



Published in final edited form as:

Nature. 2021 May ; 593(7858): 238–243. doi:10.1038/s41586-021-03446-x.

Genome-wide enhancer maps link risk variants to disease genes

Joseph Nasser^{1,*}, Drew T. Bergman^{1,*}, Charles P. Fulco^{1,24,*}, Philine Guckelberger^{1,2,*}, Benjamin R. Doughty^{1,3,*}, Tejal A. Patwardhan^{1,4}, Thouis R. Jones¹, Tung H. Nguyen¹, Jacob C. Ulirsch^{1,5}, Fritz Lekschas⁶, Kristy Mualim³, Heini M. Natri³, Elle M. Weeks¹, Glen Munson¹, Michael Kane¹, Helen Y. Kang^{3,7}, Ang Cui^{1,8}, John P. Ray^{1,25}, Thomas M. Eisenhaure¹, Ryan L. Collins^{1,9,10}, Kushal Dey¹¹, Hanspeter Pfister⁶, Alkes L. Price^{1,11,12}, Charles B. Epstein¹, 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,†}, Jesse M. Engreitz^{1,3,7,†}

¹Broad Institute of MIT and Harvard, Cambridge, MA

²Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany

³Department of Genetics, Stanford University School of Medicine, Stanford, CA

⁴Department of Statistics, Harvard University, Cambridge, MA

⁵Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA

⁶Harvard School of Engineering and Applied Sciences, Cambridge, MA

⁷BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford University School of Medicine, Stanford, CA

⁸Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA

*These authors contributed equally as first authors

†These authors jointly supervised the work

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.

Code Availability

ABC Model: <https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>. This is the codebase used to generate ABC predictions for this manuscript, and can be used to run the ABC model on new biosamples.

ABC-Max and paper-specific analyses: <https://github.com/EngreitzLab/ABC-GWAS-Paper>. This repository implements the ABC-Max pipeline can be used to reproduce specific analyses in this study.

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.

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.

- ⁹Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA
- ¹⁰Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
- ¹¹Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA
- ¹²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
- ¹³Department of Computer Science, Stanford University, Stanford, CA
- ¹⁴Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA
- ¹⁵Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA
- ¹⁶Department of Molecular Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA
- ¹⁷Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA
- ¹⁸Department of Medicine, Harvard Medical School, Boston, MA
- ¹⁹Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
- ²⁰Center for Cancer Research, Massachusetts General Hospital, Boston, MA
- ²¹Department of Biology, MIT, Cambridge, MA (on leave)
- ²²Department of Systems Biology, Harvard Medical School, Boston, MA (on leave)
- ²³Office of Science and Technology Policy and Executive Office of the President, White House
- ²⁴Present address: Bristol Myers Squibb, Cambridge, MA
- ²⁵Present address: Systems Immunology, Benaroya Research Institute at Virginia Mason, Seattle, WA

Abstract

Genome-wide association studies (GWAS) have now identified thousands of noncoding loci associated with human diseases and complex traits, each of which could reveal insights into biological mechanisms of disease¹. Many of the underlying causal variants are thought to affect enhancers^{2,3}, but we have lacked accurate maps of enhancers and their target genes to interpret such variants. We previously developed the Activity-by-Contact (ABC) Model to predict enhancer-gene connections and demonstrated that it can predict the results of CRISPR perturbations across several cell types⁴. Here, we apply this ABC Model to create enhancer-gene maps in 131 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 via variants in enhancers that act in different cell types. For inflammatory bowel disease (IBD), causal variants are >20-fold enriched in predicted enhancers in particular cell types, and ABC outperforms other regulatory methods at connecting noncoding variants to target genes. Guided by these variant-to-function maps, we show that an enhancer containing an IBD risk variant regulates the expression of *PPIF* to tune mitochondrial membrane potential in macrophages. Together, our study reveals insights into principles of genome regulation, illuminates mechanisms that influence IBD, and

demonstrates a generalizable strategy to connect common disease risk variants to their molecular and cellular functions.

Each GWAS association could provide insight into a biological mechanism underlying human disease risk^{1,5}. Yet, identifying these mechanisms has proven 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 and instead occur in noncoding gene regulatory elements such as enhancers^{2,3}, which can influence gene expression over long distances^{6,7}. Finally, common diseases appear to involve contributions from multiple cell types, and many enhancers appear to act in specific cell types or states⁸. 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 signals⁹⁻¹¹, including identifying 93 noncoding credible sets for IBD⁹. To link noncoding variants to their target genes and cell types, we recently developed the Activity-by-Contact (ABC) Model to identify enhancers in a particular cell type and predict their target genes based on data about chromatin state and 3D contacts⁴. Together, these advances suggest a new approach to connect GWAS signals to their target genes and cell types.

Here, we build ABC enhancer-gene maps in 131 biosamples and apply these maps to analyze fine-mapped genetic variants associated with 72 diseases and complex traits (Extended Data Fig. 1). ABC maps link 5,036 GWAS signals to predicted functions, with improved accuracy compared to existing approaches. By tracing the path from variant to cell type to target gene, we nominate new regulatory mechanisms for IBD and identify genes likely to influence multiple diseases through effects in different cell types, including at the 10q22.3 IBD risk locus. Together, our study demonstrates a generalizable strategy to build regulatory maps of the genome to connect genetic associations to molecular mechanisms of disease.

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 Project⁸ and other sources (Supplementary Tables 1,2, Extended Data Fig. 1). For each biosample, we calculated ABC scores for each gene and chromatin accessible element within 5 Mb by multiplying estimates of enhancer activity and 3D enhancer-promoter contact frequency. Candidate element-gene pairs that exceeded a chosen threshold were defined as “enhancer-gene connections”, and elements predicted to regulate at least one gene 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 (~12% of chromatin-accessible regions, 0.11% of the mappable genome, Supplementary Table 2, Extended Data Fig. 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 comparison 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} plus additional CRISPRi-FlowFISH experiments we performed here; Supplementary Tables 3, 4). ABC performed well at classifying regulatory connections (area under the precision-recall curve (AUPRC) = 0.64), and outperformed other methods, similar to our previous observations using a subset of this CRISPR data⁴ (Extended Data Fig. 4, Supplementary Table 5).

Enrichment of GWAS variants in enhancers

To assess the utility 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 analyses⁹, we examined 24,922 fine-mapped variants with posterior inclusion probability (PIP) $\geq 10\%$ for 72 diseases and traits, focusing on credible sets that did not contain any coding or splice site variant (Methods, Extended Data Fig. 5a).

Fine-mapped GWAS variants showed striking 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, Extended Data Fig. 5b–d), and in some cases showed evidence of allele-specific H3K27ac signals (Methods).

For example, fine-mapped variants for IBD were significantly enriched in ABC enhancers in 65 biosamples (Fisher's exact test $P_{Bonferroni} < 0.001$; "enriched biosamples"), including 56 of the 66 biosamples corresponding to immune cell types/cell lines and gut tissue (Fig. 1a; Supplementary Table 6). The most enriched biosample showed 21-fold enrichment and corresponded to activated dendritic cells, which are known to play 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 $\geq 95\%$, compared to 7.5% of all common noncoding variants (Fig. 1b, 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 $\geq 95\%$ (Fig. 1c). Importantly, this analysis likely underestimates the proportion of causal variants residing in ABC enhancers because we still lack appropriate data for many relevant cell types. We anticipate that a majority of causal noncoding GWAS variants 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 \geq 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 *Iq32.1* IBD risk locus had been previously fine-mapped to identify two independent credible sets (Extended Data Fig. 1b)⁹. Both credible sets include noncoding variants with PIP \geq 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 known to be important for IBD¹⁴ (Extended Data Fig. 6a).

To evaluate ABC-Max and other previous predictions, we examined a set of 64 genes previously linked to IBD based on coding variants or evidence from experimental models¹⁴ (Supplementary Tables 8, 9). We analyzed 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 ~70% of metabolite GWAS loci to genes with plausible biochemical functions¹⁷. Connecting the lead variant to the closest gene correctly identified the known IBD gene for 30 of 37 credible sets (81% precision, Fig. 1c). A similar approach, “closest transcription start site (TSS)”, identified the known IBD gene in 27 of 37 cases (73% precision, Fig. 1c, Supplementary Note 2).

We next evaluated other approaches to connect regulatory variants to disease genes, including predictions based on eQTL signals^{18–21}, 3D contacts²⁷, gene set enrichment²², or other enhancer-gene maps^{23–30} (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 of 18 cases (94% precision, 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. While ABC-Max had lower recall than closest gene, the fraction of loci with a prediction will likely increase upon expanding the ABC maps to include additional relevant cell types in the gut.

Because this curated gene set may harbor 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). Most notably, assigning each credible set to the gene with the strongest ABC score (“ABC-Max”; precision = 94% for known IBD 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 (*e.g.*, 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 genes³¹, and CRISPR experiments have identified individual enhancers that regulate up to 8 genes in *cis*^{4,32}. Our observations are consistent with a model where, while variants often affect the expression of multiple genes, only a subset of these effects are likely relevant to disease (Supplementary Note 1)³³.

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 <1 Kb to 1.1 Mb (median: 13 Kb), and 1,139 of 5,036 predictions involved a gene that was not the closest (Fig. 2a).

These predictions provide a resource for identifying genes, pathways, and regulatory properties at 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, Supplementary Tables 10, 11). Many of these genes have previously reported functions in immunity and inflammation, and were enriched for genes in the interferon gamma pathway (6 genes; 12-fold enrichment), lymphocyte activation (11 genes; 7-fold enrichment), and regulation of transcription from RNA polymerase II promoter (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 22q13 IBD locus, which has been annotated as corresponding to *TAB1/MAP3K7IP1*^{34,35}. Here, ABC-Max linked variants in two independent credible sets to platelet derived growth factor beta (*PDGFB*) in mononuclear phagocytes (MNPs; *e.g.*, monocytes, macrophages, and dendritic cells), supporting a causal role for PDGF signaling in IBD³⁶ (Fig. 2c). We also identified intergenic

IBD risk variants linked to *LRRC32* and *RASL11A* (Supplementary Note 3, Extended Data Fig. 7), and variants located in introns of *ANKRD55* and *ZMIZ1* 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 etiology. 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 where a variant was predicted to act only in specific cell lineages or stimulated immune cell states (Extended Data Fig. 8a, b), and allowed grouping genes by cell type to improve the detection of enriched gene sets (Extended Data Fig. 8c). At one IBD locus (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 CRISPRi perturbations, we confirmed that this predicted enhancer regulates *IL6ST* in a T cell line but not in 3 other B-cell or monocytic cell lines (Extended Data Fig. 8d).

Such cell-type specific effects appeared to lead to cases where a single gene could affect multiple traits. For example, *IKZF1* encodes a transcription factor involved in several stages of hematopoietic differentiation, and was linked by ABC to IBD and 11 other traits via different variants in 18 credible sets, including variants associated with erythrocyte, monocyte, or neutrophil count that overlapped ABC enhancers in erythroblasts, monocytes, or CD34+ hematopoietic 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, Supplementary Table 12), and where the predicted variants overlapped ABC enhancers in different sets of biosamples. These 577 genes appeared to have complex enhancer landscapes: they had (i) more predicted ABC enhancer connections (median 466 across all cell types versus 261 for other genes), (ii) more ABC enhancer connections per cell type in which the gene was expressed (median 4.8 versus 3.3), and (iii) more surrounding noncoding sequence (median 301 Kb versus 128 Kb distance to the closest neighboring TSSs, independent of ABC predictions) (Fig. 3c, 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 control³⁷.

From association to function at 10q22.3

To explore how ABC maps might accelerate experimental studies to characterize individual GWAS loci, we examined the IBD risk locus at chromosome 10q22.3, where ABC prioritized an unexpected gene. A single high-probability variant (rs1250566, PIP = 19%), located in an intron of *ZMIZ1*, overlapped an ABC enhancer in several immune cell types, including MNPs (Fig. 4a,b). Although this locus has previously been annotated as

corresponding to *ZMIZ1*^{15,34,38}, 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 3D 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 IBD risk, we used CRISPRi-FlowFISH⁴ to perturb each of the 163 accessible elements in a 712 Kb region around *PPIF* in four human immune cell lines, with and without stimulation with appropriate immune ligands. We identified 14 enhancers that regulated *PPIF* expression in at least one of these conditions (Extended Data Fig. 10a,b, Supplementary Table 4). Only one of these 14 enhancers contained a fine-mapped IBD variant (the enhancer 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 t-test $P < 10^{-111}$) (Fig. 4c, Extended Data Fig. 10b–e).

PPIF encodes cyclophilin D, a ubiquitously expressed protein that regulates metabolism, reactive oxygen species signaling, and cell death by controlling the mitochondrial permeability transition and mitochondrial membrane potential (ψ_m)³⁹. Accordingly, we tested whether the *PPIF* enhancer containing the IBD variant might tune ψ_m in THP1 cells. We infected cells with a pool of CRISPRi gRNAs targeting the *PPIF* enhancer and promoter, stained cells with MitoTracker Red (a fluorescent dye with signal that increases with ψ_m), sorted cells into 3 bins based on their level of fluorescence, and sequenced the gRNAs in each bin to infer their effects on ψ_m (Extended Data Fig. 10f). CRISPRi targeting of the *PPIF* enhancer or promoter indeed increased ψ_m in THP1 cells in LPS-stimulated, but not unstimulated, conditions (Fig. 4d, 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 responding to inflammatory stimuli. Notably, changes in ψ_m have been previously linked to inflammatory responses in macrophages⁴⁰, suggesting a path by which tuning *PPIF* expression could affect IBD risk.

Interestingly, *PPIF* has an extremely complex enhancer landscape (top 0.3% of genes with the most ABC enhancer-connections, Fig. 3c), and the *PPIF* locus also harbors 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 work creates genome-wide maps of >6 million enhancer-gene connections that illuminate the functions of disease variants. We leverage these maps to identify new genes and pathways for IBD, nominate genes that control different traits through effects in different cell types, and identify 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 hematopoietic progenitors to influence risk for clonal

hematopoiesis⁴¹. By dramatically 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). (i) ABC does not perfectly predict the effects of distal enhancers, and does not capture other types of transcriptional or post-transcriptional regulatory elements. (ii) Many of these ABC maps involve analysis of epigenomic data from a single individual and therefore miss enhancers present only in certain genotypes or environments. (iii) Assessing the performance of gene predictions requires good sets of gold-standard genes, which remain limited and may contain biases (*e.g.*, toward the closest gene, or toward genes that tolerate coding variation). (iv) ABC-Max assumes a single causal gene per variant, but enhancers containing disease variants often appear to have highly pleiotropic effects. Systematic experimental studies will be required to explore whether some variants act through effects on multiple genes and cell types.

In summary, our approach illuminates a path toward creating a comprehensive map of enhancer regulation in the human genome. By refining computational models such as ABC and collecting the needed 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 would provide a foundational reference for identifying disease 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.

Methods

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 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).

Cell culture.—We maintained cells at a density between 100K and 1M per ml (250K–1M per ml for GM12878) in RPMI-1640 (Thermo Fisher Scientific, Waltham, MA) with 10% heat-inactivated FBS (15% for GM12878, HIFBS, Thermo Fisher Scientific), 2mM L-glutamine, and 100 units/ml streptomycin and 100 mg/ml penicillin by diluting cells 1:8 in fresh media every three 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 anti-CD40 (Invitrogen-140409-82) and 10 µg/ml anti-IgM (Sigma-I0759) for 4 hours. We stimulated Jurkat cells with 5 µg/ml anti-CD3 (Biolegend-317315) and 100 ng/ml phorbol 12-myristate 13-acetate (PMA, Sigma-P1585) for 4 hours. We stimulated THP1 cells with 1 µg/ml bacterial lipopolysaccharide (LPS) from

E. coli K12 (LPS-EK Invivogen trl-pek1ps) for 4 hours. We stimulated U937 cells with 200 ng/ml LPS for 4 hours.

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

Epigenomic profiling of immune cell lines

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

ATAC-seq.—We applied ATAC-seq as previously described⁴², with modifications. Briefly, we washed 50,000 cells once with 50 µl of cold 1x PBS and added 50 µl of Nuclei Isolation EZ Lysis Buffer (SIGMA NUC101-1KT) to resuspend gently, immediately centrifuging at 500xg for 10 minutes 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 500xg for 5 minutes at 4°C and re-decant 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 signal-to-noise) and incubated the mix at 37°C for 30 minutes in a PCR block. Immediately following 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 backup, we proceeded with one by adding 1.8X SPRI beads (Agencourt A63881), waiting 5 minutes for the DNA to associate to the beads, and then washing the beads twice using 80% EtOH. We then eluted the DNA from the beads using 10 µl of water and added to it 25 µl NEBNext HiFi 2x PCR MasterMix (NEB M0541), with 2.5 uL of each of the dual-indexed Illumina Nextera primers (25 µM). We amplified the PCR reaction to 15 cycles, as previously described. We purified amplified libraries and removed adapters using two clean-ups with 1.8x 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 described³².

H3K27ac ChIP-seq.—We generated and analyzed ChIP-seq data from 5 million cells in each cell line and stimulation state, following protocols previously described⁴³. Before harvesting for ChIP-seq, cells at 1 million cells per mL were replenished by a 1:2 (v/v) split in fresh media and allowed to grow for 4 hours. 10 million cells were harvested from

each cell type at 500K cells/mL and washed 2x in cold PBS. Cells were resuspended in warm PBS with 1% formaldehyde (Cat #28906, Thermo Scientific) and incubated at 37°C for 10 minutes. Crosslinking was quenched by adding glycine to a concentration of 250 mM and incubating for 5 minutes at 37°C. Cells were placed on ice for 5 minutes, then washed 2x 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% NP40) and incubated on ice for 10 minutes. The nuclear pellet was isolated by spinning the cell lysis mix at 5,600xg at 4°C for 3.5 minutes and discarding the supernatant. Nuclear pellets were lysed by adding 1 mL nuclear lysis buffer (10 mM Tris-HCl pH 7.5 ml, 1% NP-40 alternative (CAS 9016-45-9), 0.5% Na Deoxycholate, 0.1% SDS) with protease inhibitors on ice for 10 minutes. The chromatin-containing nuclear lysate was sonicated 3x using a Branson sonifier (ON 0.7s, OFF 1.3s, TIME 2 minutes, WATTS 10–12), with 1 minute rest between sonifications. Sonicated chromatin was spun down at maximum speed. 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, and 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS). To immunoprecipitate H3K27ac, 3 µL of H3K27ac monoclonal antibody (Cat #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 (Cat #10008D, Invitrogen) and Protein G Dynabeads magnetic beads (Cat #10004D, Life Technologies) were washed with blocking buffer (PBS, 0.5% Tween20, 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% Na 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% Na Deoxycholate), twice with 200 µL LiCl buffer (250 mM LiCl, 0.5% NP40, 0.5% Na 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 cross-linking enzyme mix (250 mM Tris-HCl pH 6.5, 62.5 mM EDTA pH 8.0, 1.25 M NaCl, 5 mg/ml Proteinase K (Cat #25530-049, Invitrogen), 62.5 µg/ml RNase A (Cat #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 (Cat #A63880, Beckman Coulter) with a 2x bead:sample ratio. Sequencing libraries were prepared with KAPA Library Preparation kit (Cat #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 >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 July 17, 2017⁴⁴. We selected 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 July 12, 2017⁴⁵.

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 where 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 (*e.g.*, 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 (v0.7.17)⁴⁶, removed PCR duplicates using the MarkDuplicates function from Picard (v1.731, <http://picard.sourceforge.net>), and filtered to uniquely aligning reads using samtools (MAPQ \geq 30, <https://github.com/samtools/samtools>)⁴⁷. The resulting BAM files were used as inputs into the ABC Model.

Activity-by-Contact model predictions

We used the Activity-by-Contact (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) and histone modifications (H3K27ac ChIP-seq), as previously described⁴. 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.

Briefly, for each cell type, we:

1. Called peaks on the chromatin accessibility dataset using MACS2 with a lenient p-value cutoff of 0.1

2. 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 centered on the peak summit. To this list we added 500 bp regions centered 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*.
3. 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.
4. Calculated element-promoter Contact using the average Hi-C signal across 10 human Hi-C datasets as described below.
5. 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 KR normalized Hi-C matrices for 10 human cell types⁴. For each cell type we:

1. Transformed the Hi-C matrix for each chromosome to be doubly stochastic.
2. We then replaced the entries on the diagonal of the Hi-C matrix with the maximum of its four neighboring bins.
3. We then replaced all entries of the Hi-C matrix with a value of NaN or corresponding to KR normalization factors < 0.25 with the expected contact under the power-law distribution in the cell type.
4. We then scaled the Hi-C signal for each cell type using the power-law distribution in that cell type as previously described.
5. We then 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 here: 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 (*e.g.*, $R^2 = 0.91$ 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 (see Fulco et al 2019⁴). We also find here that using averaged Hi-C data performs similarly to using cell-type specific promoter capture Hi-C data (Extended Data Fig. 4d).

Promoter Capture Hi-C—In some evaluations of the performance of the ABC model to CRISPR data (Extended Data Fig. 4e–h), we used ABC predictions where 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:

1. We downloaded PC-HiC raw count data from the BLUEPRINT consortium⁴⁹.
2. Contacts from restriction fragments which overlap baited promoter regions were linearly adjusted based on the total number of detected contacts for the baited region(s).
3. We re-binned the data from restriction fragment sites to 5kb resolution.
4. 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 whose promoters are “active” (*i.e.*, where 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 described⁴: <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 where 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

For intersecting ABC predictions with variants, we took the predictions from the ABC Model and applied the following additional processing steps: (i) We considered all distal element-gene connections with an ABC score ≥ 0.015 (see Extended Data Fig. 4; lower threshold than our previous study⁴ 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 data⁴). (ii) We shrunk the ~500-bp regions by 150-bp on either side, resulting in a ~200-bp region centered on the summit of the accessibility peak. This is because, while 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 region⁵⁰. In practice, this adjustment does not

substantially affect the enrichment of fine-mapped IBD variants (Extended Data Fig. 5d). (iii) We included enhancer-gene connections spanning up to 2 Mb — greater than the maximum distance of the longest-range enhancer-gene connection we have identified in CRISPR experiments to date (~1.8 Mb).

CRISPRi-FlowFISH

We applied CRISPRi-FlowFISH to very sensitively test the effects of distal elements on gene expression⁴. Briefly, CRISPRi-FlowFISH involves targeting putative enhancers with many independent guide RNAs (gRNAs; median = 45) in a pooled screen using CRISPR interference (CRISPRi), which alters chromatin state via 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: (i) use fluorescence in situ hybridization (FISH, Affymetrix PrimeFlow assay) to quantitatively label single cells according to their expression of an RNA of interest; (ii) sort labeled cells with fluorescence-activated cell sorting (FACS) into 6 bins based on RNA expression; (iii) use high-throughput sequencing to determine the frequency of gRNAs from each bin; and (iv) compare the relative abundance of gRNAs in each bin to compute the effects of gRNAs on RNA expression. CRISPRi-FlowFISH provides ~300 bp resolution to identify regulatory elements; has power to detect effects of as low as 10% on gene expression; and provides effect size estimates that match those observed in genetic deletion experiments⁴.

Here, we applied CRISPRi-FlowFISH to comprehensively test all putative enhancers in a ~700-Kb region around *PPIF*, and to validate additional selected enhancers (for 12 additional genes) that contained variants 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 (+/- 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 described⁴. 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 EF1 α promoter (ClonTech, Mountain View, CA), as previously described¹². For each, we sorted polyclonal populations with high BFP expression upon addition of doxycycline. For GM12878 cells, we used an alternative lentiviral construct to express the rtTA with a hygromycin resistance gene, as GM12878 appeared resistant to selection with neomycin/G418.

We performed CRISPRi-FlowFISH using ThermoFisher PrimeFlow (ThermoFisher 88-18005-210) as previously described, using the probesets listed in Supplementary Table 13. To ensure robust data, we only included probesets with twofold signal over unstained cells, and required an uncorrected knockdown at the TSS of >20%. We analyzed these data as previously described⁴. Briefly, we counted gRNAs in each bin using Bowtie⁵¹ 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 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 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 positives are the experimentally measured element-gene pairs which are statistically significant and for which perturbation of the element resulted in a decrease in gene expression. For these comparisons, (i) we only considered experimentally tested elements in which the element is not within 500bp of an annotated gene transcription start site; (ii) for perturbations using CRISPRi we excluded pairs in which the element resides within the gene body of the assayed gene; (iii) we excluded non-significant pairs for which the power to detect a 25% change in gene expression was < 80%; and (iv) 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 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. 4e). We calculated the area under the precision-recall curve (AUPRC)

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 in 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. In order 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 ~500bp regions in one replicate, but merge them into a ~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 ~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 cutoff 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 (~200bp, see above) and considered two connections to be shared if the elements overlapped 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 (CD), and ulcerative colitis (UC) (European ancestry only)⁵² from <https://www.ibdgenetics.org/downloads.html>. We obtained fine-mapping posterior probabilities and credible sets from Huang et al.⁹, and analyzed the top two conditionally independent credible sets in each locus. We also analyzed variants from IBD GWAS loci that were not fine-mapped in this study^{52,53}; for each such locus, we analyzed all 1000 Genomes variants in LD with the lead variant ($r^2 > 0.2$) and weighted each variant evenly (probability = 1 / number of variants in LD). 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 CD, UC, or both.

We obtained fine-mapping results and summary statistics for 73 other traits based on an unpublished analysis (Jacob Ulirsch, Masahiro Kanai, and Hilary Finucane) that analyzed data from the UK Biobank (Application #31063; fine-mapping 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 (MAF) > 0.01%, and Hardy-Weinberg equilibrium (HWE) p-value > 1e-10 were included in the GWAS. Covariates for the top 20 PCs, sex, age, age², sex*age, sex*age², and dilution factor where applicable were controlled for in the association studies. Quantitative traits were inverse rank transformed and associations were estimated using BOLT-LMM⁵⁴ for quantitative traits and SAIGE⁵⁵ for binary traits. In-sample dosage LD was computed using LDStore⁵⁶, and phenotypic variance was computed empirically. Fine-mapping was performed using the Sum of Single Effects (SuSiE) method⁵⁷, allowing for up to 10 causal variants in each region. Prior variance and residual variance were estimated using the default options, and single effects (potential 95% CSs) were pruned using the standard purity filter such that no pair of variants in a CS 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 LD structure in the human population), and overlapping regions were merged. Variants in the MHC region (chr6: 25–36 Mb) were excluded as were 95% CSs containing variants with fewer than 100 minor allele counts. Coding (missense and predicted loss of function) variants were annotated using the Variant Effect Predictor (VEP) version 85⁵⁸. 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 analyzed 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” — *i.e.*, 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 24/06/2017)⁵⁹. 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 1000 Genomes as described in the S-LDSC 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 analyzed each of the previous studies from Fig. 1c reporting cell-type specific enhancer-gene predictions, and also ChromHMM enhancers in blood cells downloaded from the BLUEPRINT Project^{60,61}.

Stratified linkage disequilibrium score regression (S-LDSC)

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 genome⁶². We analyzed the ABC enhancer regions as defined above, and ran LD 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 1000 Genomes with minor allele count >5 in 379 European samples). For comparison, we also analyzed 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 24/06/2017): coding sequences (CDS), 5' and 3' untranslated regions (UTR) 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 (*e.g.*, 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 analyzed allele-specific signals called by the WASP method as previously described⁶³. 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 Signaling 1 (SOCS1) located 93 Kb away) and for rs9808651 in monocytes (linked

by ABC to ERG, located 32Kb away). This analysis indicates that some prioritized causal variants have allelic effects on enhancer activity.

Evaluating gene prediction methods

Curated genes for inflammatory bowel disease.—We analyzed a list of IBD disease genes curated by Graham and Xavier (2020).¹⁴ To evaluate methods to connect noncoding GWAS variants to genes, we analyzed 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 where the gene was curated based on evidence from coding variation, we examined nearby conditionally independent noncoding signals, which might 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 Database⁶⁴. We analyzed 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 5 gene sets with the highest enrichment that also had at least five identified genes and hypergeometric test P -value $< 10^{-4}$. We plotted a 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 $\geq 50\%$) 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 (VEP)⁵⁸ 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 where exactly 1 gene within 1 Mb carried such a coding variant, and where 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 genes.

eQTL Colocalization (Open Targets Platform).—[OpenTargets.org](https://www.opentargets.org) performed colocalization analysis between IBD GWAS signals^{52,53} and eQTLs and 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 Catalogue⁶⁵. We downloaded colocalization results from ftp://ftp.ebi.ac.uk/pub/databases/opentargets/genetics/190505/v2d_coloc/ on February 1, 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).—Chun *et al.* tested colocalization of IBD GWAS signals with eQTLs in CD4+ T cells, CD14+ monocyte, and LCLs¹⁸. We obtained their colocalized genes from Table 2. We used the JLIM p-value to rank genes within each locus.

TWAS (S-PrediXcan and multiXcan).—Barbeira *et al.* developed multiXcan and compared GTEx v7 eQTLs to IBD summary statistics²⁰. We downloaded Dataset 6 and compared genes within each locus using the multiXcan p-value.

Mendelian randomization.—Hauberg *et al.* used a Mendelian randomization based approach (SMR) to connect IBD GWAS signals to effects on gene expression using eQTL data from 24 tissues²¹. We downloaded Table S3 and defined predicted genes in any tissue. We used the SMR false discovery rate to rank genes within each locus.

COGS.—Javierre *et al.* (2016) used promoter-capture Hi-C data in many blood cell types to link GWAS variants to target genes⁴⁹. We downloaded Table S3 (Tab 2) and analyzed genes linked with COGS scores ≥ 0.5 .

In all cases, we combined predictions of disease genes for IBD, UC, and CD.

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 (i) their ability to predict enhancer-gene regulation in CRISPR datasets (Extended Data Fig. 4) and (ii) their ability to identify IBD genes (Fig. 1c, Extended Data Fig. 6b). For the latter analysis, we used the predictions from each method to overlap fine-mapped variants (PIP $\geq 10\%$) with enhancers in any cell type and assigned variants to the predicted gene(s).

Promoter-capture Hi-C.—We downloaded Data S1 peak data from Javierre *et al.* (2016)⁴⁹, representing promoter-capture Hi-C data from 9 hematopoietic 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.

DHS-promoter correlation (ENCODE2).—Thurman *et al.* (2012) linked distal accessible elements with gene promoters by looking at correlation of DNase I hypersensitivity across 125 cell and tissue types from ENCODE²⁸. 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_32celltypeCatego

[ries.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).—Andersson *et al.* (2014) linked transcriptional activity of enhancer and transcription start sites using the FANTOM5 CAGE expression atlas²⁵. We downloaded these predictions from http://enhancer.binf.ku.dk/presets/enhancer_tss_associations.bed.

Enhancer-gene correlation (Ernst Roadmap).—Liu, Ernst *et al.* (2017) correlated gene expression with five active chromatin marks (H3K27ac, H3K9ac, H3K4me1, H3K4me2, and DNase I hypersensitivity) across 56 biosamples, and then used these correlation links to make predictions for the predicted enhancers (regions with the “7Enh” ChromHMM state) in 127 biosamples from the Roadmap Epigenome Atlas^{23,45}. We downloaded these predictions from www.biolchem.ucla.edu/labs/ernst/roadmaplinking and made predictions using the “confidence score”.

Enhancer-gene correlation (Granja single-cell RNA and ATAC-seq).—Granja *et al.* (2019) analyzed single-cell ATAC-seq and RNA-seq data in peripheral blood and bone marrow mononuclear cells, CD34+ bone marrow cells, and cancer cells from leukemia patients, and correlated ATAC-seq signal in accessible elements with the expression of nearby genes²⁴. We downloaded these predictions from <https://github.com/GreenleafLab/MPAL-Single-Cell-2019>, and used the correlation in healthy samples as the quantitative score. Cell-type specific links were not reported.

EnhancerAtlas 2.0.—Gao *et al.* (2020) used *EAGLE* to predict enhancer-gene interactions across a number of human tissues and cell lines²⁹. 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).—Sheffield *et al.* (2013) correlated DNase I signal and gene expression levels using data from 112 human samples representing 72 cell types to identify regulatory elements and to predict their targets²⁶. 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.—Cao *et al.* (2017) computed correlations between gene expression and various enhancer features (*e.g.*, DNase1, H3K4me1) across multiple cell types to identify a set of putative enhancers²⁷. 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.—Whalen et al. 2016 built a model to predict whether nearby enhancer-promoter pairs are located at anchors of Hi-C loops based on chromatin features³⁰. 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 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 to two previous GWAS gene prediction methods:

MAGMA.—We applied MAGMA⁶⁶ to the summary statistics for IBD⁵² 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 chi-square 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 (LD) among SNPs within the gene body. We assigned each IBD locus to the gene with the maximum positive *z*-score.

DEPICT.—We applied DEPICT, which leverages pathway analysis and cell-type enrichment analysis from gene expression datasets to analyze genome-wide significant loci and prioritize causal genes²². We applied DEPICT to the summary statistics for each trait using the 1000 Genomes Project reference panel and DEPICT’s 14,461 reconstituted gene sets 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 LD > 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 7 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 hematopoietic 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 Database⁶⁴, 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. (*i.e.*, 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, where genes linked to multiple traits via 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 (GCGGCCGAGCGGCTTCCCGT and GAACCTGGGCAAGCCAATAA) and e-PPIF (GACTCAAGATAACCACCACCGG and GATGGCCAGTTTGGGAACGT), along with four non-targeting control guides (GAGATGAAAGCGCAGCTAGGG, GGGCGCTTACGCGCGGGCCG, GCGCGCGCTAACTGGCGCTA, GATGTGTTGTAACCTCCACT), were cloned into sgOpti as previously described¹². We generated stable cell lines expressing each sgRNA by lentiviral transduction in 8 µg/ml polybrene by centrifugation at $1200 \times g$ for 45 minutes with 200,000 CRISPRi THP1 cells in 24 well plates. After 24 hours, we selected for transduction with 1 µg/ml puromycin (Gibco) for 72 hours then maintained cells in 0.3 µg/ml puromycin. We plated sgRNA-expressing stable cell lines at 100,000 cells/ml in 1 µg/ml doxycycline and harvested cells 48 hours 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, Lexington, MA) and random nonamer primers to convert RNA to cDNA. We performed qPCR using SYBR Green I Master Mix (Roche) with primers for PPIF (AGAACTTCAGAGCCCTGTGC, CATTGTGGTTGGTGAAGTCC) and GAPDH (AGCACATCGCTCAGACAC, GCCCAATACGACCAAATCC) and calculated differences using the $\Delta\Delta CT$ 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, Inc.; 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 10M cells per replicate with 1 µg/mL doxycycline. After 44 hrs, we added 1 µg/mL LPS and harvested cells for staining 4 hrs later. For the PMA LPS condition, we plated 10M cells per replicate and added 1µg/mL

doxycycline for 48 hrs. To differentiate into macrophage-like cells, we added fresh media with 20 ng/mL PMA and 1 μ g/mL doxycycline for an additional 24 hrs, confirming that cells adhered to the tissue culture plate. We washed out the PMA and added fresh media with 1 μ g/mL doxycycline and incubated cells for 45 hrs to recover and further differentiate cells. We then added 100 ng/mL LPS for 3 hrs, harvested cells, washed 3x with cold PBS, and proceeded to mitochondrial staining.

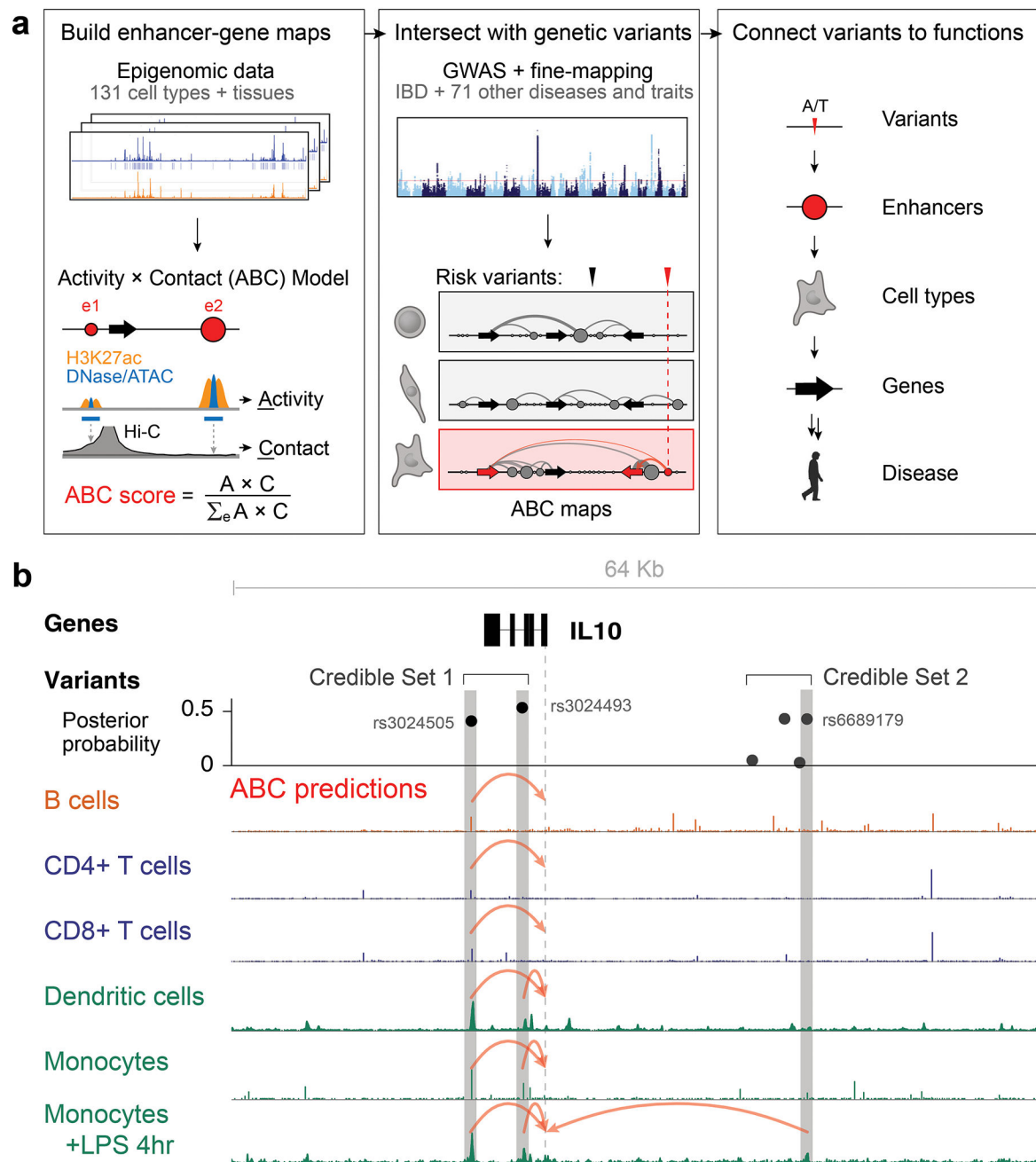
We stained cells with MitoTracker Red (200nM, Thermo Fisher, M7512) and MitoTracker Green (200nM, Thermo Fisher, M7514) according to the manufacturer's protocol and sorted cells into 3 bins according to their ratio of MitoTracker Red (which stains mitochondria dependent on ψ_m) to MitoTracker Green (which stains mitochondria independent of ψ_m), excluding a small population of depolarized cells with very low ψ_m (Extended Data Fig. 10f). We extracted genomic DNA and amplified and sequenced gRNAs from cells in each bin as previously described⁴.

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 HiGlass⁶⁷, a flexible genome browser toolkit: <https://flekschas.github.io/enhancer-gene-vis/> (Supplementary Fig. 2). The application features three linked views: the enhancer view at the top left, the gene view at 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.

Extended Data

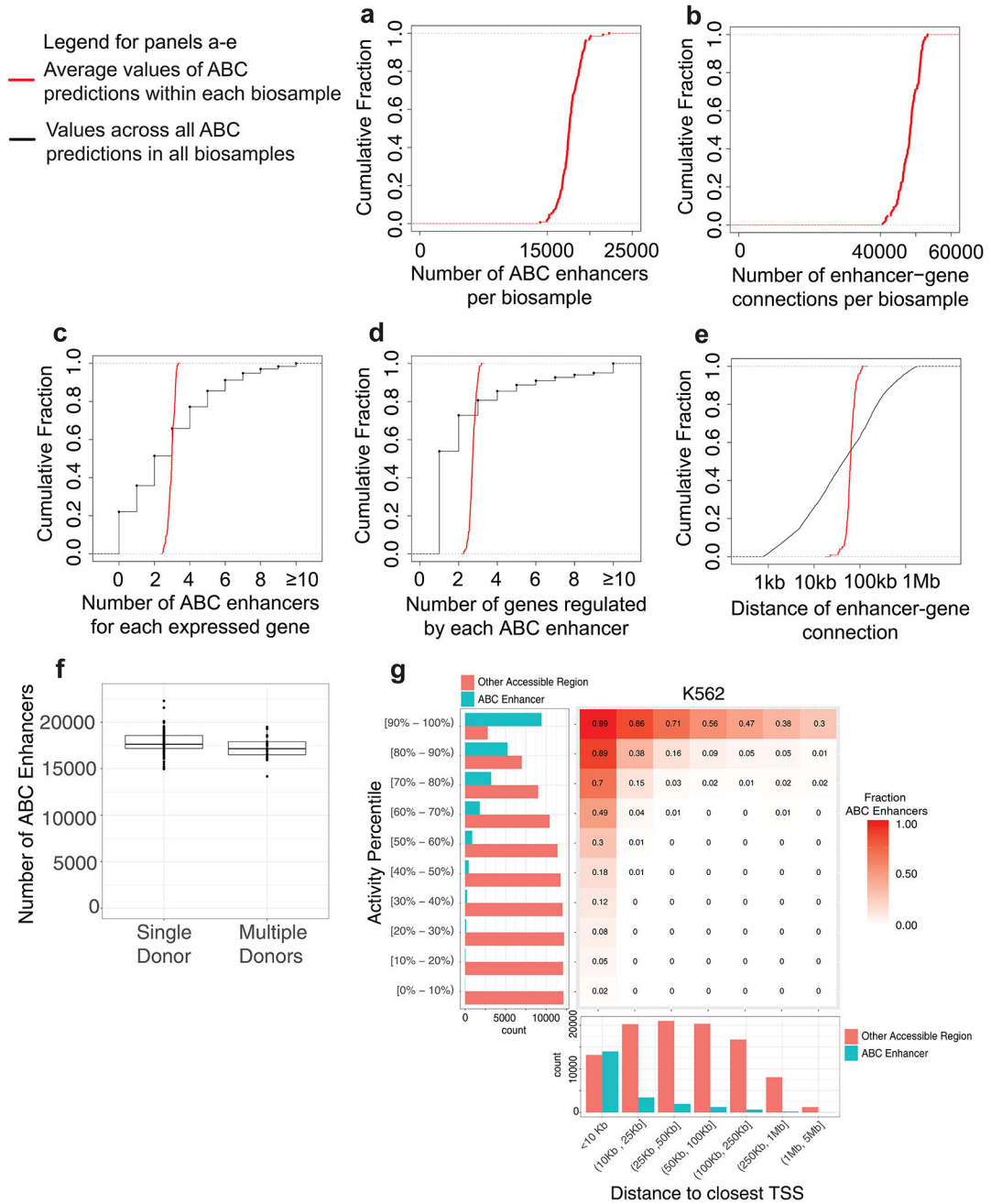


Extended Data Fig. 1. ABC maps connect fine-mapped variants to enhancers, genes, and cell types.

(a) Overview of approach.

(b) ABC predictions connect two IBD GWAS signals to *IL10*. Signal tracks show DNase- or ATAC-seq (based on availability of data). Red arrows represent ABC predictions connecting variants to *IL10*. Dashed line shows transcription start site (TSS). Gray bars highlight fine-mapped variants that overlap ABC enhancers in at least one cell type. Credible set 1 contains two variants, both of which overlap enhancers predicted to regulate *IL10* in various

cell types. Credible set 2 contains four variants, one of which overlaps an enhancer predicted to regulate *IL10* in monocytes stimulated with LPS.



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).

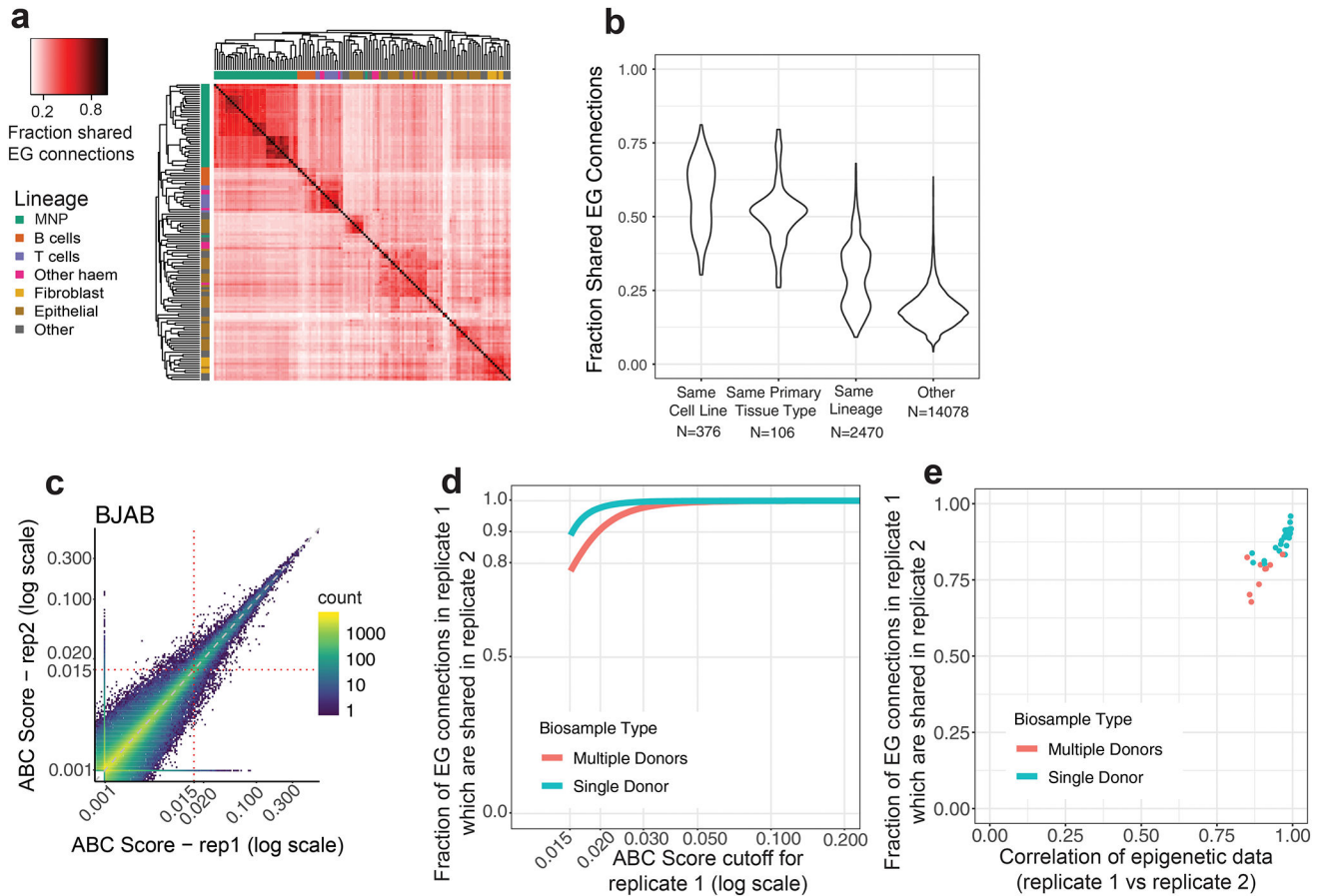
(c) Cumulative fractions of the number of enhancers predicted to regulate each gene across all biosamples (black line, median = 2, mean = 2.8) and the mean number of enhancers predicted to regulate each gene within each biosample (red line, median = 2.8).

(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 = 62,929bp) 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 p-value = 0.10). Boxplot displays median, 25th and 75th percentiles.

(g) Summary of ABC predictions in K562. Plot includes 122,410 non-promoter 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. Horizontal axis represents distance from the element to the closest transcription start site (TSS) of an expressed gene. Vertical axis represents the percentile bin of the Activity of the element (in terms of DHS and H3K27ac signals) among these 122,410 elements. The coloring of the heatmap 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

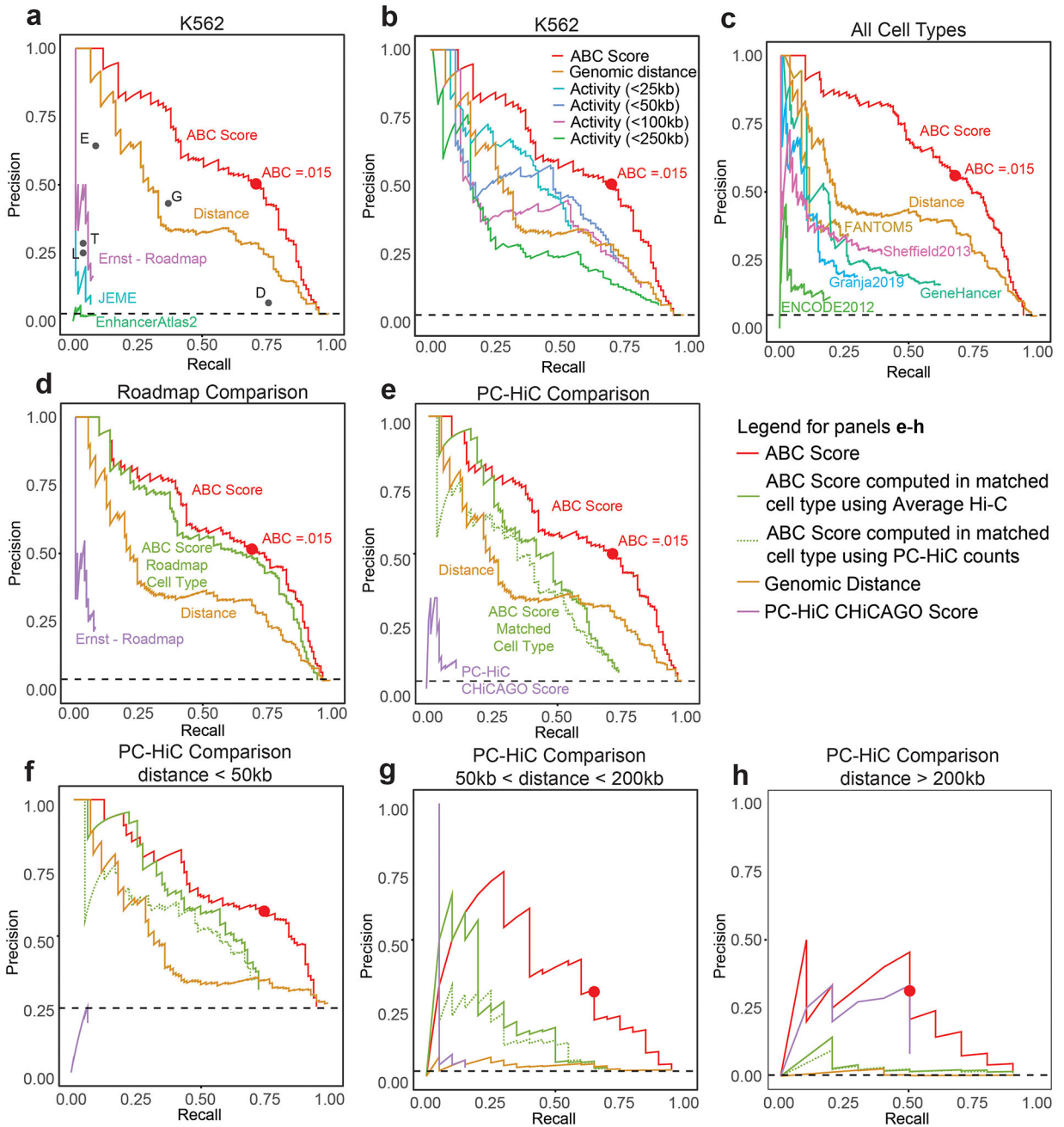
(a) Distinctness of predictions across biosamples. Biosample vs. biosample (131 × 131) heatmap. The color of the (*i,j*) pixel in the heatmap represents the fraction of enhancer-gene connections (‘EG connections’ – which are defined to be an element-gene pair whose ABC Score is greater than 0.015) in 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 cutoff increases. X-axis: Cutoff on the ABC Score. Y-axis: For a given cutoff of the ABC Score, the fraction of element-gene pairs with an ABC score greater than the cutoff 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 EG connections with ABC Score > 0.015 in replicate 1 which also have ABC Score > 0.015 in replicate 2. Colors as in (d)



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 transcription start site, (G) each element is predicted to regulate only the nearest (to TSS) expressed gene, (T) TargetFinder method³⁰, (L) elements and genes at opposite ends of HiCCUPS loops derived from Hi-C data are predicted as a connection⁶⁸, (D) an element-gene pair is a predicted positive if and only if the element and the gene are contained within the same contact domain⁶⁸. 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 is a critical component of the ABC model.

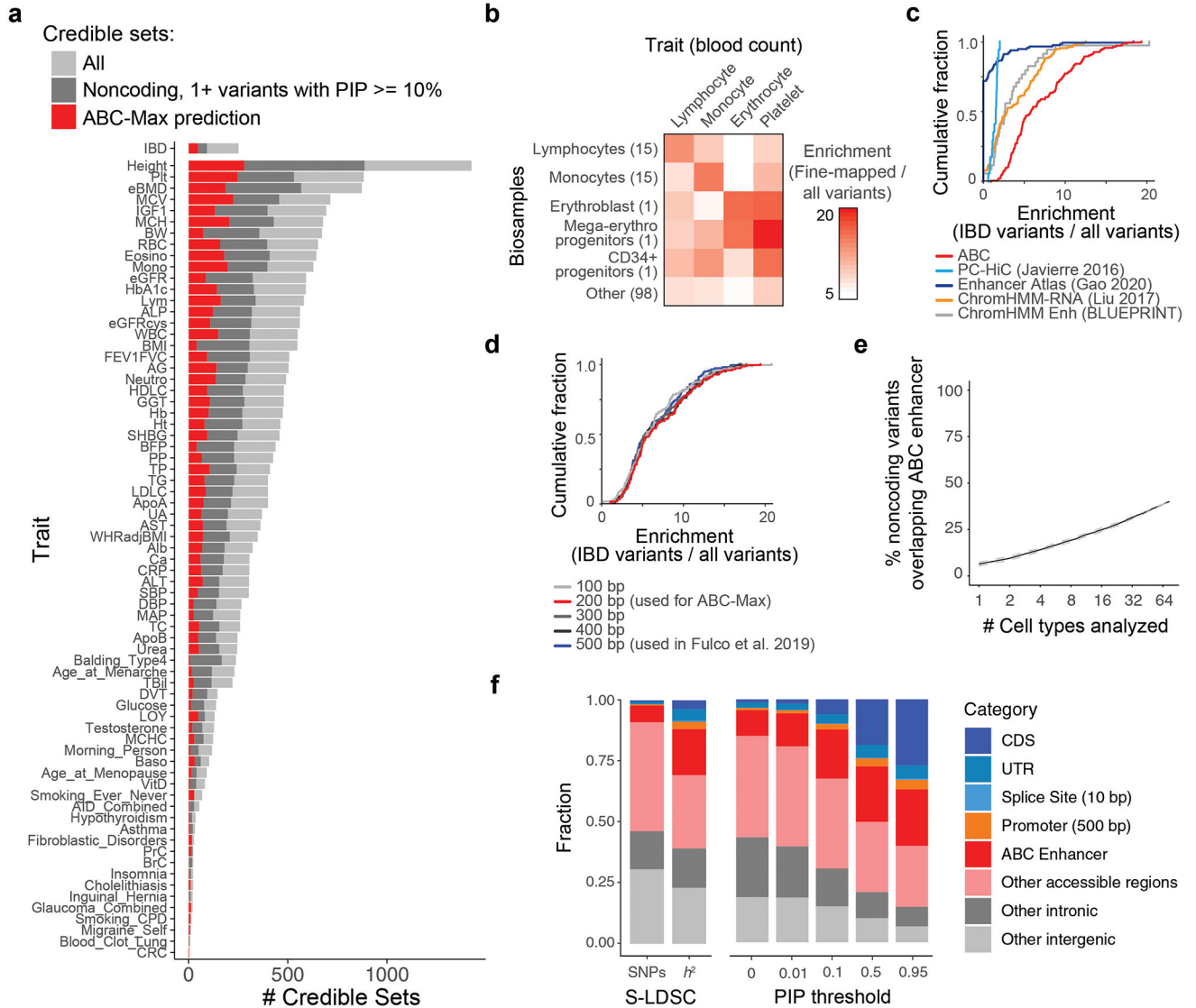
(c) Comparison of ABC and other enhancer-gene predictors in full CRISPR dataset. Comparison of enhancer-gene predictors to experimental CRISPR data in K562, GM12878, NCCIT, BJAB (+/- stimulation), Jurkat (+/- stimulation), THP1 (+/- stimulation) cells and primary hepatocytes. For ABC, we used the predictions 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. Comparison of enhancer-gene predictors to experimental CRISPR data in K562, GM12878, and unstimulated Jurkat, BJAB, THP1 cells. Red line represents 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 was compared to ABC Scores computed using epigenetic data from the Roadmap B cell sample (E032).

(e) Comparison of ABC to Promoter-Capture Hi-C. Comparison of enhancer-gene predictors to experimental CRISPR data in K562 and unstimulated BJAB, THP1 and Jurkat cells. Red line represents comparison of ABC Scores computed using epigenetic data from the same cell type as the CRISPR experiment was performed. To compare promoter-capture Hi-C CHiCAGO predictions (purple line) to CRISPR data, we made cell type substitutions because PC-HiC data is not available in 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 total 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 where the contact component is derived from the average Hi-C dataset used throughout this study. The dashed green line represents ABC scores where the contact component is derived from the raw counts in PC-HiC experiments (see Methods).

(f-h) Comparison of ABC to Promoter-Capture 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

transcription start site. Of the 4078 element-gene connections in the experimental dataset, 398 are at a distance of <50kb (of which 94 are experimental positives, 24% positive rate), 1102 are between 50kb and 200kb (20 positives, 2% positive rate), and 2578 are at a distance of >200kb (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 PR curves within each stratification, but it is not appropriate to compare the PR 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 analyzed for 72 diseases and complex traits. Light gray shows total number of fine-mapped credible sets. Dark gray shows number of such credible sets with no coding or splice site variants, and at least one variant with PIP \geq 10%. Red shows number of credible sets for which ABC-Max makes a prediction (*i.e.*, a variant with PIP \geq 10%).

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.

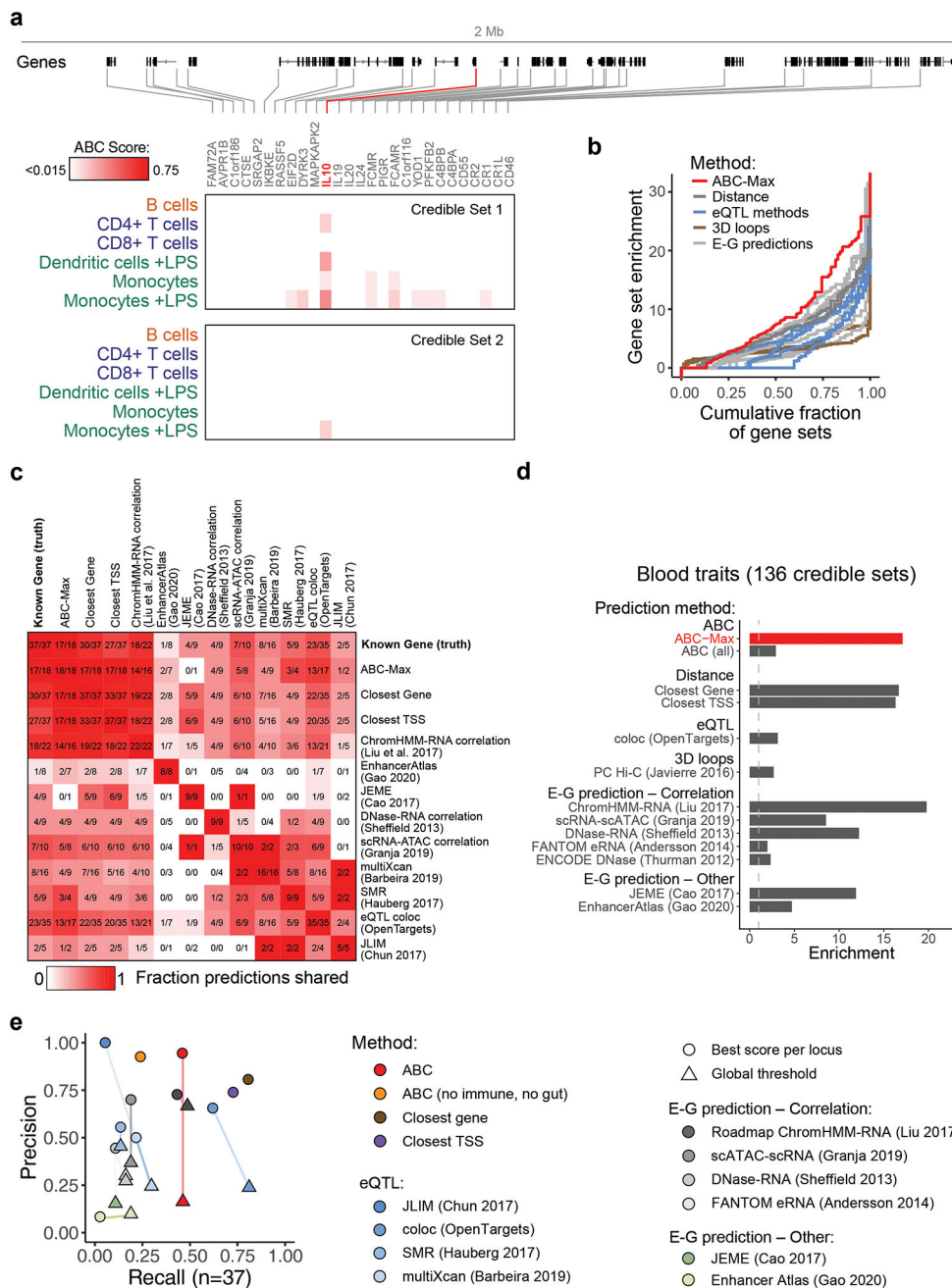
(b) Enrichment of fine-mapped variants (PIP \geq 10%) associated with 4 blood cell traits in ABC enhancers in the corresponding blood cell types or progenitors. Enrichment = (fraction of fine-mapped variants / fraction of all common variants) overlapping regions in each cell type. Numbers of biosamples in each category are shown in parentheses.

(c) Enrichment of fine-mapped IBD variants (PIP \geq 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 \geq 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. Gray lines show alternative sizes, which do not appear to notably affect enrichments of fine-mapped variants.

(e) % of noncoding variants across all traits that overlap an ABC enhancer in an enriched biosample, as a function of the number of cell types analyzed. Biosamples (131) were grouped into 74 cell types/tissues; and analyzed in random order. Black line: mean across 20 random orderings. Dashed gray 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 (\pm 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 where a variant overlaps more than one category, the variant was assigned to the first category that it overlapped (*i.e.*, variants in CDS 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 74 traits). Right: Fraction of variants above a threshold on the fine-mapping PIP.



Extended Data Fig. 6. ABC enhancer maps connect GWAS variants to known genes.

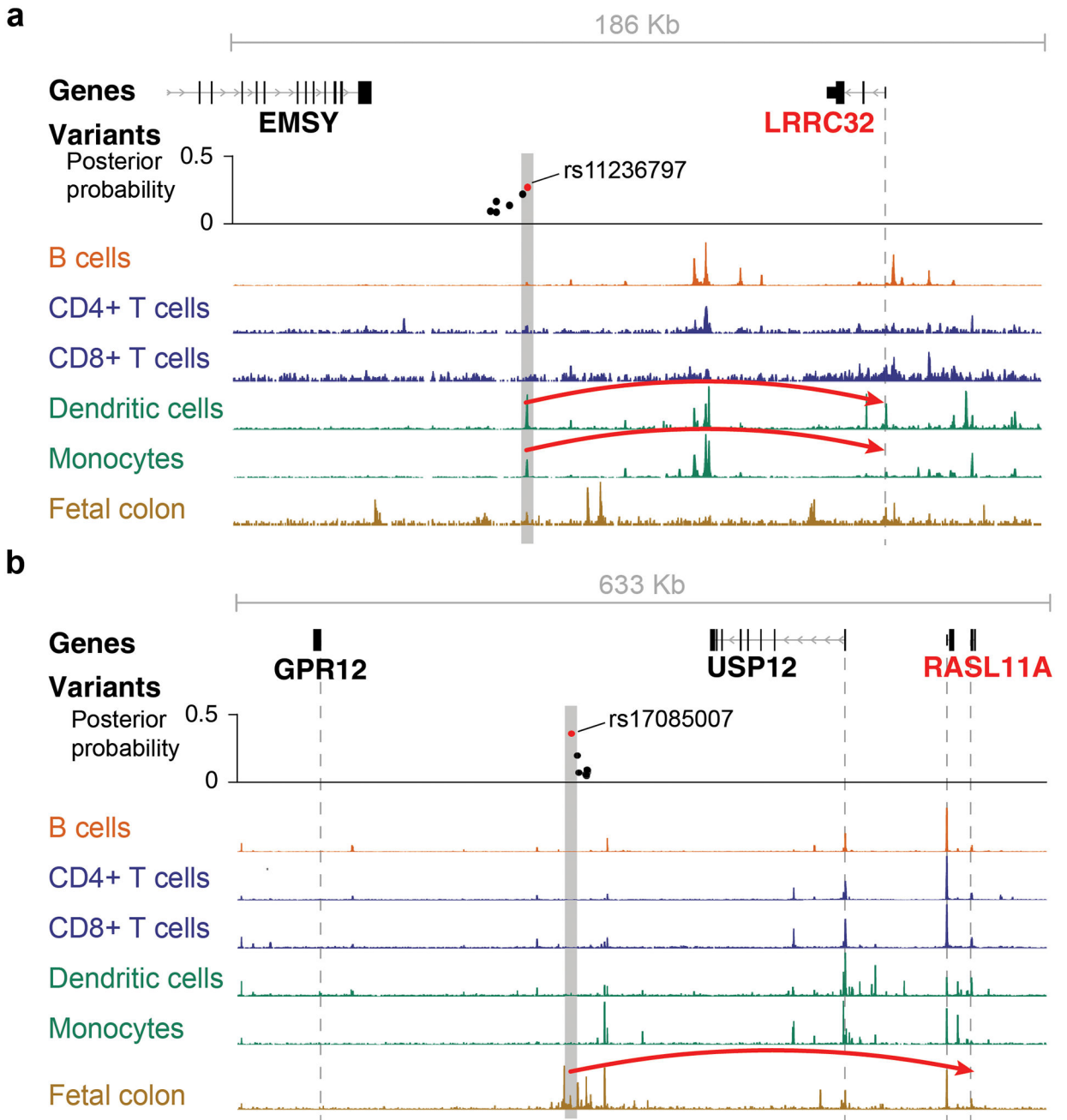
(a) ABC predictions for IBD credible sets linked to *IL10*. Heatmap shows ABC scores for each gene within 1 Mb in selected primary immune cell types. Credible Set 1 is linked by ABC to multiple genes, but *IL10* (red) has the strongest ABC score in any cell type.

(b) Cumulative density plot showing enrichment for gene sets in MSigDB among the genes prioritized by each method⁶⁴. Methods are colored 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 where both methods predict the same gene) / (# credible sets where 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 of 16 credible sets.

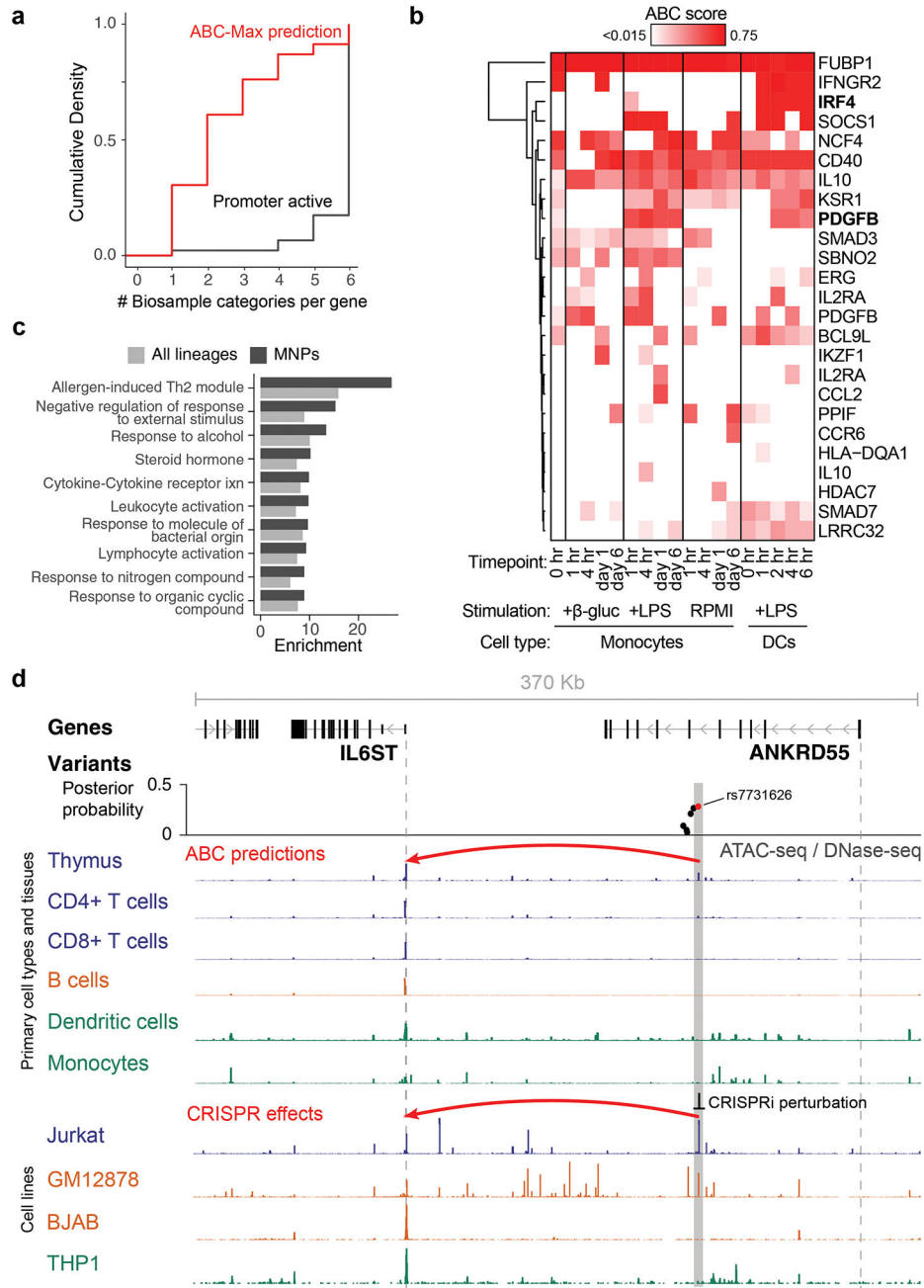
(d) Enrichment of likely causal genes for 10 blood traits (defined by common coding variants) for various prediction methods. Enrichment reflects the number of correctly predicted genes identified divided by the baseline of choosing random genes in each of the loci with a prediction.

(e) Precision-recall plot for identifying known IBD 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.

ABC-Max predictions and chromatin state in primary immune cells and fetal colon tissue at 2 IBD loci: **(a)** LRRC32 and **(b)** RASL11A. Red marks variants, enhancer-gene connections, and target genes identified by ABC-Max. Gray bars highlight the variants overlapping ABC enhancers. Vertical dotted lines represent TSSs. “DCs +LPS”: dendritic cells stimulated with bacterial lipopolysaccharide for 4 hours.



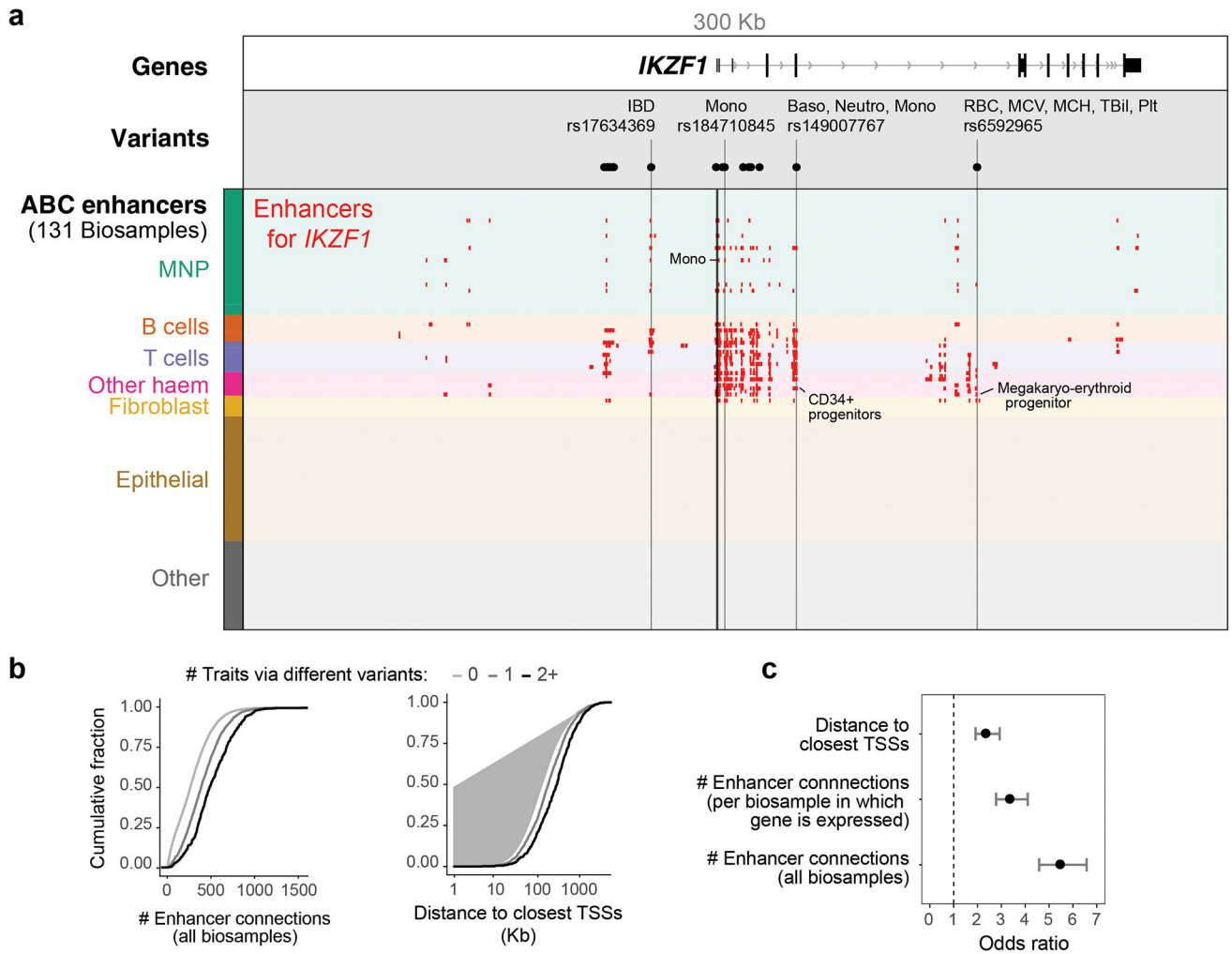
Extended Data Fig. 8. Cell-type specificity of ABC predictions.

(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) Heatmap of ABC scores for predicted IBD genes in resting and stimulated mononuclear phagocytes (from epigenomic data in monocytes⁶⁹ and dendritic cells⁷⁰). IRF4 and PDGFB (bold) are two examples where 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 genes predicted by ABC-Max in mononuclear phagocytes (MNPs, dark gray), versus when performing the same analysis among the 43 IBD genes predicted in any biosample (light gray). 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. Gray bar highlights the variant overlapping the predicted ABC enhancer. Vertical dotted lines represent TSSs. Red arc at top denotes ABC-Max prediction. Red arc at bottom denotes that CRISPRi of the highlighted enhancer significantly affects the expression of *IL6ST* only in Jurkat cells.

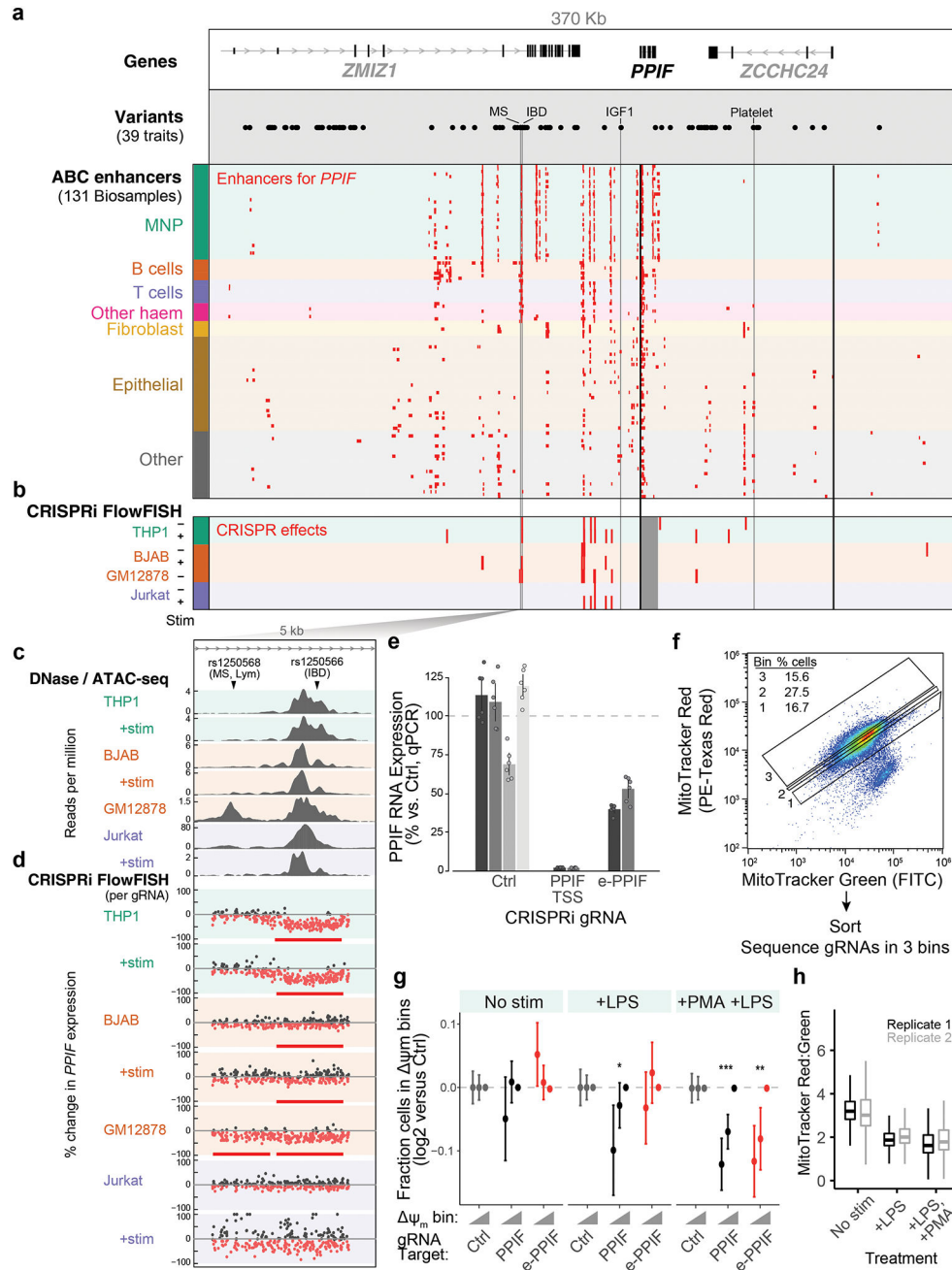


Extended Data Fig. 9. Genes linked by ABC to different traits via different variants.

(a) ABC links *IKZF1* to 2 traits via variants in 18 credible sets. Red boxes mark enhancers predicted to regulate *IKZF1*. Thick black line marks the *IKZF1* TSS. Black dots mark fine-mapped noncoding variants (PIP $\geq 10\%$) 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 (left) number of ABC enhancer-gene connections in all 131 biosamples, and (right) the distance between the TSSs of the two closest neighboring 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.

(c) The complexity of a gene's enhancer landscape is correlated with the odds of the gene being linked to multiple GWAS traits. 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 3 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 gene's promoter is active, and the genomic distance to the closest neighboring TSS on either side of the gene. Dot: mean of top decile genes (n = 1,838) versus all others (n = 16,550). Whiskers: 95% CI.



Extended Data Fig. 10. Enhancers and variants connected to *PPIF*.

(a) ABC predictions for variants near *PPIF*. Black dots represent either (i) fine-mapped variants ($PIP \geq 10\%$) for IBD and UK Biobank traits, or (ii) lead variants for any phenotype from the GWAS Catalog¹⁶ (the latter to show the approximate locations of signals for traits for which fine-mapping is not yet available). “IBD” label points to rs1250566. “MS” (multiple sclerosis) label points to rs1250568 (fine-mapped in²). 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 7 immune cell lines and stimulated states. Red boxes mark distal enhancers (CRISPR gRNAs lead to a significant decrease in the expression of PPIF). Dark gray box marks the gene body of *PPIF*, where CRISPRi cannot accurately assess the effects of putative regulatory elements⁴.
- (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 locations of variants associated with MS 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 whose effect estimate is >0% (black) or <0% (red). Red bars show regions where gRNAs have a significant effect on gene expression (FDR < 0.05), as compared by a two-sided t-test to negative control gRNAs.
- (e)** Effects of 8 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 (ψ_m). Cells expressing a pool of gRNAs were stained with MitoTracker Red and MitoTracker Green and sorted into 3 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 (Ctrl)) on ψ_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. 5d**). 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 Ctrl, PPIF, and e-PPIF, respectively. Two-sided rank-sum $P = 0.0163$ (*), 0.00426 (**), or 0.000356 (***) versus Ctrl.
- (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, $n = 8044, 99683, 99982, 99968, 99886, \text{ and } 99878$; 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 U01 CA200059 (F.L. and H.P.); U01 HG009379, R01 MH101244, and R37 MH107649 (A.K.P.); NIDDK K01DK114379 (H.H.); the Zhengxu and Ying He Foundation (H.H.); the Stanley Center for Psychiatric Research (H.H.); NIAID K22AI153648 (J.P.R.); and a Siebel Scholarship (F.L.). The authors thank Larry Schweitzer, Matteo Gentili, Moshe Biton, Chris Smillie, Aviv Regev, Masahiro Kanai, Daniel Graham, Noam Shores, Steven Gazal, Brian Cleary, Ran Cui, Patricia Rogers, Vidya Subramanian, Gavin Schnitzler, Raj Gupta, Melina Claussnitzer, Nasa Sinnott-Armstrong, Tim Majarian, Alisa 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.

Data Availability:

Immune cell line ATAC-seq and H3K27ac ChIP-seq: NCBI GEO GSE155555

GuideRNA counts from CRISPRi screens: Supplementary Tables 3, 14.

UK Biobank fine-mapping data for 71 traits: <https://www.finucanelab.org/data>

ABC predictions in 131 biosamples: <https://www.engreitzlab.org/resources/>

References

1. Claussnitzer M et al. A brief history of human disease genetics. *Nature* 577, 179–189, doi:10.1038/s41586-019-1879-7 (2020). [PubMed: 31915397]
2. Farh KK-H et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343, doi:10.1038/nature13835 (2015). [PubMed: 25363779]
3. Maurano MT et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)* 337, 1190–1195, doi:10.1126/science.1222794 (2012).
4. Fulco CP et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 51, 1664–1669, doi:10.1038/s41588-019-0538-0 (2019). [PubMed: 31784727]
5. Westra H-J & Franke L From genome to function by studying eQTLs. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842, 1896–1902, doi:10.1016/j.bbadis.2014.04.024 (2014).
6. Gasperini M, Tome JM & Shendure J Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet*, doi:10.1038/s41576-019-0209-0 (2020).
7. van Arensbergen J, van Steensel B & Bussemaker HJ In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol* 24, 695–702, doi:10.1016/j.tcb.2014.07.004 (2014). [PubMed: 25160912]
8. Consortium EP An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74, doi:10.1038/nature11247 (2012). [PubMed: 22955616]
9. Huang H et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 547, 173–178, doi:10.1038/nature22969 (2017). [PubMed: 28658209]
10. Consortium WTCC et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics* 44, 1294–1301, doi:10.1038/ng.2435 (2012). [PubMed: 23104008]
11. Ulirsch JC et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet* 51, 683–693, doi:10.1038/s41588-019-0362-6 (2019). [PubMed: 30858613]
12. Fulco CP et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769–773, doi:10.1126/science.aag2445 (2016). [PubMed: 27708057]

13. Rescigno M & Di Sabatino A Dendritic cells in intestinal homeostasis and disease. *J Clin Invest* 119, 2441–2450, doi:10.1172/JCI39134 (2009). [PubMed: 19729841]
14. Graham DB & Xavier RJ Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* 578, 527–539, doi:10.1038/s41586-020-2025-2 (2020). [PubMed: 32103191]
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. *bioRxiv* (preprint), doi: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, doi:10.1093/nar/gky1120 (2019). [PubMed: 30445434]
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, doi:10.1093/nar/gky837 (2019). [PubMed: 30239796]
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. *Nature genetics* 49, 600–605, doi:10.1038/ng.3795 (2017). [PubMed: 28218759]
19. Carvalho-Silva D et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Research* 47, D1056–D1065, doi:10.1093/nar/gky1133 (2019). [PubMed: 30462303]
20. Barbeira AN et al. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet* 15, e1007889, doi:10.1371/journal.pgen.1007889 (2019). [PubMed: 30668570]
21. Hauberg ME et al. Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. *Am J Hum Genet* 101, 157, doi:10.1016/j.ajhg.2017.06.003 (2017). [PubMed: 28686855]
22. Pers TH et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* 6, 5890, doi:10.1038/ncomms6890 (2015).
23. Liu Y, Sarkar A, Kheradpour P, Ernst J & Kellis M Evidence of reduced recombination rate in human regulatory domains. *Genome Biol* 18, 193, doi:10.1186/s13059-017-1308-x (2017). [PubMed: 29058599]
24. Granja JM et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* 37, 1458–1465, doi:10.1038/s41587-019-0332-7 (2019). [PubMed: 31792411]
25. Andersson R et al. An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461, doi:10.1038/nature12787 (2014). [PubMed: 24670763]
26. Sheffield NC et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome research* 23, 777–788, doi:10.1101/gr.152140.112 (2013). [PubMed: 23482648]
27. 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, doi:10.1038/ng.3950 (2017). [PubMed: 28869592]
28. Thurman RE et al. The accessible chromatin landscape of the human genome. *Nature* 489, 75–82, doi:10.1038/nature11232 (2012). [PubMed: 22955617]
29. 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, doi:10.1093/nar/gkz980 (2020). [PubMed: 31740966]
30. Whalen S, Truty RM & Pollard KS Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nature genetics*, doi:10.1038/ng.3539 (2016).
31. Consortium G et al. Genetic effects on gene expression across human tissues. *Nature* 550, 204–213, doi:10.1038/nature24277 (2017). [PubMed: 29022597]
32. Engreitz JM et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, doi:10.1038/nature20149 (2016).
33. Wainberg M et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* 51, 592–599, doi:10.1038/s41588-019-0385-z (2019). [PubMed: 30926968]

34. Franke A et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genetics* 42, 1118–1125, doi:10.1038/ng.717 (2010). [PubMed: 21102463]
35. Jostins L et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491, 119–124, doi:10.1038/nature11582 (2012). [PubMed: 23128233]
36. Linares PM & Gisbert JP Role of growth factors in the development of lymphangiogenesis driven by inflammatory bowel disease: a review. *Inflamm Bowel Dis* 17, 1814–1821, doi:10.1002/ibd.21554 (2011). [PubMed: 21744436]
37. Wang X & Goldstein DB Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *Am J Hum Genet* 106, 215–233, doi:10.1016/j.ajhg.2020.01.012 (2020). [PubMed: 32032514]
38. Imielinski M et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nature Genetics* 41, 1335–1340, doi:10.1038/ng.489 (2009). [PubMed: 19915574]
39. Elrod JW & Molkentin JD Physiologic functions of cyclophilin D and the mitochondrial permeability transition pore. *Circ J* 77, 1111–1122, doi:10.1253/circj.cj-13-0321 (2013). [PubMed: 23538482]
40. Ip WKE, Hoshi N, Shouval DS, Snapper S & Medzhitov R Anti-inflammatory effect of IL-10 mediated by metabolic reprogramming of macrophages. *Science (New York, N.Y)* 356, 513–519, doi:10.1126/science.aal3535 (2017).
41. Bick AG et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* 586, 763–768, doi:10.1038/s41586-020-2819-2 (2020). [PubMed: 33057201]

Extended Data References

42. Buenrostro JD, Wu B, Chang HY & Greenleaf WJ ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109, 21.29.21–21.29.29, doi:10.1002/0471142727.mb2129s109 (2015).
43. Zhu J et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654, doi:10.1016/j.cell.2012.12.033 (2013). [PubMed: 23333102]
44. Consortium EP An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74, doi:10.1038/nature11247 (2012). [PubMed: 22955616]
45. Roadmap Epigenomics C et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330, doi:10.1038/nature14248 (2015). [PubMed: 25693563]
46. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).
47. Li H et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079, doi:10.1093/bioinformatics/btp352 (2009). [PubMed: 19505943]
48. Amemiya HM, Kundaje A & Boyle AP The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 9, 9354, doi:10.1038/s41598-019-45839-z (2019). [PubMed: 31249361]
49. Javierre BM et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.e1319, doi:10.1016/j.cell.2016.09.037 (2016). [PubMed: 27863249]
50. Vierstra J et al. Global reference mapping of human transcription factor footprints. *Nature* 583, 729–736, doi:10.1038/s41586-020-2528-x (2020). [PubMed: 32728250]
51. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. *Nature methods* 9, 357–359, doi:10.1038/nmeth.1923 (2012). [PubMed: 22388286]
52. Liu JZ et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics* 47, 979–986, doi:10.1038/ng.3359 (2015). [PubMed: 26192919]
53. de Lange KM et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics* 49, 256–261, doi:10.1038/ng.3760 (2017). [PubMed: 28067908]

54. Loh PR, Kichaev G, Gazal S, Schoech AP & Price AL Mixed-model association for biobank-scale datasets. *Nat Genet* 50, 906–908, doi:10.1038/s41588-018-0144-6 (2018). [PubMed: 29892013]
55. Zhou W et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 50, 1335–1341, doi:10.1038/s41588-018-0184-y (2018). [PubMed: 30104761]
56. 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, doi:10.1016/j.ajhg.2017.08.012 (2017). [PubMed: 28942963]
57. Wang G, Sarkar A, Carbonetto P & Stephens M A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, doi:10.1111/rssb.12388 (2020).
58. McLaren W et al. The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122, doi:10.1186/s13059-016-0974-4 (2016). [PubMed: 27268795]
59. Fujita PA et al. The UCSC Genome Browser database: update 2011. *Nucleic acids research* 39, D876–882, doi:10.1093/nar/gkq963 (2011). [PubMed: 20959295]
60. Carrillo-de-Santa-Pau E et al. Automatic identification of informative regions with epigenomic changes associated to hematopoiesis. *Nucleic Acids Research* 45, 9244–9259, doi:10.1093/nar/gkx618 (2017). [PubMed: 28934481]
61. Astle WJ et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* 167, 1415–1429.e1419, doi:10.1016/j.cell.2016.10.042 (2016). [PubMed: 27863252]
62. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* 47, 1228–1235, doi:10.1038/ng.3404 (2015). [PubMed: 26414678]
63. Chen L et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 167, 1398–1414.e1324, doi:10.1016/j.cell.2016.10.026 (2016). [PubMed: 27863251]
64. Liberzon A et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740, doi:10.1093/bioinformatics/btr260 (2011). [PubMed: 21546393]
65. Kerimov N et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. *bioRxiv* (preprint), doi:10.1101/2020.01.29.924266 (2021).
66. de Leeuw CA, Mooij JM, Heskes T & Posthuma D MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 11, e1004219, doi:10.1371/journal.pcbi.1004219 (2015). [PubMed: 25885710]
67. Kerpedjiev P et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* 19, 125, doi:10.1186/s13059-018-1486-1 (2018). [PubMed: 30143029]
68. Rao SS et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680, doi:10.1016/j.cell.2014.11.021 (2014). [PubMed: 25497547]
69. Novakovic B et al. β -Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance. *Cell* 167, 1354–1368.e1314, doi:10.1016/j.cell.2016.09.034 (2016). [PubMed: 27863248]
70. Donnard E et al. Comparative Analysis of Immune Cells Reveals a Conserved Regulatory Lexicon. *Cell Syst* 6, 381–394 e387, doi:10.1016/j.cels.2018.01.002 (2018). [PubMed: 29454939]

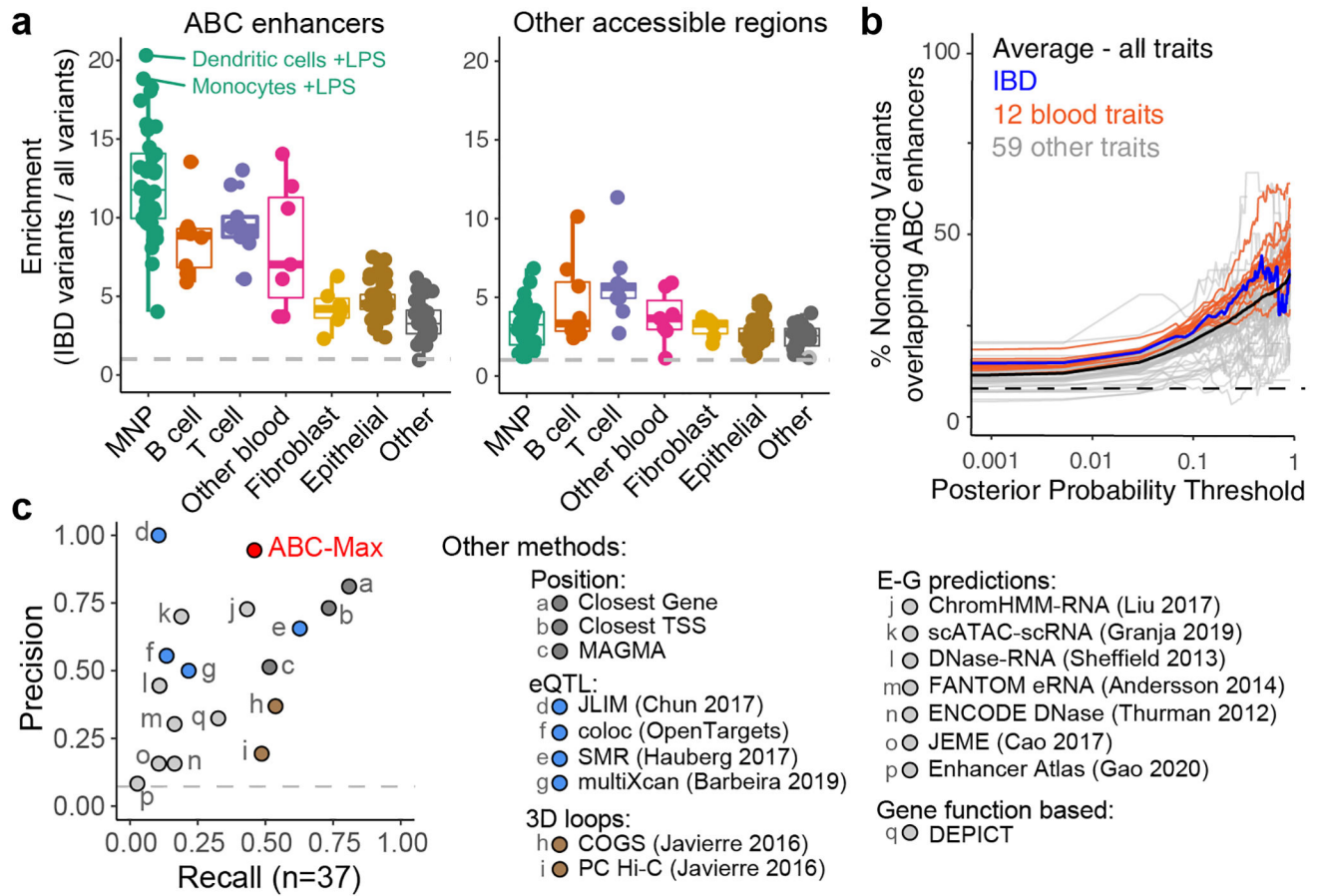


Figure 1. ABC maps connect fine-mapped variants to enhancers, genes, and cell types.
(a) Enrichment of fine-mapped IBD variants (PIP $\geq 10\%$) in ABC enhancers (left) and all other accessible regions (right) in each of 131 biosamples. MNPs: mononuclear phagocytes. Box: median and interquartile range. Whiskers: observation less than or equal to quartile $\pm 1.5 \times$ IQR.
(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 5 variants. Dashed line: fraction of all common noncoding variants that overlap ABC enhancers.
(c) Precision-recall for connecting noncoding IBD credible sets to known IBD genes¹⁴, considering 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. Where quantitative scores were available (*e.g.*, colocalization probability), plot presents the performance of choosing the gene with the best score per locus (see also Extended Data Fig. 6b).

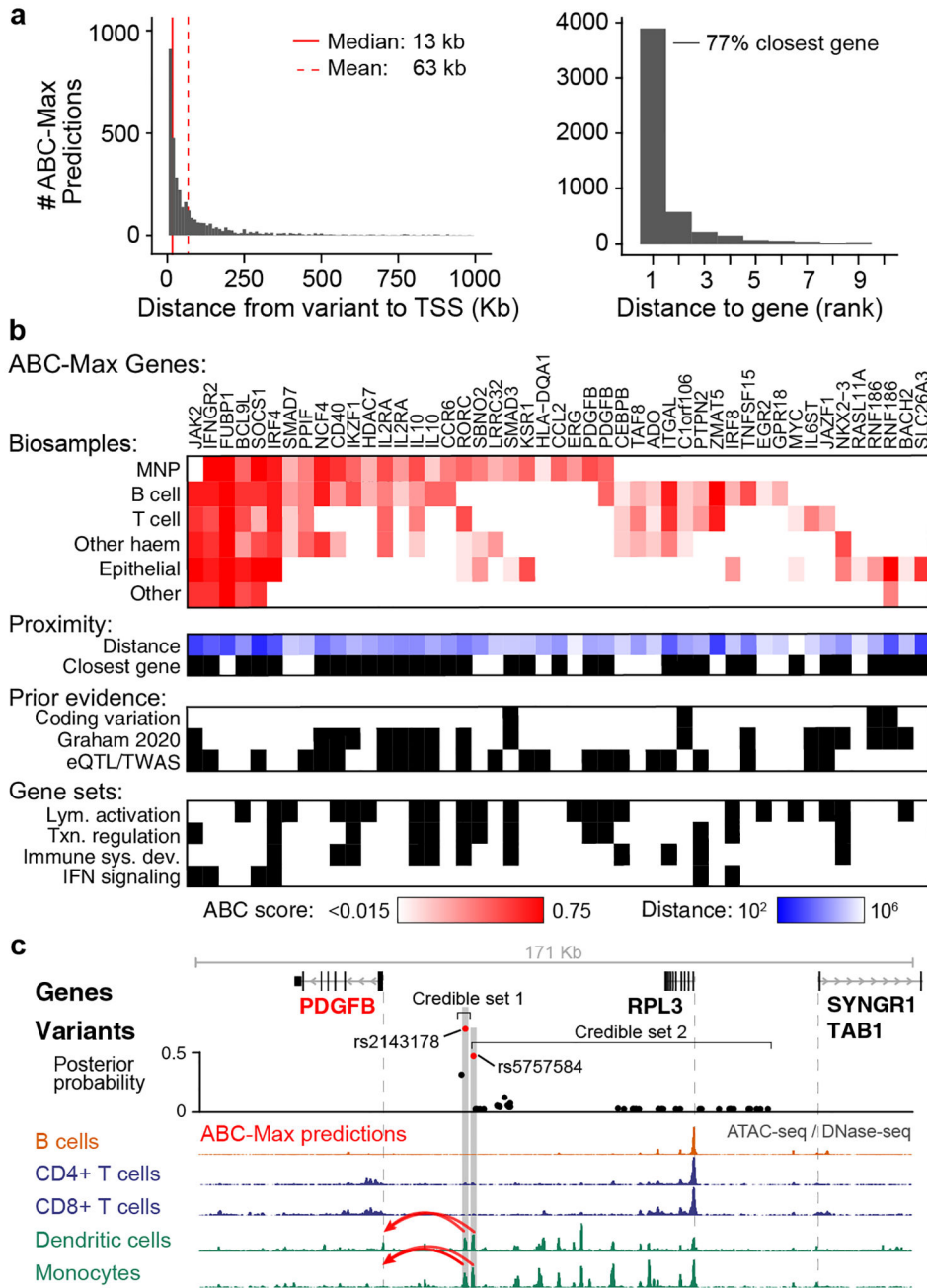


Figure 2. Connecting variants to target genes

(a) Histograms of the (left) distances from the predicted variant to the TSS of the ABC-Max target gene and (right) distance rank of the gene in the locus. Data includes 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). Heatmap: ABC scores in 6 biosample categories (maximum value within each category). Red scale: ABC score. Blue scale: \log_{10} genomic distance from variant to gene TSS. Black boxes indicate that the gene is the closest to the lead SNP, was implicated in IBD risk based on coding variation or experimental evidence about

gene function¹⁴, was identified by prior eQTL colocalization or TWAS analyses, or is in an enriched gene set (Methods).

(c) ABC-Max predictions and chromatin state at the *PDGFB* locus. Red color denotes variants, enhancer-gene connections, and target genes identified by ABC-Max. Gray bars: variants in two credible sets overlap ABC enhancers. Vertical dotted lines: TSSs.

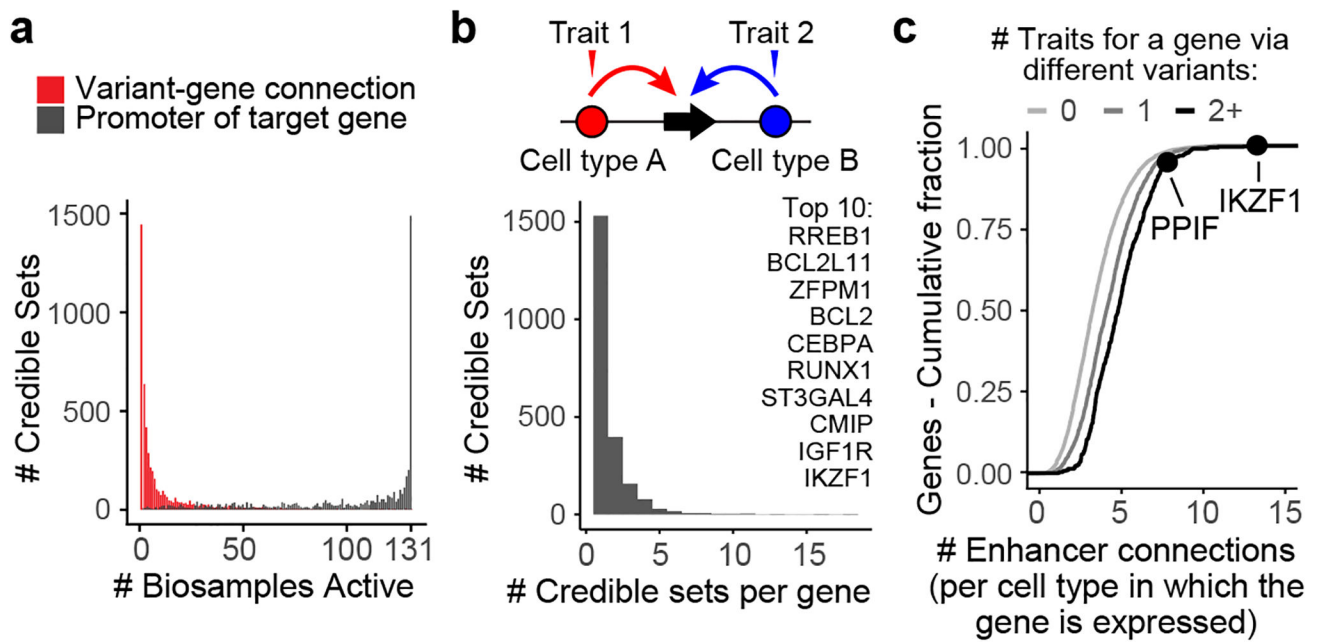


Figure 3. Cell-type specificity of ABC predictions

(a) Histogram of the number of biosamples in which (red) a variant-gene connection is predicted by ABC-Max (*i.e.*, an ABC enhancer regulates the target gene in a given biosample) and (gray) 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 \geq 10%, Methods). Model at top depicts a gene linked to different traits via different variants. Circles: enhancers. Black arrows: gene. Colored 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.

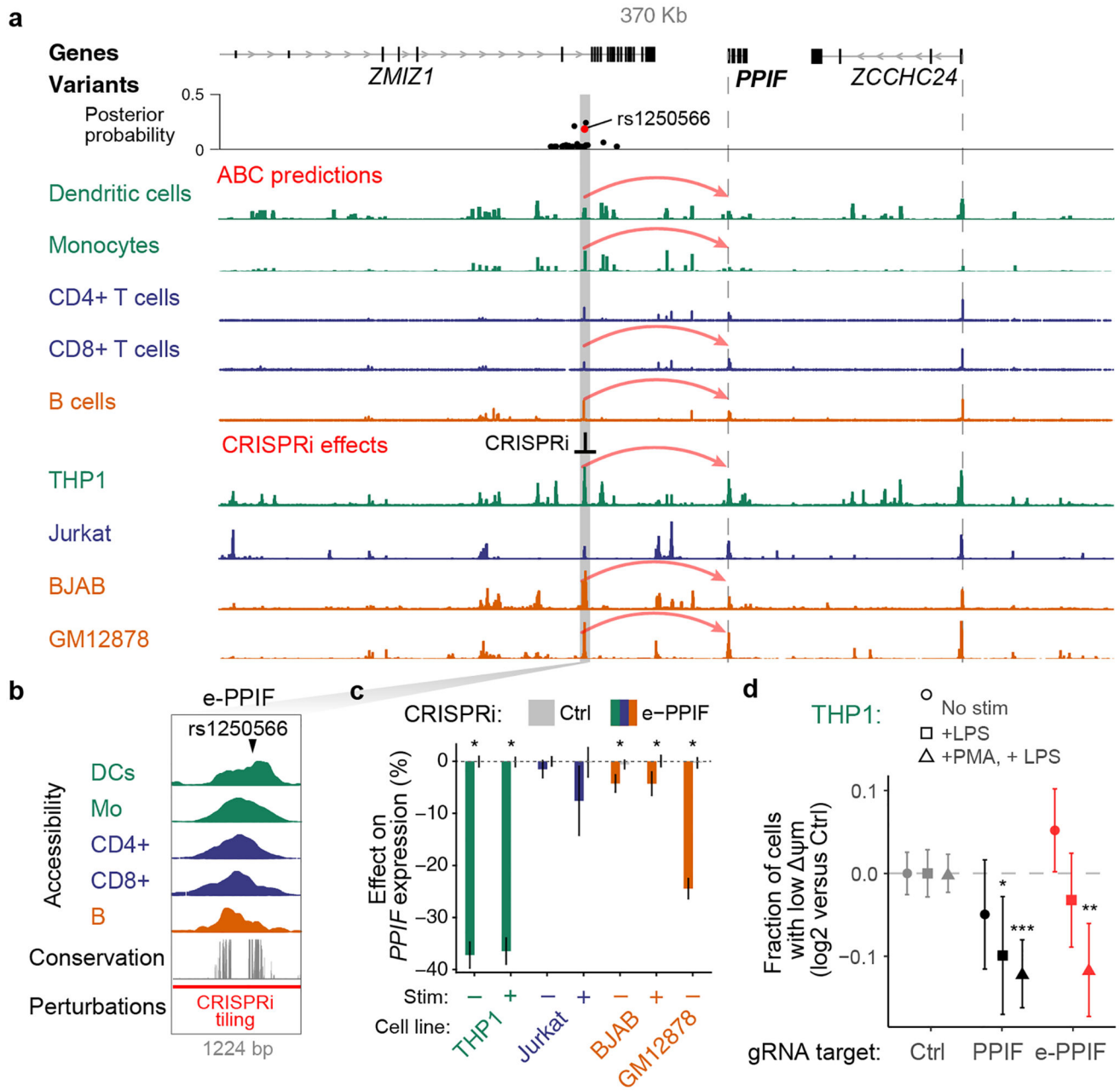


Figure 4. An enhancer regulates *PPIF* expression and mitochondrial function.

(a) An IBD risk variant (rs1250566) overlaps an enhancer predicted to regulate *PPIF*. Signal tracks: ATAC-seq or DNase-seq. Gray bar: enhancer containing rs1250566. Dashed lines: TSSs. Red arcs at top: ABC-Max predictions. Red arcs at bottom: CRISPRi leads to a significant decrease in *PPIF* expression.

(b) 1224-bp region at the *PPIF* enhancer (e-PPIF). Accessibility: DNase- or ATAC-seq from 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 (stim) conditions. Error bars show 95% confidence intervals of the mean.

*: two-sided t-test $P_{\text{Benjamini-Hochberg}} < 0.05$ for 164 CRISPRi gRNAs targeting e-PPIF compared to 814 negative control (Ctrl) gRNAs (adjusted P values from left to right: 4.68×10^{-101} , 4.86×10^{-112} , 0.019, 0.044, 1.48×10^{-71}).

(d) Effects of CRISPRi gRNAs (targeting e-PPIF, PPIF promoter, or negative controls (Ctrl)) on ψ_m , quantified as the frequency of THP1 cells carrying those gRNAs with low versus high MitoTracker Red signal (see Extended Data Fig. 10f–h). We tested THP1 cells in unstimulated conditions, stimulated with LPS, and differentiated with phorbol 12-myristate 13-acetate (PMA) and stimulated with LPS (Methods). Error bars: 95% confidence intervals for the mean of 40, 9, and 5 gRNAs for Ctrl, PPIF, and e-PPIF, respectively. Two-sided rank-sum $P = 0.0163$ (*), 0.00426 (**), or 0.000356 (***) versus Ctrl.