Description of Additional Supplementary Files

Supplementary data 1: Description of the CEDAR2 cohort

Supplementary data 2: Definition of sorted circulating immune cells

Supplementary data 3: NGS sequencing metrics and QC information. nB: naive B cells, mB: memory B cells, pB: plasmocytes, nT4: naive CD4 lymphocytes, mT4: memory CD4 lymphocytes, Th1: T helper type 1 CD4 lymphocytes, Th2: T helper type 2 CD4 lymphocytes, Th17: T helper type 17 CD4 lymphocytes, Th1_17: T helper type 1_17 CD4 lymphocytes, nTreg: naive regulatory CD4 lymphocytes, mTreg: memory regulatory CD4 lymphocytes, nT8: naive CD8 lymphocytes, mT8:memory CD8 lymphocytes, MAIT: mucosal associated invariant T cells, gdT: TCR gamma-delta T cells, NKT: natural killer T cells, NK: cytotoxic natural killer cells, ILC: innate lymphoid cells, cMO: classical monocytes, iMO: intermediate monocytes, ncMO: non-classical monocytes, pDC: plasmacytoid dendritic cells, mDC: myeloid dendritic cells, PBMC: peripheral blood mononuclear cells, EOS: eosinophils, NEUT: neutrophils, GRAN: granulocytes.

Supplementary data 4: Number of individuals kept, number of coding and non-coding genes analyzed, number of expression principal components (PC) considered in the model, number of significant cis-eQTL detected. Cell types are abbreviated as in Supplementary data 2.

Supplementary data 5: 60,113 cis-eQTL mapped in 27 sorted circulating immune cell populations. P-values are computed using QTLtools as detailed in methods. We report nominal p-values, p-values adjusted for the number of tests in the 2Mb window, and q-values computed using the latter (by cell type) following Storey & Tibshirani (ref. 20).

Supplementary data 6: 23,831 gene-specific cis-acting regulatory modules (cRM) detected in 27 sorted circulating immune cell populations. Cell type abbreviations are as in Supplementary Data 2.

Supplementary data 7: Proportion of shared eQTL with opposite sign, ontogenic distance (distance in ontogenic tree shown in SFig. 2C, , and log(1/p) value of excess module sharing, for all pairs of cell types. The p-values were computed using the binomial test described in methods, and are one-sided.

Supplementary data 8: GO analysis for e-genes assigned to the root of the ontogenic tree hence assumed to be enriched in regulatory modules active in hematopoietic progenitor cells.

Supplementary data 9: 22,067 across-gene cis-acting regulatory modules (cRM) detected in 27 sorted circulating immune cell populations. Cell type abbreviations are as in Supplementary Data 2.

Supplementary data 10: Statistics of scRNA Seq experiment. Gut biopsies (IL: terminal ileum; TC: transverse colon; RE: rectum) were collected from 60 healthy donors (Biopsy_ID, SYSCID_ID). Electronic medical records of the donors is found in Suppl. Table 1. Peripheral blood from thirty-two donors were also subjected to the blood eQTL study (Blood_eQTL). scRNA-seq data from three donors colored in gray (B2, B8, B15) were removed from further analysis, and parts of B4 and B54 samples had problems during sampling (Sample note). Single cell suspensions were prepared using a two-step dissociation protocol with some modifications (Protocol 1 - 4). We firstly separated epithelial cell layer of cells (EC) from lamina propria (LP) by incubating biopsies in the epithelial strip buffer containing EDTA and DTT. LP was dissociated in the enzyme solution containing Liberase TL. These cell suspensions were further treated with TrypLE Express enzyme. Cell concentrations and proportions of living cells after dissociation are shwon (Concentration, Viability). In Protocol 1, we enriched EC, intraepithelial lymphocytes (IEL), LP lymphocytes (LPLC), and LP myeloid cells (LPMC) using FACS. In Protocol 2 and 3, living cells were enriched using EasySep Dead Cell Removal kit and FACS, respectively. Each fraction of cells was labeled using TotalSeq anti-human antibodies conjugated with oligonucleotides (Hashtag 1 - 13). To increase complexity of multiplexing, LP cells were divided into two or three and labeled separately with distinct hashtags (LP1, LP2, LP3). The hash-tagged cells were pooled

(Pooled cell number) and loaded on one lane of 10X Genomics Chromium Single Cell Partitioning System (Single cell partitioning). scRNA-Seq libraries were prepared with 10X Genomics Chromium Cell 3' kits (v3 or v3.1) and sequenced using Illumina sequencers with indicated length of reads (Sequencing). Mapping of the reads on human genome assembly GRCh38 (Ensembl v110 annotation) or hashtag references was performed using 10X Genomics Cell Ranger software (v 8.0.1). Summary statistics of libraries from the software are shown. Cells were demultiplexed based on HTO counts using two algorithms, HTOdemux and MULTIseqDemux. We classified cells as singlets when both algorithms assigned them as singlets (Singlet Seurat). NA: not applicable; NR: not recorded.

Supplementary data 11: Intestinal cell type annotation of the hierarchical tree of single-cell clusters and utilized abbreviations. Backbone of the CEDAR2 website (Intestinal scRNA-Seq data) and for assignment of gut regulatory modules to intestinal cell types.

Supplementary data 12: 35,010 eQTL detected in scRNA-Seq data of intestinal biopsies from terminal ileum (IL), transverse colon (TC), and rectum (RE). P-values are computed using QTLtools as detailed in methods. We report nominal p-values, p-values adjusted for the number of tests in the 2Mb window, and q-values computed using the latter (by cell type) following Storey & Tibshirani (ref. 20).

Supplementary data 13: Gene-specific cis-acting regulatory modules detected by scRNA-Seq analysis of intestinal biopsies.

Supplementary data 14: Across-gene cis-acting regulatory modules detected by scRNA-Seq analysis of intestinal biopsies.

Supplementary data 15: Gene-specific regulatory modules (cRM) obtained when analyzing eQTL from circulating immune cells and gut biopsies jointly.

Supplementary data 16: Across-gene regulatory modules (cRM) obtained when analyzing eQTL from circulating immune cells and gut biopsies jointly.

Supplementary data 17: Regulatory modules in ST15 were assigned to specific cell types, separately for blood (Y axis) and biopsies (X axis). The number of combinations of cell type assignment in blood and biopsies are provided in this table. Most "unassigned" cases correspond to modules that are not active in the corresponding sample type. Cell type abbreviations are as in ST2 and ST11.

Supplementary data 18: Gene-specific modules that are assigned either to the lymphoid or myeloid category in gut, while being undetected in blood.

Supplementary data 19: Boundaries of the IBD risk loci (and sub-loci) considered for colocalization analysis.

Supplementary data 20: Matching DAP-EAP with |theta| >= 0.6 and FDR <= 0.10.

Supplementary data 21: IBD risk loci with matching DAP-EAP for IBD, CD and/or UC. Cell type abbreviations are as defined in ST2 and ST11.

Supplementary data 22: List of genes with matching DAP-EAP for IBD, CD and/or UC. Cell type abbreviations are defined as in ST2 and ST11.

Supplementary data 23: Comparison of the frequency of DAP-EAP matches in the CEDAR2 dataset, between IBD, CD and UC on the one hand, and dilated cardiomyopathy (DCM) on the other.

Supplementary data 24: Top Reactome pathways identified when submitting the list of 483 protein coding genes with matching DAP-EAP (FDR <= 0.10) for IBD, CD, and/or UC. P-values and FDR-values are from Reactome (https://reactome.org) and are one-sided.

Supplementary data 25: Genes with DAP-matching EAP in corresponding IBD risk locus. Genes = all genes. Genes_with_function = genes with functional evidence supporting causality in the literature.

Supplementary data 26: DAP-matching e-genes with reported functional evidence in favor of a role in IBD pathogenesis.

Supplementary data 27: Results of a differential expression analysis of the transcriptome of 27 circulating immune cell populations between 55 active IBD cases and 79 controls. Analyses were conducted using DESeq2 that uses a 2-sided Wald test.

Supplementary data 28: Drugs used to treat IBD.

Supplementary data 29: Panels of antibodies used for FACS sorting of PBMCs.