trait. Specifically, we took the list of MACS2 peaks (FDR < 0.05), removed those that overlapped ABC enhancers, and used these regions in S-LDSC.

Partitioning the genome into disjoint functional categories

To compare the frequency of variants occurring in ABC enhancers as opposed to other functional elements such as coding sequences

and splice sites (Extended Data Fig. 5f), we partitioned the genome into the following functional categories, using the RefGene database (downloaded from UCSC Genome Browser on 24 June 2017): coding sequences, 5′ and 3′ untranslated regions of protein-coding genes, splice sites (within 10 bp of a intron–exon junction of a protein-coding gene) of protein-coding genes, promoters (±250 bp from the gene TSS) of protein-coding genes, ABC enhancers in 131 biosamples, other accessible regions in the same biosamples not called as ABC enhancers, and other intronic or intergenic regions. These categories may overlap; a disjoint annotation was created by assigning each nucleotide to the first of any overlapping categories in the order above (for example, nucleotides in both coding sequences and ABC enhancers were counted as coding sequences).

Overlap with H3K27ac QTLs

We downloaded H3K27ac data in monocytes and T cells from the Blueprint Project and analysed allele-specific signals called by the WASP method as previously described[62]. We examined variants associated with allelic effects on H3K27ac where FDR < 0.05 and the variant was located within the associated peak. Of 52 fine-mapped IBD variants that overlapped ABC enhancers in any T cell or myeloid biosample, 10 variants had genome-wide significant allelic effects on H3K27ac ChIP–seq (3.6-fold enrichment versus other common variants that overlap ABC enhancers in T cells or myeloid cells). For example, we found significant allelic effects for rs11643024 in T cells (linked by ABC to suppressor of cytokine signalling 1 (*SOCS1*) located 93 kb away) and for rs9808651 in monocytes (linked by ABC to *ERG*, located 32 kb away). This analysis indicates that some prioritized causal variants have allelic effects on enhancer activity.

Evaluating gene prediction methods

Curated genes for IBD. We analysed a previously curated list of IBD disease-associated genes[14]. To evaluate methods to connect noncoding GWAS variants to genes, we analysed credible sets within 1 Mb of exactly 1 of these known genes that did not contain any protein-coding or splice site variants. In cases in which the gene was curated based on evidence from coding variation, we examined nearby conditionally independent noncoding signals, which may act via regulatory effects on the same gene that carries the coding variant.

Gene set enrichment for IBD predictions. As a second approach for comparing methods for identifying causal genes in IBD GWAS loci, we examined the extent to which the predicted genes were enriched for any gene sets. To do so, we downloaded curated and Gene Ontology gene sets from the Molecular Signatures Database[63]. We analysed all 93 noncoding IBD credible sets. For each gene set, we tested whether it was enriched in the genes predicted by a given method, using the set of all genes within 1 Mb of IBD credible sets as the background, excluding HLA genes. For Extended Data Fig. 6b, we applied this approach to each of the methods described in Fig. 1c, selected the five gene sets with the highest enrichment that also had at least five identified genes and hypergeometric test $P < 10^{-4}$. We plotted the cumulative density function (CDF) of the enrichments for each of the methods across the union of the top 5 gene sets identified by any of the methods.

Likely causal gene for blood traits. We identified genes carrying fine-mapped coding variants with high posterior probability (PIP ≥ 50%) associated with one of 10 blood cell traits (Baso, Eosino, Hb, LOY, Lym, MCH, Mono, Neutro, RBC, WBC), for which our ABC maps and other previous predictions include many of the relevant cell types. We used the Variant Effect Predictor[57] to identify protein-truncating variants and damaging missense variants. Because of the large number of total genome-wide significant associations, many loci had multiple known genes within 1 Mb of the signal, which may or may not point to the same gene. Accordingly, we examined noncoding credible sets in which exactly 1 gene within 1 Mb carried such a coding variant, and in which that gene was not more than the tenth closest gene to the variant with the highest PIP. To compute the enrichment for ABC and other methods in identifying such genes, we calculated: Enrichment = (true positive predictions/predictions)/(positive genes in the considered credible sets/all genes near the considered credible sets).

Comparisons to alternative variant-to-gene predictions

We compared ABC-Max to previously published results from alternative methods to link regulatory variants to disease-associated genes.

eQTL colocalization (Open Targets Platform). OpenTargets.org performed colocalization analysis between IBD GWAS signals[51,52] and eQTLs and protein QTLs (pQTLs) using coloc[15]. This analysis involved QTL datasets from a variety of sources including dozens of human tissues and many immune cell types, including from the eQTL Catalogue[64]. We downloaded colocalization results from ftp://ftp.ebi.ac.uk/pub/databases/opentargets/genetics/190505/v2d_coloc/ on 1 February 2020, and examined genes showing colocalization with an eQTL or pQTL in any biosample. We considered genes with coloc h4 probability ≥ 0.9, and h4/h3 ratio ≥ 2. We used the coloc h4 probability to rank genes within each locus.

eQTL colocalization (JLIM). The colocalization of IBD GWAS signals with eQTLs in $CD4^+$ T cells, $CD14^+$ monocyte and lymphoblastoid cell lines was analysed previously[18]. We obtained their colocalized genes from table 2 of ref. [18]. We used the JLIM *P* value to rank genes within each locus.

TWAS (S-PrediXcan and multiXcan). multiXcan was developed previously and was used to compare GTEx v.7 eQTLs to IBD summary statistics[20]. We downloaded dataset 6 from ref. [20] and compared genes within each locus using the multiXcan *P* value.

Mendelian randomization (SMR). A Mendelian randomization-based approach (SMR) was previously used to connect IBD GWAS signals to effects on gene expression using eQTL data from 24 tissues[21]. We downloaded supplementary table 3 from ref. [21] and defined predicted genes in any tissue. We used the SMR FDR to rank genes within each locus.

COGS. PC-Hi-C data in many blood cell types were previously analysed to link GWAS variants to target genes[22]. We downloaded supplementary table 3 from ref. [22] (tab 2) and analysed genes linked with COGS scores ≥ 0.5.

In all cases, we combined predictions of disease-associated genes for IBD, ulcerative colitis and Crohn's disease.

Comparisons to previous enhancer–gene predictions

We compared the ABC model to methods using alternative enhancer–gene linking approaches. For each of the methods below, we downloaded previous predictions of enhancer–gene links, and assessed their ability to predict enhancer–gene regulation in CRISPR datasets (Extended Data Fig. 4) and their ability to identify IBD-associated genes (Fig. 1c and Extended Data Fig. 6b). For the latter analysis, we used the predictions from each method to overlap fine-mapped variants (PIP ≥ 10%) with enhancers in any cell type and assigned variants to the predicted gene(s).

PC-Hi-C gene predictions. We downloaded data S1 peak data from a previously published study[22], representing PC-Hi-C data from 9 haematopoietic cell types, and selected the promoter–distal region pairs with CHiCAGO score ≥ 5. For comparison to CRISPR data we used the CHiCAGO score as a quantitative predictor.

ENCODE DNase correlation. Distal accessible elements with gene promoters were previously linked by looking at correlation of DNase I hypersensitivity across 125 cell and tissue types from ENCODE[29]. We downloaded these links from ftp://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/byDataType/openchrom/jan2011/dhs_gene_connectivity/genomewideCorrs_above0.7_promoterPlusMinus500kb_withGeneNames_32celltypeCategories.bed8.gz. GWAS loci with high-confidence fine-mapped variants that overlapped these regions were assigned to the linked gene(s).

eRNA–mRNA correlation (FANTOM5). The transcriptional activity of enhancers and TSSs was previously linked using the FANTOM5 CAGE expression atlas[27]. We downloaded these predictions from http://enhancer.binf.ku.dk/presets/enhancer_tss_associations.bed.

Enhancer–gene correlation (ChromHMM-RNA). Gene expression was previously correlated with five active chromatin marks (H3K27ac, H3K9ac, H3K4me1, H3K4me2 and DNase I hypersensitivity) across 56 biosamples, and these correlation links were then used to make predictions for the predicted enhancers (regions with the '7Enh' ChromHMM state) in 127 biosamples from the Roadmap Epigenome Atlas[25,45]. We downloaded these predictions from www.biolchem.ucla.edu/labs/ernst/roadmaplinking and made predictions using the confidence score.

Enhancer–gene correlation in single-cell RNA and ATAC data. Single-cell ATAC-seq and RNA-sequencing data in peripheral blood and bone marrow mononuclear cells, $CD34^+$ bone marrow cells and cancer cells from patients with leukaemia were previously analysed and the ATAC-seq signal in accessible elements was correlated with the expression of nearby genes[26]. We downloaded these predictions from https://github.com/GreenleafLab/MPAL-Single-Cell-2019 and used the correlation in samples from healthy individuals as the quantitative score. Cell-type-specific links were not reported.

EnhancerAtlas 2.0. EAGLE was previously used to predict enhancer–gene interactions across a number of human tissues and cell lines[30]. The method calculates a score based on six features obtained from the information of enhancers and gene expression: correlation between enhancer activity and gene expression across cell types, gene expression level of target genes, genomic distance between an enhancer and its target gene, enhancer signal, average gene activity in the region between the enhancer and target gene, and enhancer–enhancer correlation. We downloaded enhancer annotations for 104 cell types from http://www.enhanceratlas.org/.

Enhancer–gene correlation (DNase-seq and microarray gene expression). The DNase I signal and gene expression levels were previously correlated using data from 112 human samples representing 72 cell types to identify regulatory elements and to predict their targets[28]. We downloaded these predictions from http://dnase.genome.duke.edu/ and used the correlation as the quantitative score. Cell-type-specific links were not reported.

JEME (joint effect of multiple enhancers). Correlations between gene expression and various enhancer features (for example, DNase1 and H3K4me1) were previously computed across multiple cell types to identify a set of putative enhancers[24]. Then, a sample-specific model is used to predict the enhancer gene connections in a given cell type. We downloaded the lasso-based JEME predictions in all ENCODE+Roadmap cell types from http://yiplab.cse.cuhk.edu.hk/jeme/. We used the JEME confidence score as a quantitative score.

TargetFinder. A model was previously generated to predict whether nearby enhancer–promoter pairs are located at anchors of Hi-C loops based on chromatin features[31]. We downloaded the TargetFinder predictions from https://raw.githubusercontent.com/shwhalen/targetfinder/master/paper/targetfinder/combined/output-epw/predictions-gbm.csv. For each distal element–gene pair in our dataset, we searched to see if the element and gene TSS overlapped with an enhancer and promoter loop listed in this file. If so, we assigned the pair a score corresponding to the 'prediction' column from this file; otherwise the pair received a score of 0.

Comparisons to previous GWAS gene prediction methods

Finally, we compared our results to two previous GWAS gene prediction methods.

MAGMA (Multi-marker Analysis of GenoMic Annotation). We applied MAGMA[65] to the summary statistics for IBD[51] using the 1000 Genomes Project reference panel to compute gene-level association statistics and gene–gene correlations using the SNP-wise mean gene analysis and a 0-kb window around the gene body for mapping SNPs to genes. For each gene, MAGMA computes a gene P value from the mean χ^2 statistic of SNPs in the gene body and its approximate sampling distribution. The gene P value is converted to a z-score using the probit function. The resulting z-score reflects the gene–trait association after correcting for linkage disequilibrium among SNPs within the gene body. We assigned each IBD locus to the gene with the maximum positive z-score.

DEPICT (Data-driven Expression Prioritized Integration for Complex Traits). We applied DEPICT, which leverages pathway analysis and cell-type enrichment analysis from gene expression datasets to analyse genome-wide significant loci and prioritize causal genes[23]. We applied DEPICT to the summary statistics for each trait using the 1000 Genomes Project reference panel and the 14,461 reconstituted gene sets from DEPICT to prioritize genes in genome-wide significant loci. First, we performed PLINK clumping with a P-value threshold of 5×10^{-8}, r^2 threshold of 0.05, and distance threshold of 500 kb as recommended by the DEPICT software to identify associated variants. Loci are defined by taking all genes that reside within boundaries defined by the most distal variants in either direction with linkage disequilibrium of more than 0.5 to the lead variant identified by PLINK clumping. DEPICT then scores genes by correlating their membership to reconstituted gene sets to those of other genes in genome-wide significant loci and performs a bias adjustment for the scores. Finally, to prioritize genes in each locus, we prioritized the single gene in each genome-wide significant locus with the most-significant P value.

Cell-type-specific gene set enrichments

We assessed whether the cell-type specificity of the ABC predictions for IBD variants could aid in identifying gene pathways enriched in IBD GWAS loci. To do so, we defined seven cell-type categories based on the biosamples available in our compendium and based on biological categories relevant to IBD: mononuclear phagocytes, B cells, T cells, other haematopoietic cells, fibroblasts, epithelial cells or tissues, and other cells or tissues. We then examined the extent to which the genes predicted by ABC in any cell type category, or in each individual cell type category, were enriched for gene sets from the Molecular Signatures Database[63], as described above.

Assessing pleiotropy across 72 traits

We identified genes linked to multiple traits through different variants. To identify such genes, we identified genes that were predicted by ABC-Max to be linked to at least two different traits by two different variants, where that gene was not linked to the same two traits by any single variant (that is, a gene linked to two traits by each of two variants would not fit this criteria). Because some of the 72 traits show high genetic correlation, we repeated these analyses in a subset of 36 traits that were selected to show pairwise genetic correlation below

a threshold ($|r_g| < 0.2$), plus IBD. We observed similar effects in this subset of the data, in which genes linked to multiple traits by different variants were more likely to have complex enhancer landscapes and large amounts of nearby noncoding genomic sequence.

Single-guide qPCR validation of e-PPIF

Two non-overlapping guides against *PPIF* TSS (GCGGCCGAGCGGC TTCCCGT and GAACCTGGGCAAGCCAATAA) and e-PPIF (GACTCAAGA TACCACCACCGG and GATGGCCAGTTTGGGAACGT), along with four non-targeting control guides (GAGATGAAAGCGCAGCTAGGG, GGGCGCTTACGCGCGGGCCG, GCGCGCGCTAACTGGCGCTA, GATGTG TTGTAACCTCCACT), were cloned into sgOpti as previously described[12]. We generated stable cell lines expressing each sgRNA by lentiviral transduction in 8 μg ml^{-1} polybrene by centrifugation at 1,200*g* for 45 min with 200,000 CRISPRi THP1 cells in 24-well plates. After 24 h, we selected for transduction with 1 μg ml^{-1} puromycin (Gibco) for 72 h then maintained cells in 0.3 μg ml^{-1} puromycin. We plated sgRNA-expressing stable cell lines at 100,000 cells per ml in 1 μg ml^{-1} doxycycline and collected cells 48 h later by lysing in buffer RLT (Qiagen). For each sgRNA, we generated three independent polyclonal cell populations through triplicate infections and treated each cell population with doxycycline twice, for a total of six biological replicates per sgRNA. We extracted RNA from 20,000 cells per experiment in buffer RLT (Qiagen) using Dynabeads MyOne Silane beads (Thermo Fisher), treated samples with TURBO DNase (Thermo Fisher), and cleaned again with Dynabeads MyOne Silane beads. We used AffinityScript reverse transcriptase (Agilent Technologies) and random nonamer primers to convert RNA to cDNA. We performed qPCR using SYBR Green I Master Mix (Roche) with primers for *PPIF* (AGAACTTCAGAGCCCTGTGC, CATTGTGGTTGG TGAAGTCG) and *GAPDH* (AGCACATCGCTCAGACAC, GCCCAATACGACCA AATCC) and calculated differences using the $\Delta\Delta C_t$ method.

Assessing the effect of *PPIF* and e-PPIF on mitochondrial membrane potential

We synthesized a pool of 105 gRNAs including 40 negative control gRNAs, 9 gRNAs targeting the promoter of *PPIF* and 5 gRNAs targeting the *PPIF* enhancer (Agilent Technologies; see Supplementary Table 14), cloned these gRNAs into CROP-seq-opti (Addgene, 106280), and transduced THP1 cells at a multiplicity of infection of 0.3 to ensure most cells contained 1 gRNA integration.

For untreated and LPS-stimulated conditions, we plated 10,000,000 cells per replicate with 1 μg ml^{-1} doxycycline. After 44 h, we added 1 μg ml^{-1} LPS and collected cells for staining 4 h later. For the PMA LPS condition, we plated 10,000,000 cells per replicate and added 1 μg ml^{-1} doxycycline for 48 h. To differentiate into macrophage-like cells, we added fresh media with 20 ng ml^{-1} PMA and 1 μg ml^{-1} doxycycline for an additional 24 h, confirming that cells adhered to the tissue-culture plate. We washed out the PMA and added fresh medium with 1 μg ml^{-1} doxycycline and incubated cells for 45 h to recover and further differentiate cells. We then added 100 ng ml^{-1} LPS for 3 h, collected cells, washed 3 times with cold PBS and proceeded to mitochondrial staining.

We stained cells with MitoTracker Red (200 nM, Thermo Fisher, M7512) and MitoTracker Green (200 nM, Thermo Fisher, M7514) according to the manufacturer's protocol and sorted cells into 3 bins according to their ratio of MitoTracker Red (which stains mitochondria dependent on $\Delta\psi_m$) to MitoTracker Green (which stains mitochondria independent of $\Delta\psi_m$), excluding a small population of depolarized cells with very low $\Delta\psi_m$ (Extended Data Fig. 10f). We extracted genomic DNA and amplified and sequenced gRNAs from cells in each bin as previously described[4].

We aligned and counted gRNAs in each bin as described above for FlowFISH experiments. For each gRNA, we summed counts across the two biological replicates. We then calculated the frequency fold change in Fig. 4d and Extended Data Fig. 10g by dividing gRNA reads per million by the mean value for negative-control gRNAs, and dividing values in each bin by the value for bin 3.

Data visualization

We developed a web application for interactively exploring ABC enhancer–gene connections, by extending HiGlass[66], a flexible genome browser toolkit: https://flekschas.github.io/enhancer-gene-vis/ (Supplementary Fig. 2). The application features three linked views: the enhancer view in the top left, the gene view in the bottom left and the DNA accessibility view on the right. The enhancer view supports pan-and-zoom for navigation and allows the user to focus on a gene or genomic region. The gene and DNA accessibility views are linked to the enhancer view and update automatically. Each view is interactive, customizable and exportable. The design of the user interface and visualizations have been refined through several participatory exploration sessions.

Reporting summary

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

Data availability

Data for the immune cell line ATAC-seq and H3K27ac ChIP–seq analyses can be found in the NCBI GEO under accession number GSE155555. gRNA counts from CRISPRi screens can be found in Supplementary Tables 3, 14. UK Biobank fine-mapping data for 71 traits are available from https://www.finucanelab.org/data. ABC predictions in 131 biosamples can be found at https://www.engreitzlab.org/abc/.

Code availability

The ABC model is available on GitHub (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction). This is the codebase used to generate the ABC predictions for this manuscript, and can be used to run the ABC model on new biosamples. ABC-Max and paper-specific analyses can be found on GitHub (https://github.com/EngreitzLab/ABC-GWAS-Paper). This repository implements the ABC-Max pipeline and can be used to reproduce specific analyses in this study.

43. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol .Biol.* **109**, 21.29.1–21.29.9 (2015).
44. Zhu, J. et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).
45. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
46. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
47. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
49. Vierstra, J. et al. Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. 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).
52. 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).
53. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
54. Zhou, W. et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
55. Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
56. 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. B* **82**, 1273–1300 (2020).
57. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
58. Fujita, P. A. et al. The UCSC genome browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).
59. Carrillo-de-Santa-Pau, E. et al. Automatic identification of informative regions with epigenomic changes associated to hematopoiesis. *Nucleic Acids Res.* **45**, 9244–9259 (2017).

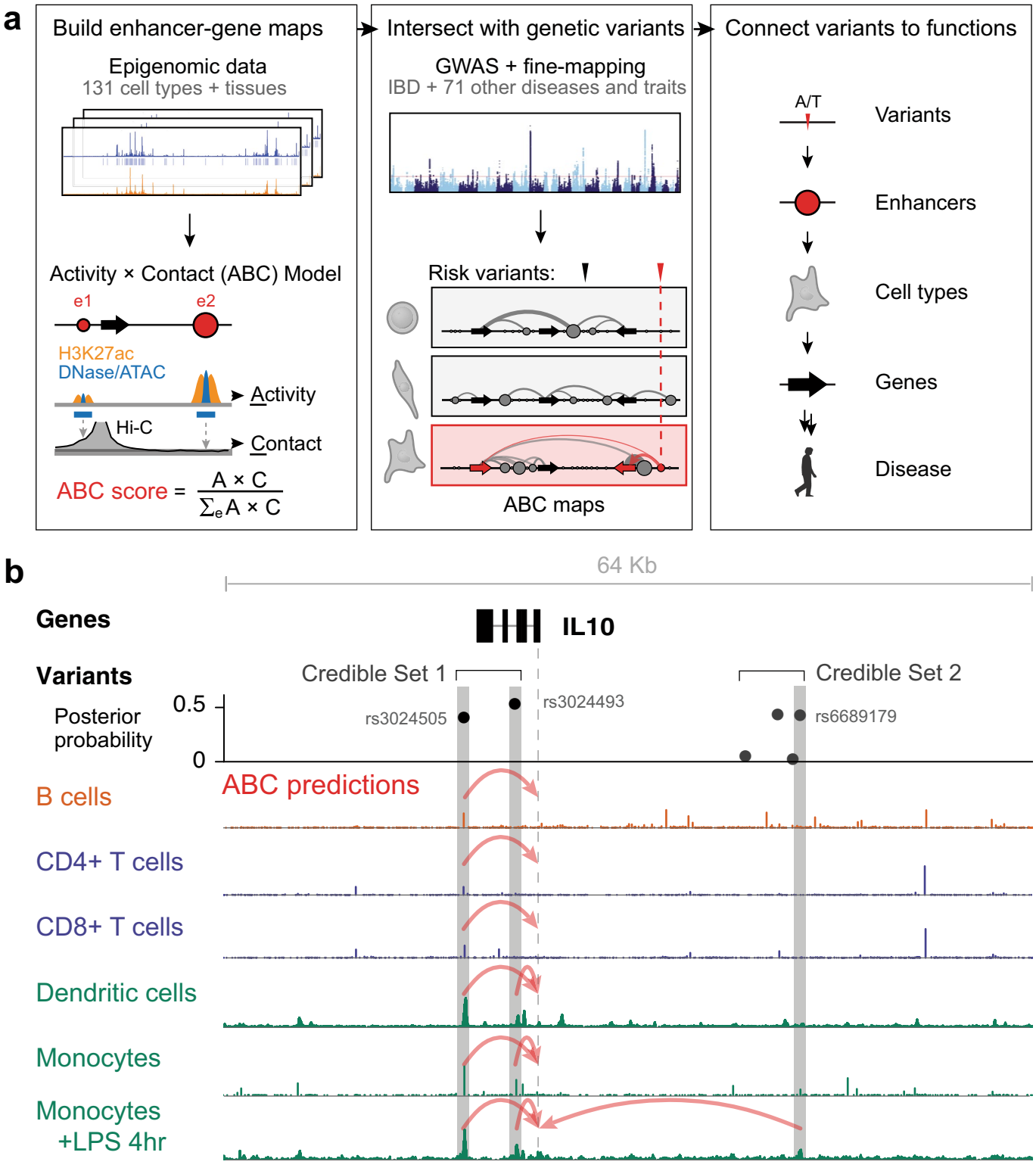
60. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
61. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
62. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414 (2016).
63. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
64. Kerimov, N. et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. Preprint at https://doi.org/10.1101/2020.01.29.924266 (2021).
65. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
66. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
67. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
68. Novakovic, B. et al. β-Glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* **167**, 1354–1368 (2016).
69. Donnard, E. et al. Comparative analysis of immune cells reveals a conserved regulatory lexicon. *Cell Syst.* **6**, 381–394 (2018).

Acknowledgements This work was supported by the Broad Institute (E.S.L.); an NIH Pathway to Independence Award (K99HG009917 and R00HG009917 to J.M.E.); an NHGRI Genomic Innovator Award (R35HG011324 to J.M.E.); the Harvard Society of Fellows (J.M.E.); Gordon and Betty Moore and the BASE Research Initiative at the Lucile Packard Children's Hospital at Stanford University (J.M.E.); NHGRI P50HG006193 (N.H.); NIDDK P30DK043351 (R.J.X.); NIH U01CA200059 (H.P.); NHGRI U01HG009379, NIMH R01MH101244 and NIMH R37MH107649 (A.K.P.); NIDDK K01DK114379 (H.H.); the Zhengxu and Ying He Foundation, the Stanley Center for Psychiatric Research and NIAID K22AI153648 (J.P.R.); NHGRI U24HG009446 (A.K.); an NSF Graduate Research Fellowship (DGE-1656518 to B.R.D.); and a Siebel Scholarship (F.L.). We thank L. Schweitzer, M. Gentili, M. Biton, C. Smillie, A. Regev, M. Kanai, D. Graham, N. Shoresh, S. Gazal, B. Cleary, R. Cui, P. Rogers, V. Subramanian, G. Schnitzler, R. Gupta, M. Claussnitzer, N. Sinnott-Armstrong, T. Majarian, A. Manning and members of the Lander lab, Hacohen lab and Variant-to-Function Initiative for discussions or technical assistance. This research has been conducted using the UK Biobank Resource.

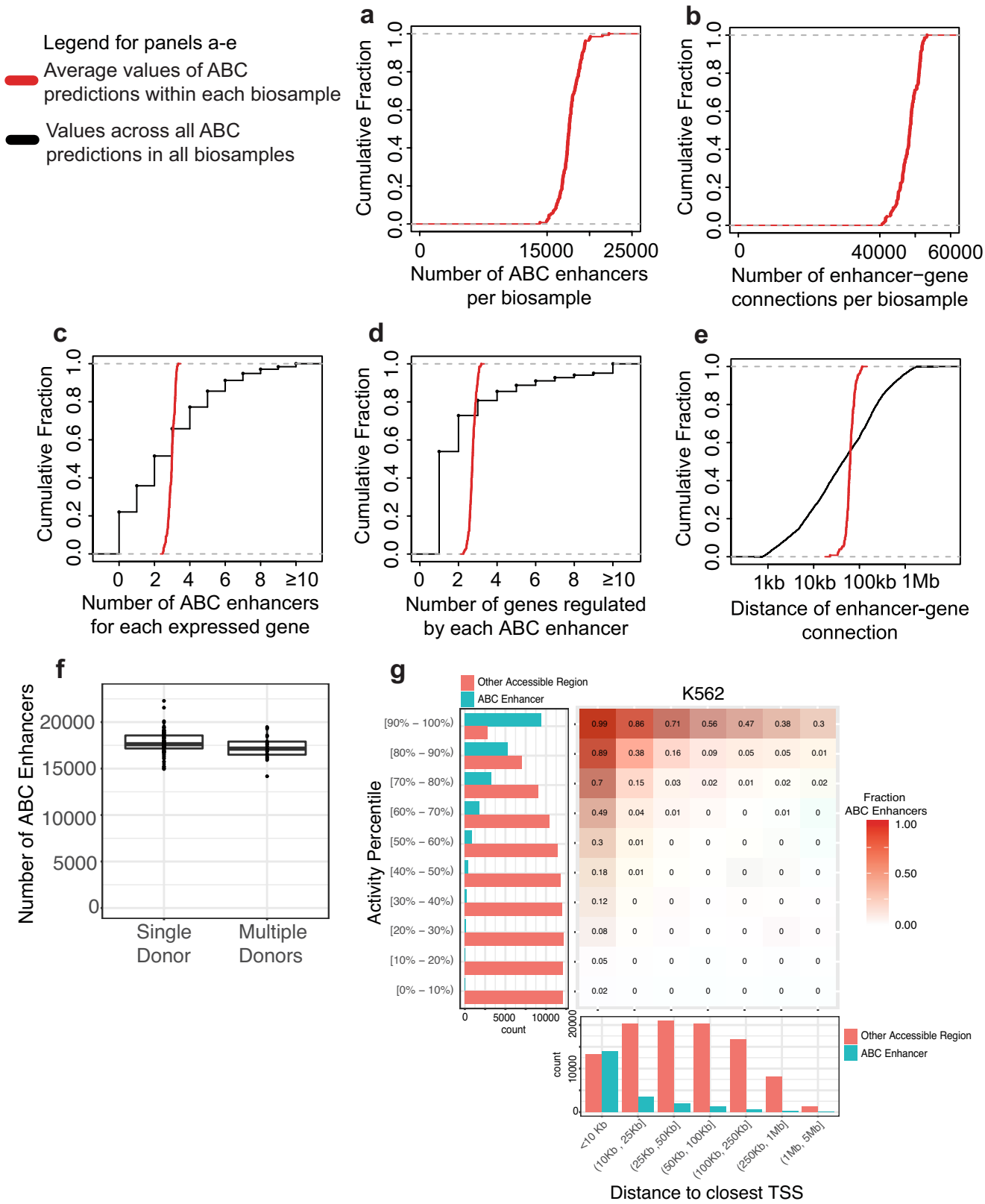
Author contributions J.N. and J.M.E. developed computational methods. J.N., D.T.B., J.M.E., C.P.F., B.R.D., T.A.P., T.R.J., K.M., H.M.N., H.Y.K., A.C. and R.L.C. conducted data analysis. D.T.B., C.P.F., P.G., B.R.D., G.M. and T.H.N. conducted CRISPR experiments. T.H.N., J.P.R., T.M.E., C.B.E. and M.K. collected epigenomic datasets. J.C.U., E.M.W., M.J.D., H.H. and H.K.F. contributed fine-mapping analysis. F.L. and H.P. built visualization tools. K.D., A.L.P., A.K., R.J.X., M.J.D., H.H. and H.K.F. contributed to data interpretation and design of analyses. N.H., E.S.L. and J.M.E. supervised the study. J.N., D.T.B., C.P.F., P.G., B.R.D., E.S.L. and J.M.E. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests J.M.E., C.P.F. and E.S.L. are inventors on a patent application on CRISPR methods filed by the Broad Institute related to this work (16/337,846). Until recently, E.S.L. served on the Board of Directors for Codiak BioSciences and Neon Therapeutics; served on the Scientific Advisory Board of F-Prime Capital Partners and Third Rock Ventures; was affiliated with several non-profit organizations including serving on the Board of Directors of the Innocence Project, Count Me In and Biden Cancer Initiative, and the Board of Trustees for the Parker Institute for Cancer Immunotherapy; and served on various federal advisory committees. E.S.L. is currently on leave from MIT and Harvard. C.P.F. is now an employee of Bristol Myers Squibb. T.A.P. is now an employee of Boston Consulting Group. R.J.X. is a cofounder of Jnana Therapeutics and Celsius Therapeutics. M.J.D. is a founder of Maze Therapeutics. N.H. holds equity in BioNTech and consults for Related Therapeutics. All other authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03446-x.
Correspondence and requests for materials should be addressed to E.S.L. or J.M.E.
Peer review information *Nature* thanks Annique Claringbould, Judith Zaugg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

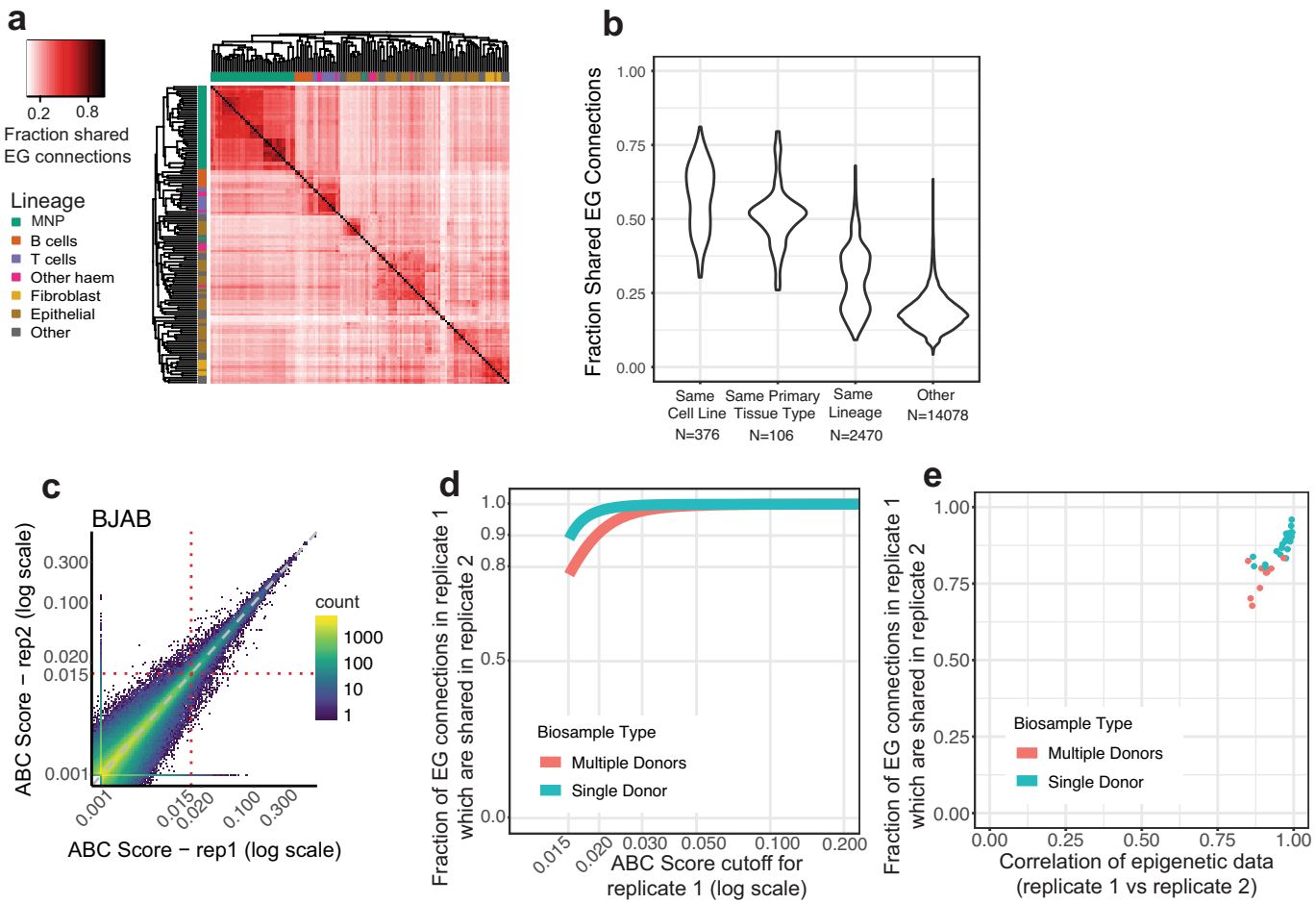


Extended Data Fig. 1 | ABC maps connect fine-mapped variants to enhancers, genes and cell types. a, Overview of approach. **b**, ABC predictions connect two IBD GWAS signals to *IL10*. Signal tracks show DNase-seq or ATAC-seq (based on availability of data). Red arrows represent ABC predictions connecting variants to *IL10*. Dashed line shows the TSS. Grey bars highlight fine-mapped variants that overlap with ABC enhancers in at least one cell type. Credible set 1 contains two variants, both of which overlap with enhancers predicted to regulate *IL10* in various cell types. Credible set 2 contains four variants, one of which overlaps with an enhancer predicted to regulate *IL10* in monocytes stimulated with LPS.

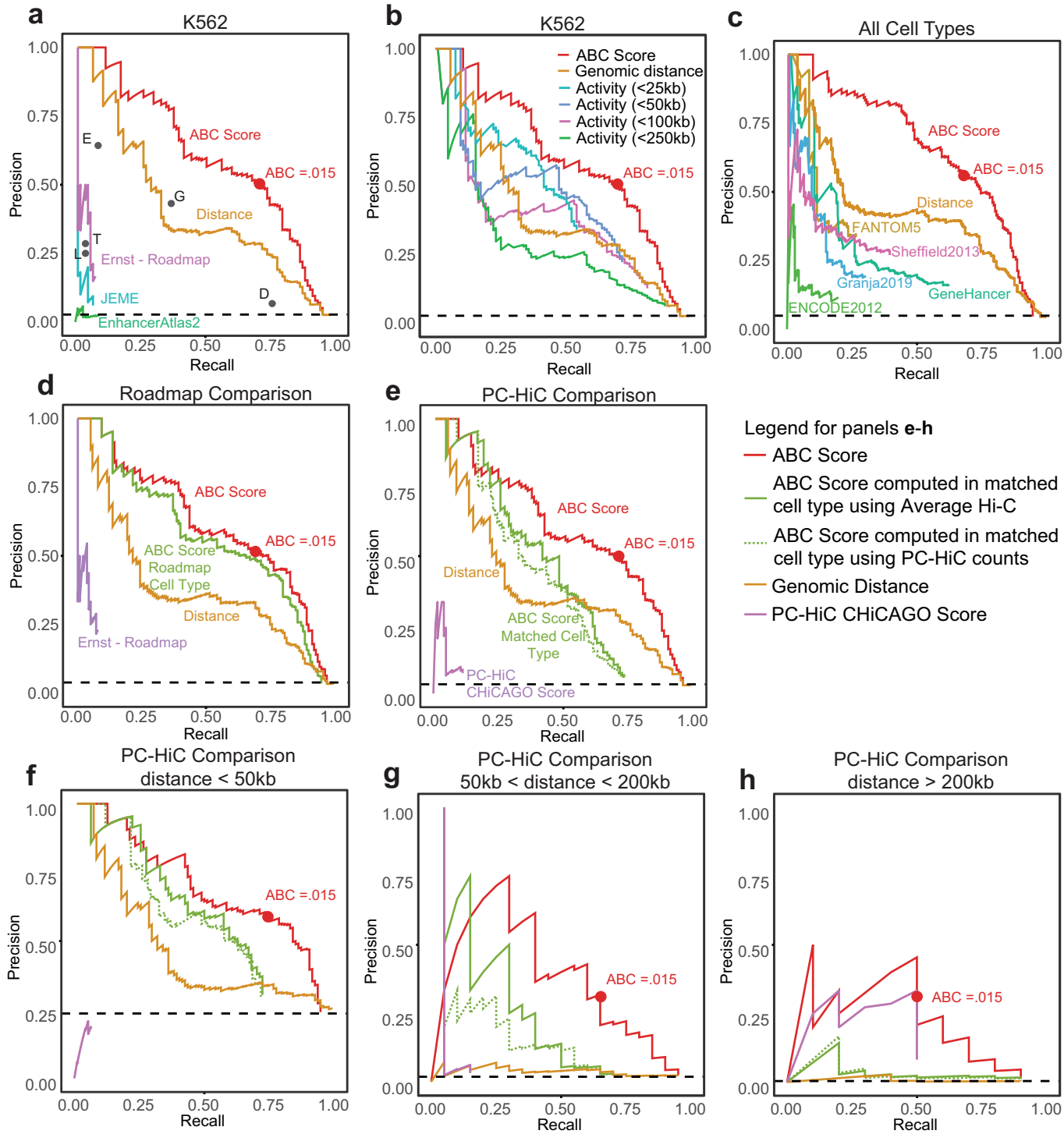


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Properties of ABC predictions. a, Cumulative fraction of the number of ABC enhancers within each biosample (median = 17,605). **b**, Cumulative fraction of the number of enhancer–gene connections within each biosample (median = 48,441). **c**, Cumulative fractions of the number of enhancers predicted to regulate each gene across all biosamples (black line; median = 2, mean = 2.8) and the mean number of enhancers predicted to regulate each gene within each biosample (red line; median = 2.8). **d**, Cumulative fractions of the number of genes regulated by each ABC enhancer across all genes and all biosamples (black line; median = 1, mean = 2.7) and the mean number of genes regulated by each ABC enhancer within each biosample (red line; median = 2.7). **e**, Cumulative fractions of the genomic distances between the enhancer and the gene for each predicted enhancer–gene connection across all genes and all biosamples (black line; median = 62,929 bp) and the median genomic distance between each enhancer–gene connection within each biosample (red line; median = 62,782 bp). **f**, Number of ABC enhancers predicted in 131 biosamples stratified by whether the epigenomic data for the biosample is derived from one or multiple donors. We do not observe significant differences between these distributions (two-sided Wilcoxon rank-sum test, $P = 0.10$). Box plot displays median, 25th and 75th percentiles. **g**, Summary of ABC predictions in K562. Plot includes 122,410 non-promoter Dnase hypersensitive sites (DHS elements) in K562. Each element is classified as an 'ABC enhancer' if the element is predicted to regulate at least one gene, or 'other accessible region' otherwise. The *x* axis represents the distance from the element to the closest TSS of an expressed gene. The *y* axis represents the percentile bin of the activity of the element (in terms of DHS and H3K27ac signals) among these 122,410 elements. The colouring of the heat map represents the fraction of elements in the corresponding distance and activity bins that are ABC enhancers.

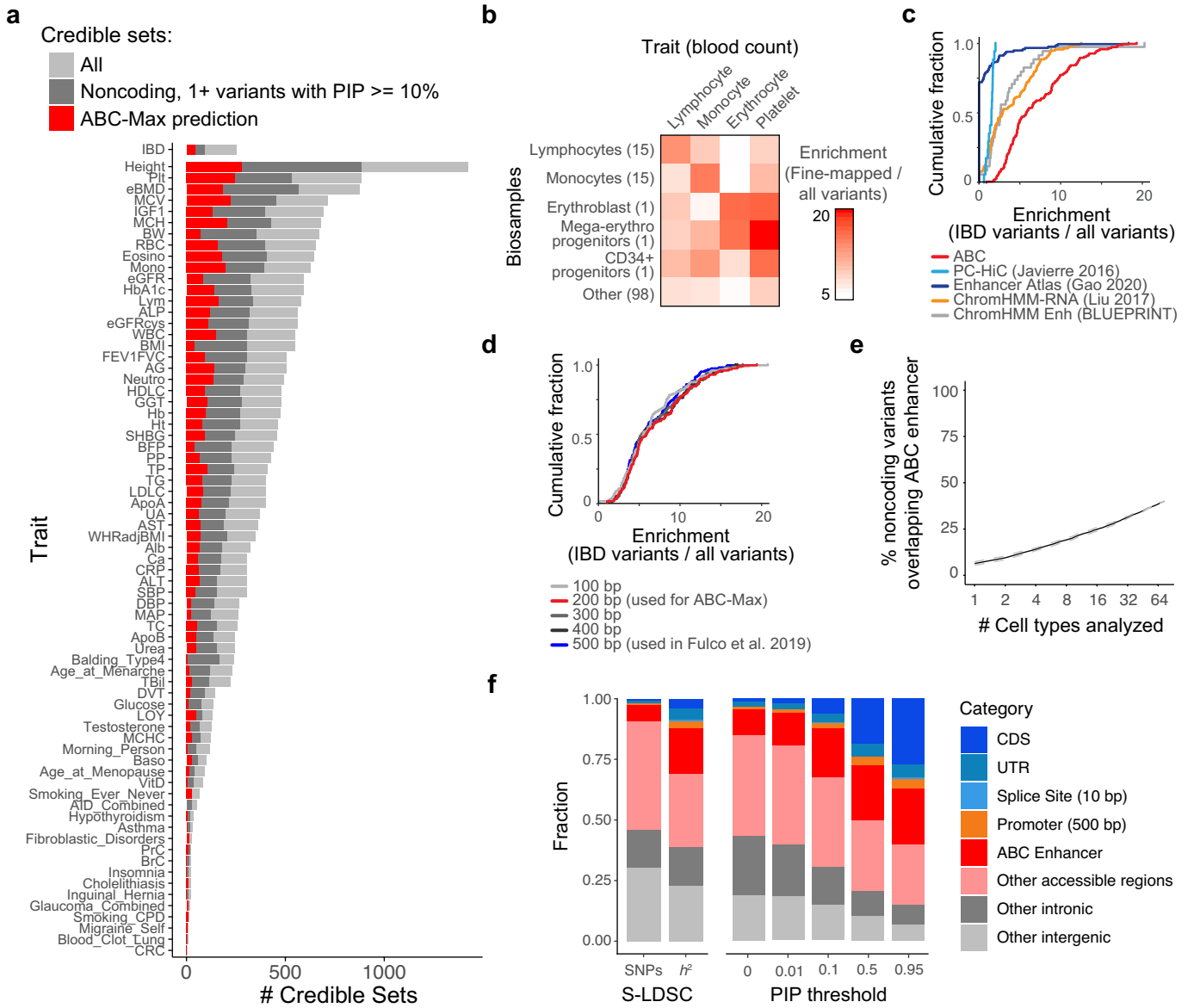


Extended Data Fig. 3 | Distinctness and reproducibility of ABC predictions. **a**, Distinctness of predictions across biosamples. Biosample versus biosample (131 × 131) heat map. The colour of the (*i*,*j*) pixel in the heat map represents the fraction of enhancer–gene connections ('E-G connections'—which are defined to be an element–gene pair for which the ABC score is greater than 0.015) in biosample *i* that have a corresponding overlapping prediction in biosample *j*. Two connections are considered overlapping if the predicted genes are the same and the enhancer elements overlap. Rows and columns are ordered by hierarchical clustering. A median of 19% (median of row medians) of enhancer–gene connections are shared across distinct biosamples. **b**, Distribution of shared connections by relatedness of samples. Distribution of the fraction of shared connections in **a** stratified by the relatedness of the samples. Each pair of biosamples is classified as: 'same cell line', which indicates the same cell line under different perturbation conditions or from different compendia; 'same primary tissue type', which indicates the same tissue type from different compendia; 'same lineage', which indicates samples from the same lineage classification as in **a**; 'other' refers to all other pairs of samples. **c**, Quantitative reproducibility of ABC predictions. ABC scores computed using independent biological replicates of epigenomic data (ATAC-seq and H3K27ac ChIP–seq) from the BJAB cell line. Each data point is an element–gene pair. **d**, Fraction of shared enhancer–gene connections between replicates increases as ABC score cut-off increases. *x* axis, cut-off on the ABC score; *y* axis, for a given cut-off of the ABC score, the fraction of element–gene pairs with an ABC score greater than the cut-off in sample 1 that have an ABC score > 0.015 in sample 2. Each biosample is classified as: 'multiple donors', which indicates that the epigenetic data for this biosample is derived from different donors, or 'single donor', which indicates that the epigenetic data for this biosample is derived from the same donor or cell line. For 'single donor' biosamples, replicates represent independent epigenomic experiments from the same donor or cell line; for 'multiple donor' biosamples, replicates represent epigenomic experiments from different donors. Separate curves are computed for each biosample and then the average across biosamples is plotted. **e**, Fraction of shared enhancer–gene connections increases as reproducibility of underlying epigenetic data increases. Each data point represents a biosample. *x* axis, geometric mean of correlation of ATAC-seq (or DNase-seq) and H3K27ac ChIP–seq signal in candidate regions computed using replicate epigenetic experiments. *y* axis, fraction of enhancer–gene connections with ABC score > 0.015 in replicate 1 that also have ABC score > 0.015 in replicate 2. Colours as in **d**.

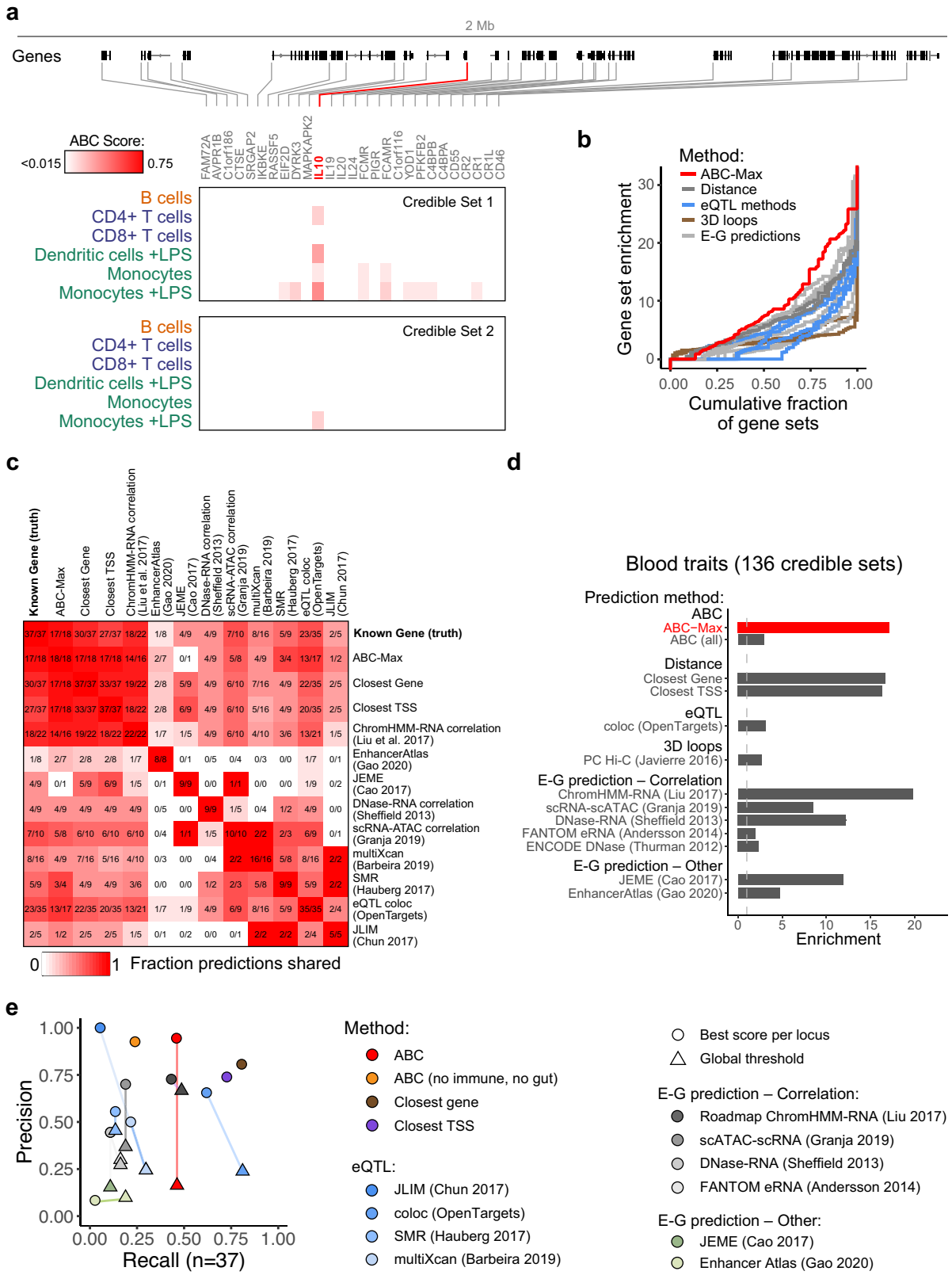


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | ABC performs well at identifying regulatory enhancer–gene connections in CRISPR datasets. a, Comparison of enhancer–gene predictors to experimental CRISPR data in K562 cells. Each of these predictors makes K562-specific predictions. Curves represent continuous predictors. Dots represent binary predictors as follows: E, each gene is predicted to be regulated only by the element closest to its TSS; G, each element is predicted to regulate only the nearest (to TSS) expressed gene; T, TargetFinder method[31]; L, elements and genes at opposite ends of HiCCUPS loops derived from Hi-C data are predicted as a connection[67]; D, an element–gene pair is a predicted positive if and only if the element and the gene are contained within the same contact domain[67]. The red dot on ABC score curve: precision and recall achieved using a threshold on the ABC score of 0.015. Dashed black line, rate of experimental positives. **b**, Comparison of ABC predictions using a binary distance threshold to experimental CRISPR data in K562 cells. 'Activity (<*X* kb)' represents a model in which the score for an element–gene pair is the activity of the element (in terms of DHS and H3K27ac signals) multiplied by a binary indicator (1 if the distance is <*X* kb, and 0 otherwise). The ABC model using quantitative Hi-C outperforms the models based on binary thresholds indicating that Hi-C data are a critical component of the ABC model. **c**, Comparison of ABC and other enhancer–gene predictors in full CRISPR dataset. Comparison of enhancer–gene predictors to experimental CRISPR data in K562, GM12878, NCCIT, BJAB (with or without stimulation), Jurkat (with or without stimulation), THP1 (with or without stimulation) cells and primary hepatocytes. For ABC, we used the predictions in the cell type corresponding to the CRISPR experiments. Because ABC is the only method that makes predictions in all of these cell types, we used this plot to compare ABC to other methods that make predictions without cell-type information. We consider each enhancer–gene pair predicted by these methods to be a prediction in all cell types. **d**, Comparison of ABC and Ernst-Roadmap predictions[25]. Comparison of enhancer–gene predictors to experimental CRISPR data in K562, GM12878 and unstimulated Jurkat, BJAB and THP1 cells. The red line represents a comparison of ABC scores computed using epigenetic data from the same cell type as the CRISPR experiment was performed. To compare Roadmap predictions to CRISPR data, we made cell-type substitutions because the Roadmap predictions did not include BJAB, Jurkat and THP1 cells: for BJAB CRISPR data we compared to predictions in the Roadmap B cell sample (E032); for THP1 data we used the Roadmap monocyte sample (E124); and for Jurkat data we used the Roadmap T cell sample (E034). To directly compare the performance of ABC and Ernst-Roadmap methods in matched cell types, we also calculated ABC performance using the same cell type substitutions (green line)—for example, CRISPR data in BJAB cells were compared to ABC scores computed using epigenetic data from the Roadmap B cell sample (E032). **e**, Comparison of ABC to PC-Hi-C. Comparison of enhancer–gene predictors to experimental CRISPR data in K562 and unstimulated BJAB, THP1 and Jurkat cells. The red line represents a comparison of ABC scores computed using epigenetic data from the same cell type as the CRISPR experiment was performed. To compare PC-Hi-C CHiCAGO predictions (purple line) to CRISPR data, we made cell-type substitutions because PC-HiC data are not available for K562, BJAB, Jurkat and THP1 cells: for K562 CRISPR data we compared to CHiCAGO scores in erythroblasts; for BJAB CRISPR data we compared to total B cells; for THP1 data we compared to monocytes; and for Jurkat data we compared to activated $CD4^+$ T cells. To directly compare the performance of ABC and PC-HiC methods in matched cell types, we also calculated ABC performance using the same cell-type substitutions (green lines). The solid green line represents ABC scores for which the contact component is derived from the average Hi-C dataset used throughout this study. The dashed green line represents ABC scores for which the contact component is derived from the raw counts in PC-HiC experiments (see Methods). **f**–**h**, Comparison of ABC to PC-Hi-C stratified by distance. These panels represent the comparison of the same predictors as in **e** while stratifying the experimental dataset in **e** based on the distance between the tested element and gene TSS. Of the 4,078 element–gene connections in the experimental dataset, 398 are at a distance of <50 kb (of which 94 are experimental positives, 24% positive rate), 1,102 are between 50 kb and 200 kb (20 positives, 2% positive rate) and 2,578 are at a distance of >200 kb (10 positives, 0.4% positive rate). Given the differences in positive rates between the stratifications (indicated by dashed black lines), it is appropriate to compare precision–recall curves within each stratification, but it is not appropriate to compare the precision–recall curve of the same predictor across stratifications.

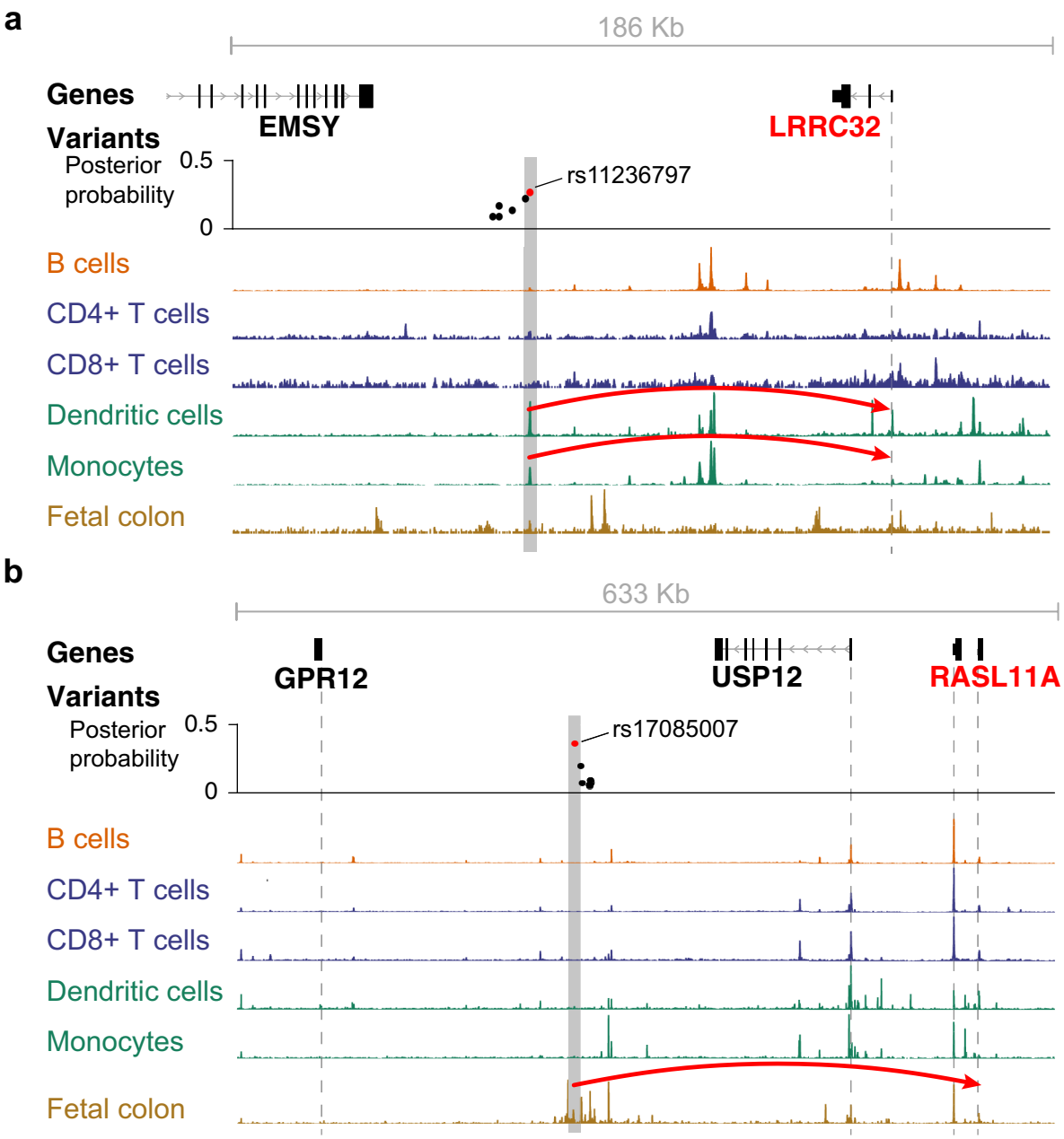


Extended Data Fig. 5 | Fine-mapped GWAS variants are highly enriched in ABC enhancers. a, Number of credible sets analysed for 72 diseases and complex traits. Light grey shows the total number of fine-mapped credible sets. Dark grey shows the number of such credible sets with no coding or splice site variants, and at least one variant with PIP ≥ 10%. Red shows the number of credible sets for which ABC-Max makes a prediction (that is, a variant with PIP ≥ 10% overlaps an ABC enhancer in a biosample that shows global enrichment for that trait). See Supplementary Table 7 for trait descriptions and additional statistics. **b**, Enrichment of fine-mapped variants (PIP ≥ 10%) associated with four blood cell traits in ABC enhancers in the corresponding blood cell types or progenitors. Enrichment = (fraction of fine-mapped variants/fraction of all common variants) overlapping regions in each cell type. Numbers of biosamples in each category are shown in parentheses. **c**, Enrichment of fine-mapped IBD variants (PIP ≥ 10%) in ABC enhancers and other sets of previously defined enhancers. Cumulative density function shows distribution across cell types. **d**, Enrichment of fine-mapped variants (PIP ≥ 10%) in ABC enhancers resized in different ways. Regions of at least 500-bp (blue line) are used to count reads, as defined previously. Regions were then shrunk by 150-bp on each side (minimum size of element = 200 bp) for overlapping with variants. Grey lines show alternative sizes, which do not appear to notably affect enrichments of fine-mapped variants. **e**, Percentage of noncoding variants across all traits that overlap an ABC enhancer in an enriched biosample, as a function of the number of cell types analysed. Biosamples (131) were grouped into 74 cell types or tissues; and analysed in random order. Black line, mean across 20 random orderings. Dashed grey lines, 95% confidence intervals. **f**, Fraction of variants or heritability for all 72 traits contained in different categories of genomic regions: coding sequences (CDS), untranslated regions (UTR), splice sites (within 10 bp of an intron–exon junction of a protein-coding gene), promoters (±250 bp from the gene TSS), ABC enhancers in 131 biosamples, other accessible regions not called as ABC enhancers, and other intronic or intergenic regions. In cases in which a variant overlaps more than one category, the variant was assigned to the first category that it overlapped (that is, variants in coding sequences were not also counted in the ABC category; Methods). Left, all common variants or heritability (h^2, as estimated by S-LDSC in inverse-variance-weighted meta-analysis across 72 traits). Right, fraction of variants above a threshold on the fine-mapping PIP.

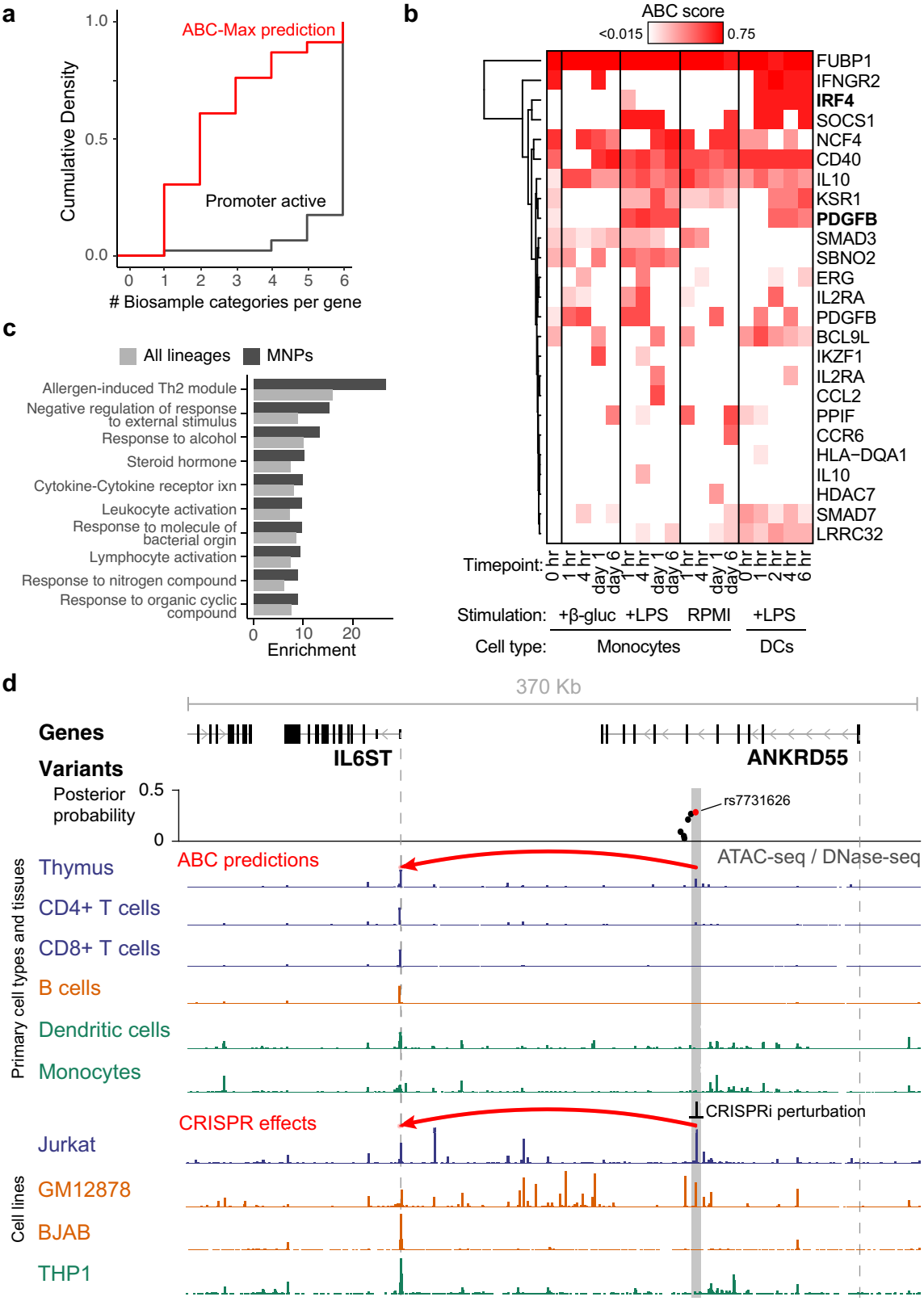


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | ABC enhancer maps connect GWAS variants to known genes. a, ABC predictions for IBD credible sets linked to *IL10*. Heat map shows ABC scores for each gene within 1 Mb in selected primary immune cell types. Credible set 1 is linked by ABC to multiple genes, but *IL10* (red) has the strongest ABC score in any cell type. **b**, Cumulative density plot showing enrichment for gene sets in MSigDB among the genes prioritized by each method[63]. Methods are coloured and categories as in Fig. 1c. For each method, we first identified the top 5 most enriched significant gene sets in the predictions of that method (82 gene sets total). Then, we calculated the levels of enrichment of all 82 gene sets in the predictions of each method. **c**, Comparison of predictions for the 37 IBD credible sets near known genes. Fraction predictions shared = (credible sets for which both methods predict the same gene)/(credible sets for which both methods make a prediction). For example, 16 credible sets have predictions from both ABC-Max and ChromHMM-RNA correlation, and the two methods predict the same gene in 14 out of 16 credible sets. **d**, Enrichment of likely causal genes for 10 blood traits (defined by common coding variants) for various prediction methods. Enrichment reflects the number of correctly predicted genes identified divided by the baseline of choosing random genes in each of the loci with a prediction. **e**, Precision–recall plot for identifying known IBD-associated genes, comparing additional variations on the prediction methods (related to Fig. 1c). For ABC, we compared ABC-Max (assigning each credible set to the gene with the maximum ABC score, red circle), ABC-Max excluding all immune and gut tissue biosamples (orange circle) and ABC-All (assigning each credible set to all genes linked to enhancers, red triangle). For other methods that provided quantitative scores, we similarly compared choosing the gene with the best score per locus (circles) with choosing all genes above the global thresholds previously reported in each study (triangles). In most cases, the best gene per locus outperformed using a global threshold.

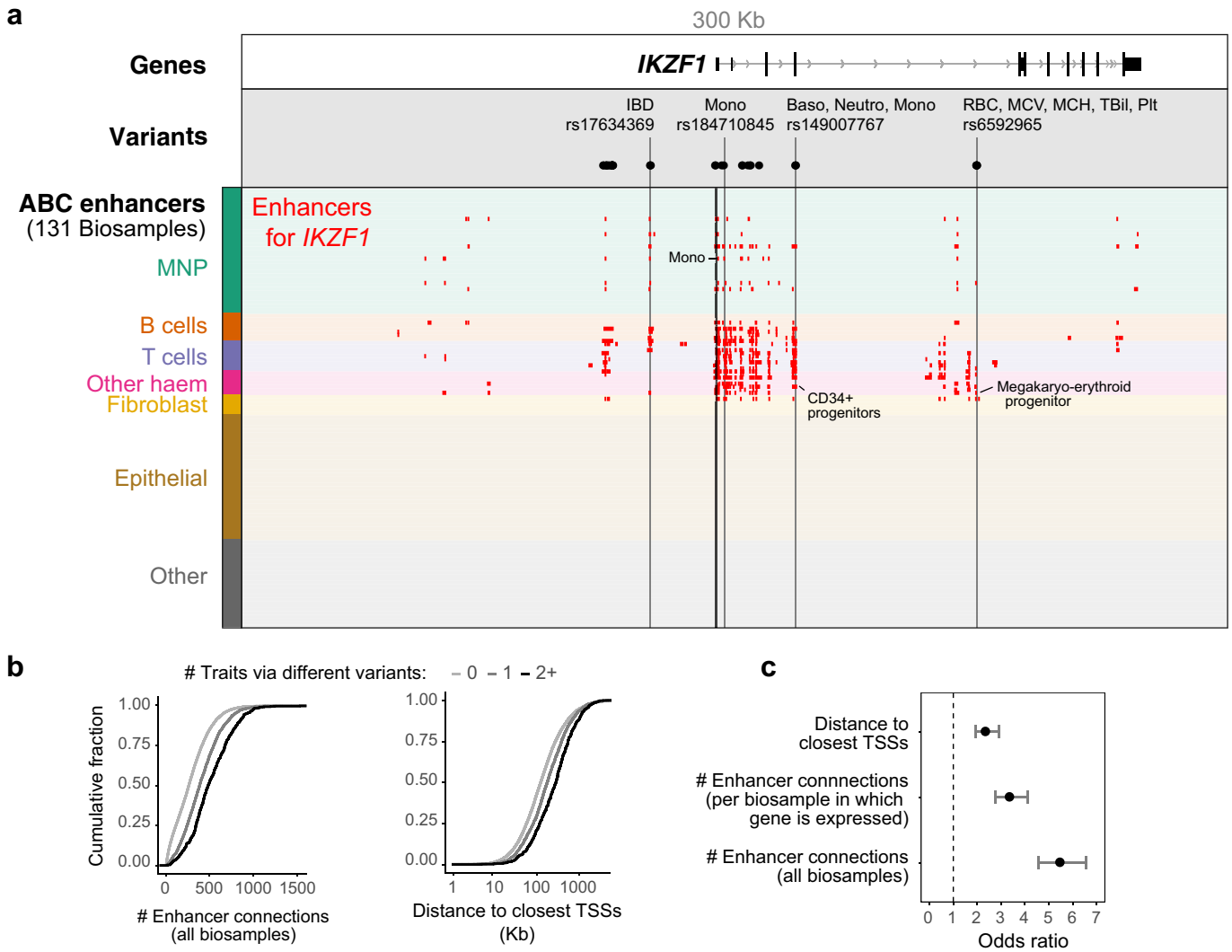


Extended Data Fig. 7 | ABC-Max predictions at *LRRC32* and *RASL11A* loci. **a**, **b**, ABC-Max predictions and chromatin state in primary immune cells and fetal colon tissue at two IBD loci: *LRRC32* (**a**) and *RASL11A* (**b**). Red marks variants, enhancer–gene connections and target genes identified by ABC-Max. Grey bars highlight the variants overlapping ABC enhancers. Vertical dotted lines represent TSSs. 'DCs + LPS', dendritic cells stimulated with bacterial LPS for 4 h.

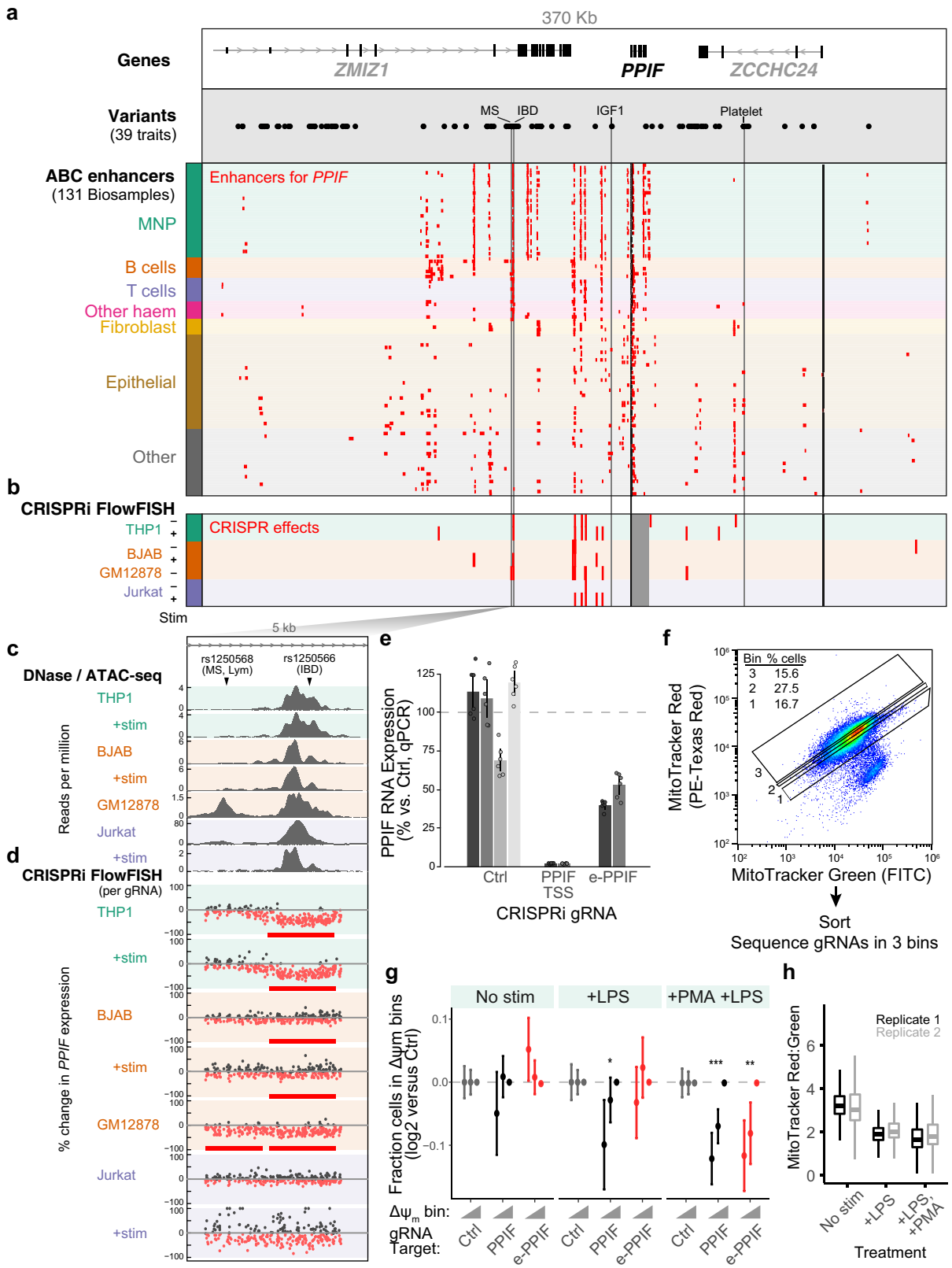


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cell-type specificity of ABC predictions. a, A comparison of the number of biosample groups (cell type lineages) in which the gene promoter is active versus the number of categories in which a variant is predicted to regulate the gene by ABC-Max. **b**, Heat map of ABC scores for predicted IBD-associated genes in resting and stimulated mononuclear phagocytes (from epigenomic data in monocytes[68] and dendritic cells[69]). *IRF4* and *PDGFB* (bold) are two examples for which ABC predictions are specific to a particular stimulated state (+LPS) and are not observed in unstimulated states. **c**, Enrichment for top gene sets identified when performing enrichment analysis among the 23 IBD-associated genes predicted by ABC-Max in mononuclear phagocytes (dark grey), versus when performing the same analysis among the 43 IBD-associated genes predicted in any biosample (light grey). The enrichment for a given gene is calculated as the ratio of the frequency at which ABC-predicted genes belong to the gene set, compared to the frequency at which all genes within 1 Mb of these loci belong to the gene set (Methods). **d**, A variant in an intron of *ANKRD55* is predicted by the ABC model to regulate *IL6ST* in thymus. The grey bar highlights the variant overlapping the predicted ABC enhancer. Vertical dotted lines represent TSSs. The red arc at the top denotes the ABC-Max prediction. The red arc at the bottom denotes that CRISPRi of the highlighted enhancer significantly affects the expression of *IL6ST* only in Jurkat cells.



Extended Data Fig. 9 | Genes linked by ABC to different traits by different variants. a, ABC links *IKZF1* to 2 traits by variants in 18 credible sets. Red boxes mark enhancers predicted to regulate *IKZF1*. The thick black line marks the *IKZF1* TSS. Black dots mark fine-mapped noncoding variants (PIP ≥ 10%) associated with one or more traits linked to *IKZF1* by ABC-Max. **b**, Genes linked to different traits via different variants have more complex enhancer landscapes. Cumulative distribution plots show the number of ABC enhancer–gene connections in all 131 biosamples (left) and the distance between the TSSs of the two closest neighbouring genes on either side of a gene, for each gene linked by ABC-Max to zero traits, one trait, or two or more traits through different variants (right). **c**, The complexity of the enhancer landscape of a gene is correlated with the odds of the gene being linked to multiple GWAS traits. The *x* axis shows the Wald odds ratio that a gene is connected to multiple GWAS traits, comparing genes in the top decile versus all other deciles of the corresponding enhancer complexity metric. The three enhancer complexity metrics are defined for each gene: the total number of enhancers linked to the gene by ABC in any biosample, the number of enhancers linked to a gene per biosample in which the promoter of the gene is active, and the genomic distance to the closest neighbouring TSS on either side of the gene. Dot, mean of the top decile genes ($n = 1,838$) versus all others ($n = 16,550$). Whiskers, 95% confidence intervals.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Enhancers and variants connected to *PPIF*. **a**, ABC predictions for variants near *PPIF*. Black dots represent either fine-mapped variants (PIP ≥ 10%) for IBD and UK Biobank traits, or lead variants for any phenotype from the GWAS Catalog[16] (the latter to show the approximate locations of signals for traits for which fine-mapping is not yet available). The 'IBD' label points to rs1250566. The 'MS' (multiple sclerosis) label points to rs1250568 (fine-mapped in ref. [2]). Red boxes mark enhancers predicted to regulate *PPIF*. Thick back lines mark TSSs. Thin black lines mark selected variants. **b**, CRISPRi-FlowFISH data for *PPIF* in seven immune cell lines and stimulated states. Red boxes mark distal enhancers (CRISPR gRNAs lead to a significant decrease in the expression of *PPIF*). Dark grey box marks the gene body of *PPIF*, for which CRISPRi cannot accurately assess the effects of putative regulatory elements[4]. **c**, Chromatin accessibility in 5-kb regions around the *PPIF* enhancer (e-PPIF). Signal tracks show ATAC-seq (for THP1 and BJAB) or DNase-seq (for GM12878 and Jurkat) data in reads per million. Arrows show the locations of variants associated with multiple sclerosis and lymphocyte count (Lym, rs1250568) and with IBD (rs1250566), which overlap with enhancers that regulate *PPIF* in distinct sets of cell types. **d**, Effect of each tested gRNA on *PPIF* expression, as measured by CRISPRi-FlowFISH (Methods). Dots, gRNAs for which the effect estimate is >0% (black) or <0% (red). Red bars show regions for which gRNAs have a significant effect on gene expression (FDR < 0.05), compared by a two-sided *t*-test to negative control gRNAs. **e**, Effects of eight individual gRNAs on *PPIF* expression in THP1 cells, as measured by CRISPRi and qPCR (Methods). *PPIF* expression is normalized to expression of *GAPDH* and to cells expressing negative control, non-targeting gRNAs (Ctrl). Error bars, 95% confidence intervals of the mean ($n = 6$ replicates per gRNA). **f**, Schema of pooled CRISPRi screen to examine the effects of *PPIF* and e-PPIF on mitochondrial membrane potential ($\Delta\psi_m$). Cells expressing a pool of gRNAs were stained with MitoTracker Red and MitoTracker Green and sorted into three bins of increasing red:green ratios. gRNAs from cells in each bin were PCR-amplified, sequenced and counted. **g**, Effects of CRISPRi gRNAs (targeting e-PPIF, *PPIF* promoter or negative controls) on $\Delta\psi_m$, quantified as the frequency of THP1 cells carrying those gRNAs with low or medium versus high MitoTracker Red signal (corresponding to bins 1, 2 and 3, respectively; superset of data in Fig. 4d). We tested THP1 cells in unstimulated conditions, stimulated with LPS, and differentiated with PMA and stimulated with LPS (Methods). Error bars, 95% confidence intervals for the mean of 40, 9, and 5 gRNAs for control, *PPIF* and e-PPIF, respectively. Two-sided Wilcoxon rank-sum test versus control; $^{*}P = 0.0163$, $^{**}P = 0.00426$, $^{***}P = 0.000356$. **h**, Ratios of MitoTracker Red (mitochondrial membrane potential) to MitoTracker Green (mitochondrial mass) signal in THP1 cells at baseline, stimulated with LPS and differentiated into macrophages with PMA and stimulated with LPS in biological duplicate (from left to right, $n = 8{,}044$, 99,683, 99,982, 99,968, 99,886 and 99,878; replicates were cultured, stimulated, stained and flow-sorted independently). Box represents median and interquartile range; whiskers show minimum and maximum. Stimulation with either LPS alone or both PMA and LPS leads to a reduction in red:green signal, indicating a reduction in mitochondrial membrane potential normalized to mitochondrial mass.

nature research

Corresponding author(s): Jesse Engreitz

Last updated by author(s): Mar 8, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	All published, open-source code (see Methods section)
Data analysis	Data was processed using published, open-source code (see Methods section). ABC prediction custom software is available at https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction. Code for additional analyses is available at: https://github.com/EngreitzLab/ABC-GWAS-Paper

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

A data statement is included.
Immune cell line ATAC-seq and H3K27ac ChIP-seq: NCBI GEO GSE155555
ABC predictions in 131 biosamples: https://www.engreitzlab.org/abc/

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	For CRISPRi-FlowFISH, we chose the number of gRNAs per element and number of negative control gRNAs according to power calculations based on prior similar datasets (Fulco, Nasser et al Nat Genet 2019) to provide >80% power for >25% effects on gene expression
Data exclusions	N/A
Replication	Epigenomic and CRISPRi experiments were performed in biological replicates and found to be reproducible.
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	H3K27ac monoclonal antibody (Cat #39685, Active Motif). anti-CD3 (Biolegend-317315). anti-CD40 (Invitrogen-140409-82).
Validation	H3K27ac antibody (ChIP-seq) was validated by comparing our ChIP-seq maps to prior tracks, and verifying enrichment of signal at promoters. Anti-CD3 and anti-CD40 antibodies (stimulation of immune cell lines) were validated by examining up-regulation of stimulus-responsive genes.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	THP1 (monocytic-like cell line, acute monocytic leukemia), BJAB (B cell-like cell line, EBV-negative inguinal Burkitt's lymphoma), GM12878 (EBV-immortalized lymphoblastoid cell line), U937 (monocytic-like cell line, histiocytic lymphoma), and Jurkat (T cell-like, T cell leukemia)
Authentication	Cell lines were authenticated by examining ChIP-seq signals at known marker genes
Mycoplasma contamination	Confirmed negative
Commonly misidentified lines (See ICLAC register)	None

ChIP-seq

Data deposition

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

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

Data access links
May remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155555

Files in database submission

GSM4706074 BJAB-ctrl-1_H3K27ac
GSM4706075 BJAB-ctrl-2_H3K27ac
GSM4706076 BJAB_ctrl_1-Tag-30_ATAC
GSM4706077 BJAB_ctrl_2-Tag-30_ATAC
GSM4706078 BJAB-stim-1_H3K27ac
GSM4706079 BJAB-stim-2_H3K27ac
GSM4706080 BJAB_stim_1-Tag-30_ATAC
GSM4706081 BJAB_stim_2-Tag-30_ATAC
GSM4706082 Jurkat-ctrl-1_H3K27ac
GSM4706083 Jurkat-ctrl-2_H3K27ac
GSM4706084 Jurkat_ctrl_1-Tag-30_ATAC
GSM4706085 Jurkat_ctrl_2-Tag-30_ATAC
GSM4706086 Jurkat-stim-1_H3K27ac
GSM4706087 Jurkat-stim-2_H3K27ac
GSM4706088 Jurkat_stim_1-Tag-30_ATAC
GSM4706089 Jurkat_stim_2-Tag-30_ATAC
GSM4706090 THP-ctrl-2_H3K27ac
GSM4706091 THP1_ctrl_1-Tag-30_ATAC
GSM4706092 THP1_ctrl_2-Tag-30_ATAC
GSM4706093 THP-stim-1_H3K27ac
GSM4706094 THP-stim-2_H3K27ac
GSM4706095 THP_stim_1-Tag-30_ATAC
GSM4706096 THP_stim_2-Tag-30_ATAC
GSM4706097 THP1pmaLPS_K27ac_1_0h_H3K27ac
GSM4706098 THP1pmaLPS_K27ac_2_0h_H3K27ac
GSM4706099 THP_pmaLPS_ATAC_1_0h_ATAC
GSM4706100 THP_pmaLPS_ATAC_2_0h_ATAC
GSM4706101 THP1pmaLPS_K27ac_19_120h_H3K27ac
GSM4706102 THP1pmaLPS_K27ac_20_120h_H3K27ac
GSM4706103 THP_pmaLPS_ATAC_19_120h_ATAC
GSM4706104 THP_pmaLPS_ATAC_20_120h_ATAC
GSM4706105 THP1pmaLPS_K27ac_10_12h_H3K27ac
GSM4706106 THP1pmaLPS_K27ac_9_12h_H3K27ac
GSM4706107 THP_pmaLPS_ATAC_10_12h_ATAC
GSM4706108 THP_pmaLPS_ATAC_9_12h_ATAC
GSM4706109 THP1pmaLPS_K27ac_3_1h_H3K27ac
GSM4706110 THP1pmaLPS_K27ac_4_1h_H3K27ac
GSM4706111 THP_pmaLPS_ATAC_4_1h_ATAC
GSM4706112 THP1pmaLPS_K27ac_11_24h_H3K27ac
GSM4706113 THP1pmaLPS_K27ac_12_24h_H3K27ac
GSM4706114 THP_pmaLPS_ATAC_11_24h_ATAC
GSM4706115 THP_pmaLPS_ATAC_12_24h_ATAC
GSM4706116 THP1pmaLPS_K27ac_5_2h_H3K27ac
GSM4706117 THP1pmaLPS_K27ac_6_2h_H3K27ac
GSM4706118 THP_pmaLPS_ATAC_5_2h_ATAC
GSM4706119 THP_pmaLPS_ATAC_6_2h_ATAC
GSM4706120 THP1pmaLPS_K27ac_13_48h_H3K27ac
GSM4706121 THP1pmaLPS_K27ac_14_48h_H3K27ac
GSM4706122 THP_pmaLPS_ATAC_13_48h_ATAC
GSM4706123 THP_pmaLPS_ATAC_14_48h_ATAC
GSM4706124 THP1pmaLPS_K27ac_7_6h_H3K27ac
GSM4706125 THP1pmaLPS_K27ac_8_6h_H3K27ac
GSM4706126 THP_pmaLPS_ATAC_7_6h_ATAC
GSM4706127 THP1pmaLPS_K27ac_15_72h_H3K27ac
GSM4706128 THP1pmaLPS_K27ac_16_72h_H3K27ac
GSM4706129 THP_pmaLPS_ATAC_15_72h_ATAC
GSM4706130 THP1pmaLPS_K27ac_17_96h_H3K27ac
GSM4706131 THP1pmaLPS_K27ac_18_96h_H3K27ac
GSM4706132 THP_pmaLPS_ATAC_17_96h_ATAC
GSM4706133 U937-ctrl-1_H3K27ac
GSM4706134 U937-ctrl-2_H3K27ac
GSM4706135 U937_ctrl_1-Tag-30_ATAC
GSM4706136 U937_ctrl_2-Tag-30_ATAC

GSM4706137 U937-stim-1_H3K27ac
GSM4706138 U937-stim-2_H3K27ac
GSM4706139 U937_stim_1-Tag-30_ATAC
GSM4706140 U937_stim_2-Tag-30_ATAC

Genome browser session (e.g. UCSC)

http://genome.ucsc.edu/cgi-bin/hgTracks?hubUrl=ftp://ftp.broadinstitute.org/outgoing/lincRNA/Nasser2020/UCSCHub.txt&genome=hg19

Methodology

Replicates	Biological duplicates
Sequencing depth	ChIP-Seq experiments were sequenced using 75bp single end reads to a depth of >30 million reads per sample. ATAC-Seq experiments were sequenced using 50bp paired end reads to a depth of >20 million reads per sample
Antibodies	H3K27ac monoclonal antibody (Cat #39685, Active Motif)
Peak calling parameters	The 'callpeak' function in MACS2 was used to call peaks for both ChIP-Seq and ATAC-Seq experiments using default parameters.
Data quality	Data quality was assessed by computing the aggregated TSS enrichment of each sample. For ATAC-Seq samples there were 60,000 - 90,000 peaks called at an FDR of 5%. For ChIP-Seq samples there were 25,000 - 60,000 peaks called at an FDR of 5%.
Software	We aligned reads using BWA (v0.7.17), removed PCR duplicates using the MarkDuplicates function from Picard (v1.731, http://picard.sourceforge.net), and filtered to uniquely aligning reads using samtools (MAPQ >= 30, https://github.com/samtools/samtools)

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	After stimulation and incubation of THP1 cells we harvested cells, resuspended in FACS buffer (0.4% BSA, PBS) and proceeded to mitochondrial staining (see methods).
Instrument	MA900 Multi-Application Cell Sorter
Software	FlowJo v10.7
Cell population abundance	For THP1 mitochondrial CRISPR screen experiments, unfragmented cells were gated using FSC/SSC and cells were split into three populations based on mitochondrial membrane potential (ratio of MitoTracker Red to MitoTracker Green, top 15%, middle 25%, bottom 15%), excluding a population with depolarized mitochondria (Figure S8).
Gating strategy	See previous

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