Genome-wide association meta-analysis identifies 126 novel loci for diverticular disease and implicates connective tissue and colonic motility

Christopher J. Neylan M.D.[1], Michael G. Levin M.D.[2,3], Katherine Hartmann M.D. Ph.D.[4], Katherine Beigel M.S.[5], Sam Khodursky Ph.D.[1], John S. DePaolo M.D. Ph.D.[1], Sarah Abramowitz[1,6], Emma E. Furth M.D.[7], Robert O. Heuckeroth M.D. Ph.D.[8,9,10], Scott M. Damrauer M.D.[1,2,3,11], Lillias H. Maguire M.D.[3,12]

Affiliations:
[1] Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
[2] Cardiovascular Institute, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
[3] Corporal Michael Crescenz VA Medical Center, Philadelphia, PA 19104
[4] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
[5] Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, Philadelphia, PA 19104
[6] The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY 11549
[7] Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104
[8] Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, 3401 Civic Center Blvd, Philadelphia, PA 19104
[9] The Children's Hospital of Philadelphia Research Institute and Abramson Research Center, 3615 Civic Center Blvd, Philadelphia, PA 19104, USA
[10] Perelman School of Medicine at the University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104
[11] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
[12] Division of Colon and Rectal Surgery, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Corresponding author:
Christopher J. Neylan, MD
Department of Surgical Education
Hospital of the University of Pennsylvania
3400 Spruce Street, 4 Maloney
Philadelphia, PA 19104
Telephone: (215) 847-3524
Fax: (215) 662-7983
Email: Christopher.neylan@pennmedicine.upenn.edu

ABSTRACT

Diverticular disease is a common and morbid complex phenotype influenced by both innate and environmental risk factors. We performed the largest genome-wide association study meta-analysis for diverticular disease, identifying 126 novel loci. Employing multiple downstream analytic strategies, including tissue and pathway enrichment, statistical fine-mapping, allele-specific expression, protein quantitative trait loci and drug-target investigations, and linkage disequilibrium score regression, we prioritized causal genes and produced several lines of evidence linking diverticular disease to connective tissue biology and colonic motility. We substantiated these findings by integrating single-cell RNA sequencing data, showing that prioritized diverticular disease-associated genes are enriched for expression in colonic smooth muscle, fibroblasts, and interstitial cells of Cajal. In quantitative analysis of surgical specimens, we found a substantial reduction in the density of elastin present in the sigmoid colon in severe diverticulitis.

INTRODUCTION

Diverticula are protrusions of the mucosa and submucosa through the muscularis propria in the human colon. Diverticulosis (the asymptomatic presence of diverticula) occurs in the sigmoid colon of over 60% of Americans by the age of 60 years.[1-6] Diverticulitis, the inflammation of a diverticular colon, occurs in a minority of patients with diverticulosis and manifests with symptoms ranging from mild abdominal pain to fatal bowel perforation. Diverticulitis recurs after the first episode in 20-30% of people and can have sequela such as fistulization and abscess. Due to the ubiquity of the precursor lesion, the burden of disease caused by diverticulitis is high: in aggregate, diverticulitis is annually responsible for nearly 300,00 inpatient admissions and $2 billion of direct costs in the United States.[7-13]

Diverticulitis has historically been understood as an acquired disease attributable to environmental risk factors.[14] Generations of healthcare providers attributed diverticulitis to constipation and food (e.g. nuts and seeds) lodging in a diverticulum leading to inflammation and physical disruption of the colon.[15] While environmental factors are important, evidence now suggests that these hypotheses are overly simplistic, and a more thorough understanding of the etiology of diverticular disease (DD) is greatly needed.[16-21]

Genetic liability to diverticular disease (an umbrella term for diverticulosis and diverticulitis) has recently been demonstrated. Twin studies indicate 40-53% heritability. Several genome wide association studies (GWASs) have been performed.[22-26] The largest study to date included over 724,000 individuals and identified 150 loci associated with DD.[25,27,28] However, in these studies, most participants were genetically similar to the European reference population.[25,26]

In the current study, we 1) perform the largest and most diverse GWAS meta-analysis for diverticular disease to date identifying 126 novel loci and 159 novel genes, 2) prioritize causal genes, pathways, and involved proteins, including protein targets for drug repurposing, 3) integrate single cell RNA sequencing data to localize gene expression to cell types, 4) investigate the potentially causal role of the gut microbiome with Mendelian randomization, and 5) analyze the extracellular matrix in surgical specimens of patients with diverticulitis. Multiple lines of evidence in our study indicate that DD may best be understood as a complex phenotype strongly linked to structural factors (e.g. connective tissue such as elastin, as well as smooth muscle) and colonic motility.

METHODS

GWAS meta-analysis

To identify genetic variants associated with diverticular disease, we performed a meta-analysis of two publicly available genome-wide association studies (GWASs) conducted in separate discovery cohorts (the Million Veteran Program (MVP) and the United Kingdom Biobank (UKB)) (Figure 1).[29,30] The Department of Veterans Affairs (VA) MVP is a longitudinal health cohort that has enrolled > 1 million Veteran volunteers. It utilizes a harmonized ancestry and race/ethnicity (HARE) approach to integrate self-identified racial background information with genetic data.[29,31] The MVP-HARE cohort consisted of 638,797 participants: 457,880 in the European American population, 121,703 in the African American population, 51,153 in the Hispanic/Latino American population, and 8,061 in the Asian American population.[29,31] The UKB is a collection of >500,000 genotyped individuals that links genetic and phenotypic information in a large-scale biomedical database, and the Pan-UKB is an initiative that performs pan-ancestry genetic analysis within the UKB.[32] Genetic similarity to reference populations was

determined in the source GWAS and is described elsewhere.[30] The Pan-UKB cohort consisted of 381,541 individuals in the following populations: 365,887 European (EUR), 7,587 Central or south Asian (CSA), 5,835 African (AFR), 1,370 middle eastern (MID), and 862 admixed American (AMR).[33] Population-specific meta-analyses of the MVP-HARE and Pan-UKB cohorts were performed for each of the above populations (data not shown) in addition to a meta-analysis of the overall cohort. MVP-HARE populations were mapped to Pan-UKB populations in the following manner: European American to EUR, African American to AFR, Hispanic/Latino to AMR, and Asian American to CSA. In each source GWAS, diverticular disease was identified by phecode 562, a parent code for a group of nine International Classification of Disease (ICD)-10 codes and nineteen ICD-9 codes specific to diverticulosis and diverticulitis.[34]

Quality control was performed on each set of summary statistics to remove variants with poor imputation scores ($r^2 < 0.3$) or erroneous p-values or allele frequencies (e.g. > 1 or < 0). Meta-analysis was performed with a fixed-effects, inverse-variance weighted model using METAL.[35] Meta-analyzed data was then quality-controlled and cleaned using GWASinspector (Supplemental Figure 1).[36] SNPs with a minor allele frequency (MAF) < 0.001 were discarded, strand flips were corrected, and variants were removed if they were missing crucial values, duplicated, or ambiguous.

Definition of loci and secondary signals

From the meta-analyzed overall cohort, loci were defined around genome-wide significant variants ($p < 5 \times 10^{-8}$) via local clumping with a 500kb window and an r^2 threshold of 0.01. A single lead variant (or index variant) was retained per locus (Supplemental Table 1). To

identify additional, independently significant variants within a locus (secondary signals), we performed statistical fine-mapping using CARMA.[37] A fine-mapped variant was considered a secondary signal if it fell within a defined credible set and had the highest PIP within its credible set. Index variants and secondary signals were then carried forward for further analysis. Loci were additionally defined in each population-specific meta-analysis (EUR, CSA, AFR, MID, AMR).

Gene mapping and prioritization

An ensemble approach considering several methods was used to map each signal to a gene. First, the nearest gene to the signal was identified using *gwasRtools* package in R.[38] Next, MAGMA was run using the *magmaR* package to identify a gene associated with the signal based on the 1000 Genomes Project Phase 3 EUR imputation panel.[39] Third, OpenTargets annotations were analyzed to determine whether the signal lies in a known coding region or regulatory region for a gene.[40] Finally, multi-trait colocalization was performed using HyPrColoc to identify colocalization between the signals and eQTLs in the eQTLGen Phase I cis-eQTL blood data.[41] All genes mapped to each signal by these methods were then tabulated and a single gene was prioritized for each signal based on a process that prioritized coding variants over non-coding variants, and gave preference to genes identified by multiple methods at the same signal (Supplemental Table 2).

Enrichment analysis

The list of genes to which signals were mapped was then subjected to a suite of enrichment analyses to gain insight into the biological function of the genes. First, utilizing

Enrichr, the set genes was analyzed for enrichment against numerous gene set libraries (e.g. the Gene Ontology Cellular Component dataset) (Figure 2a, b).[42-44] Then, FUMA (*GENE2FUNC* feature) was used to analyze tissue expression with the following parameters: all background genes, GTEx v8 tissue types (excluding MHC region), FDR multiple test correction (adjusted $p < 0.05$) (Figure 2c).[45]

Linkage Disequilibrium Score Regression (LDSC)

LDSC was used to estimate the heritability of diverticular disease and genetic correlation with select phenotypes.[36] LD scores derived from the 1000 Genomes Project European samples were utilized. SNP heritability was transformed to the liability scale using a sample prevalence of 12.3% (the calculated prevalence in our data) and a population prevalence of 30% (the approximated epidemiologic prevalence of disease).[46] To account for the potential inflationary effect of assumed population prevalence on liability-transformed SNP heritability, we also report the unadjusted, observed h^2 as well as the unbiased h^2-Z score. Publicly available summary statistics were used for phenotypes other than diverticular disease, and sources are listed in Supplemental Table 3.

Microbiome Mendelian randomization (MR)

To investigate whether a potentially causal relationship between the microbiome and diverticular disease exists, we performed univariable two-sample MR and bidirectional MR, using the TwoSampleMR package (v. 0.6.8) in R.[47] Inverse variance weighting was used for MR estimation in cases where the instrument comprised > 1 SNP; otherwise Wald ratios were utilized. Genetic instruments for microbiome exposures were taken from a GWAS (listed in

Supplemental Table 3) of host genetic variation effects on gut microbiome composition among 18,340 participants, which revealed significant associations between host genetics and 385 bacterial genera.[48] The outcome of diverticular disease was instrumented using variants discovered in our full cohort GWAS meta-analysis. The two sets of summary statistics were harmonized prior to running MR, and variants with missing data were excluded. We filtered variants in both sets of summary statistics to a p-value threshold of $p < 5 \times 10^{-6}$, with an r^2 threshold of 0.001. This p-value threshold is often used in Mendelian randomization and fulfills the instrumental variable assumption of relevance.[49,50] This MR analysis meets the three core instrumental variable assumptions of relevance, independence, and exclusion. Sensitivity analysis was performed for all exposure instruments via leave-one-out analysis.

Allele-specific expression

At heterozygous DNA positions, allele-specific expression can quantify the RNA expression of one allele relative to the other and thereby interrogate *cis*-regulatory variation.[51] We analyzed allele-specific expression at each independent index variant and secondary signal across all GTEx (v8) dbGaP tissue types, filtering to variant-person-tissue combinations with at least five reference and five alternate alleles.

Protein drug targets

We performed a proteome-wide plasma Mendelian randomization screen to identify proteins with putatively causal effects on the development of diverticular disease, which could serve as targets for the repurposing of existing drugs. First, we identified genetic variants associated with plasma protein levels (protein quantitative trait loci, or pQTLs) in two large-scale

plasma-proteomics studies (deCODE and UK Biobank Pharma Proteomics Project, hereafter UKB-PPP).[52,53] Then we performed Mendelian randomization using pQTLs as instrumental variables and our diverticular disease GWAS meta-analysis summary statistics to instrument the outcome of diverticular disease. Instrumental variables were selected if they met the following criteria: genome-wide significance ($p < 5 \times 10^{-8}$), maximum linkage disequilibrium r^2 of 0.01, and *cis* variant within a 500kb window of the respective gene. Wald ratios were used for instruments comprised of a single SNP, and inverse-variance weighted MR was used otherwise. Proteins were considered significant if the adjusted p-value was < 0.05, according to the Benjamini-Hochberg correction to maintain FDR < 0.05. Using a previously-generated list of 1,263 actionable targets, defined as proteins targeted by approved drugs or drugs in the clinical phase of development, we then identified actionable targets within our own results.[54] To understand the biological implication of the significant proteins, we performed enrichment analysis of the identified proteins in the Reactome pathway database.

Immunohistochemistry

Formalin-fixed, paraffin embedded tissue blocks from surgical resection of sigmoid colons were obtained from the Pathology Department of the Hospital of the University of Pennsylvania. Patient data were reviewed in the electronic medical record (chart review) to classify the indication for surgery as incidental (sigmoidectomy for reasons unrelated to diverticular disease, such as iatrogenic injury to the colon), elective quiescent (patient with recurrent diverticulitis, but asymptomatic at the time of surgery), acutely symptomatic (patient with symptomatic diverticulitis at the time of surgery), colonic fistulization, or emergent colonic perforation. These indications were then collapsed into a variable termed ‘Patient Classification’

which took the values of 'No disease' for an incidental indication in which the patient did not have diverticular disease, 'Diverticulosis' for an incidental indication in which the patient had diverticulosis but not diverticulitis, 'Emergent diverticulitis' for an indication of emergent colonic perforation, or 'Non-emergent diverticulitis' for any of the following indications: acutely symptomatic diverticulitis, elective quiescent diverticulitis, or colonic fistulization. Patients with known connective tissue disorders or confounding colonic pathology (e.g. Crohn's disease or ulcerative colitis) were excluded. A subset of patients with an incidental indication were undergoing colectomy for colon cancer and did not have DD; these were used as controls. For those with DD, slides were prepared from tissue samples taken from areas of grossly apparent disease and from the margins of the surgical specimen (typically the area, in the estimation of the operating surgeon, furthest from disease). All slides from patients without DD were considered normal tissue. Stains for elastin, the smooth muscle marker MYH11, and the collagen stain Sirius Red were applied using the protocols given in Supplemental Figure 2.

After staining, to quantify the amount of elastin and collagen present, stained sections were imaged at ×40 magnification using Hamamatsu NanoZoomerS360 to generate whole slide digital images, which were uploaded to QuPath 5.1 for analyses.[55] We developed pixel classifiers for the elastin and collagen stains through several rounds of annotation and training using the Random Forests algorithm from a compiled region image set from the cohort. This classifier was then applied to scanned images of the annotated inner and outer muscularis propria for each section to determine the percentage of elastin and collagen per square millimeter (mm) of muscularis propria in the region adjacent to the diverticulum. The color intensity per square mm was derived by running the "Compute intensity features" application and choosing the diaminobenzidine (color deconvoluted) channel.

The average percent of elastin and collagen per square mm of muscularis propria on each slide was calculated by averaging the percentage of elastin or collagen per square mm of inner and outer muscularis propria layers. A multivariable linear regression was then constructed with average elastin as the dependent variable and patient classification, age, and sex as the independent variables.

Single cell RNAseq analysis

Cell-level gene expression was ascertained from single cell RNA-sequencing (RNAseq) of the human colon in two previously published datasets.[56,57] Hickey et al. (2023) used the 10x Genomics platform to process samples surgically removed from sites throughout the human small and large intestine (see Supplemental Table 3 for data availability of both studies). In this dataset, we restricted our analysis to samples from the sigmoid colon and re-annotated "Myofibroblasts/SM 1," "Myofibroblasts/SM 2," and "Myofibroblasts/SM 3" as 'Myofibroblasts' for our analyses.[56] Wright et al. (2021) was produced from manual dissection of the human colonic myenteric plexus, resulting in RNA sequencing of 3798 glia, 5568 smooth muscle cells, 377 interstitial cells of Cajal, and 2153 macrophages.[57] In both datasets, we employed the single-cell disease-relevance score (scDRS) method to evaluate expression of our GWAS-associated genes in individual cells. scDRS is an analytic method that identifies differential expression of GWAS-implicated genes at single cell level.[58] We ran scDRS on each published dataset with our list of GWAS-identified genes. scDRS was run using the python (v3.10) package scdrs (v1.0.2).

To determine whether GWAS-identified genes are expressed in specific cell types, for the Wright et al. and the Hickey et al. datasets, we applied k-means clustering and generated

clustered dot plots using the R package scCustomize.[59] The k-means clustering and relative expression were used to assess patterns of gene expression, and we identified specific k-means clusters of genes that were more predominantly expressed in specific cell types. We visually identified clusters of genes with enriched expression in any cell type noted to be significant in scDRS, and then generated heatmaps of gene expression for all genes that comprised the relevant clusters. We finally performed pathway analysis of the genes the comprised the relevant clusters, using Enrichr.

RESULTS

GWAS meta-analysis identified 236 loci

The meta-analysis consisted of 934,030 individuals in the combined MVP and UKB cohorts, of whom 114,806 (12%) had diverticular disease and 819,224 (88%) did not (Figure 1). In the overall cohort, 763,970 (82%) individuals were genetically similar to the European reference population (EUR); 114,236 (12%) to the African reference population (AFR); 46,867 (5.0%) to the admixed American population (AMR); 7,587 (0.81%) to the Central/South Asian (CSA) population; and 1,370 (0.15%) to the Middle Eastern (MID) population. We obtained single-variant association statistics for a total of 45,061,282 SNPs, of which 99.8% passed GWASinspector variant processing, with λ=1.16 (Supplemental Figure 1). We identified 236 loci defined around index variants that were genome-wide significant ($p < 5 \times 10^{-8}$) for their association with diverticular disease, of which 126 (53%) were novel (Supplemental Table 1). Statistical fine-mapping identified an additional 105 secondary signals, for a total of 341 independent signals (Supplemental Table 2). Population-specific meta-analyses revealed 22 loci present in the overall cohort meta-analysis that were not present in the EUR population meta-

analysis. However, all loci present in a non-EUR population were also present in the EUR population (i.e. there were no loci unique to non-EUR populations).

Gene mapping and biological function

Of 341 independent signals, gene mapping was successful for 309; several signals mapped to the same gene, and ultimately the 341 signals mapped to 214 distinct genes. Of the 214 genes, 7 (3%) were implicated in human disease by damaging coding variants (missense or stop-gain), while 20 (9%) were prioritized by coding or well-annotated regulatory variants (Supplemental Table 2). Interrogation of the Reactome (2024) pathway database showed the prioritized genes were significantly enriched in "extracellular matrix organization" and "elastic fiber formation", as well as "disorders of nervous system development" (Figure 2a), with similar results seen in gene ontology analysis (Figure 2b). Specific mapped genes involved in extracellular matrix remodeling include *ELN*, *LTBP4*, *LOXL1, COL* family and *ADAM* family genes, while examples of mapped genes involved in smooth muscle function, which implicates both structure and motility, included *ANO1, RYR2, and VIPR1*. GTEx (v8) analysis performed using FUMA revealed significant enrichment of gene expression for mapped DD-associated genes in visceral smooth muscle containing organs (e.g. sigmoid colon, transverse colon, esophagus, bladder, uterus), as well as to arteries, the tibial nerve and visceral adipose (Figure 2c).

Linkage Disequilibrium Score Regression

The liability-scale SNP heritability (h^2_{SNP}) was 0.33 for diverticular disease. The unadjusted observed h^2 was 0.10, while the h^2-Z score was 15.7. Moderate genetic correlation

was observed with connective tissue phenotypes such as hernia and pelvic organ prolapse, weak genetic correlation with abdominal aortic aneurysm (AAA), and negligible correlation with ulcerative colitis and Crohn's disease (Figure 3).

Microbiome Mendelian randomization

Dietary factors, such as low fiber intake or high red meat consumption, are established risk factors for diverticulitis, but it is unclear whether these associations are causative or correlative.[17-19,60] Additionally, recent evidence suggested that microbiome composition may be implicated in the development of diverticular disease.[20,21] We hypothesized that dietary factors may mediate diverticulitis risk through changes in the microbiome, and utilized Mendelian randomization to elucidate the relationship between microbiome composition and diverticular disease based on a GWAS that links host genetic variants to specific gut microbes.

Genetically predicted increases in the abundances of the phylum *Actinobacteria,* the genera *Desulfovibrio and Intestinibacter*, and several other taxa in the gut microbiome were associated with decreased genetic liability to diverticular disease, while genetically predicted increases in the genera *Flavonifractor, Lachnoclostridium*, *Butyricimonas*, and other taxa were associated with increased liability (Figure 4). Significant effects were not seen in the reverse direction, suggesting a causal relationship between bacterial abundance and diverticular disease. The most intriguing result came from *Ruminococcus*, which was associated with both increased and decreased liability to diverticular disease depending upon the particular SNPs used to instrument the exposure. A possible explanation is that *Ruminococcus* exerts a species-specific effect, whereby some *Ruminococcus* species increase the risk of diverticular disease while others are protective against it. The putative *Ruminococcus*-mediated increase in DD risk may be

attributable to *Ruminoccocus gnavus,* which produces a pro-inflammatory polysaccharide that is hypothesized to change the intraluminal milieu in a way that increases susceptibility to diverticular disease.[61] Indeed, several reports in the literature confirm this disease-promoting effect of *Ruminococcus gnavus* and even demonstrate decreased *Ruminococcus gnavus* in response to the initiation of a high-fiber diet.[60-62] Taken together, these data suggest the possibility that the protective effect of a high-fiber diet is mediated through a decrease in *Ruminococcus gnavus*. Our study provides only limited evidence for this hypothesis, and the protective effect of *Ruminococcus* that is simultaneously seen in MR is not concordant, though this result may be confounded by pleiotropy in addition to multiple *Ruminococcus* species.

Allele specific expression

At heterozygous DNA sites in diploid individuals, equal expression of haplotypes is expected except in cases of *cis*-regulatory variation (e.g. expression or splicing quantitative trait loci) and certain other rare regulatory phenomena (e.g. nonsense-mediated decay or parent-of-origin epigenetic imprinting).[44,63] Using RNA sequencing data, the quantity of RNA representing each nucleotide can be measured to detect allele-specific expression, which thus implies a functional regulatory variant.[64] We investigated allele specific expression data available in GTEx RNA sequencing data and found data for 64 (19%) of 341 SNPs representing independent GWAS signals (most variants identified by GWAS are non-coding and thus not well-captured by RNA sequencing data). One variant, rs34093919 which is a missense variant mapped to *LTBP4,* demonstrated clear patterns of imbalanced expression (Figure 5) in the sigmoid colon of all observed individuals. The dramatically imbalanced expression is suggestive of (likely *cis*-) regulatory variation, providing evidence that this variant or those in near perfect linkage

disequilibrium are functional regulatory variants of *LTBP4*. An additional SNP, rs1705003, which is a missense variant mapped to *CUTA*, demonstrated imbalanced expression in only some individuals suggesting that the variant itself is not functional but rather serves as a marker of a functional variant(s) in the locus. That we have specific evidence of this imbalanced expression in the colon is consistent with a potential role of these genes in the development of diverticular disease.

Protein drug targets

Proteome-wide Mendelian randomization analysis using deCODE and UKB-PPP pQTLs to link diverticular disease GWAS associations to serum protein levels identified a total of 157 distinct proteins significantly associated with diverticular disease (Figure 6), from a possible 1703 proteins in deCODE and 1963 proteins in UKB-PPP (see Supplemental Tables 4 and 5 for full results). Many of these proteins whose abundance in plasma is associated with diverticular disease-associated SNPs directly overlapped with our independently mapped genes in GWAS analysis, including *ELN, LTBP4, COL15A1*, and *COL6A1*, while other proteins corresponded to the same sub-families as mapped genes such as *LOXL1/LOXL3* and *SEMA3G/SEMA3D/SEMA3E/SEMA3F*. Analysis of the 157 proteins using the Reactome 2024 database (via Enrichr) revealed enrichment in the extracellular matrix proteins and elastin-forming processes, mirroring the enrichment analysis for our mapped genes (Supplemental Figure 3). It should be noted that this plasma MR experiment identifies circulating proteins, not proteins measured in colonic tissue. However, extracellular matrix proteins are often released into circulation during disease states, making a segment of colon with diverticulitis a plausible

source of the observed proteins.[65] Moreover, even if released into circulation from other tissues, the observed plasma levels may reflect a global process that would also occur in colonic tissue.

Of the 157 proteins found to associate with diverticular disease, 21 (13.4%) were known drug targets, including *ELN* and *TNF* products (Figure 6). These provide the possibility of actionable targets, allowing the repurposing of existing drugs against these gene products. For example, the mTOR inhibitor everolimus improves *in vitro* smooth muscle cell differentiation of stem cells derived from individuals with elastin insufficiency because of *ELN* mutations.[66] Our results suggest that targeting DD with everolimus is a compelling topic of future research.

Immunohistochemistry

To further define anatomic findings in diverticular disease, assess the plausibility of links to elastin, and appreciate differences between individuals with diverticulitis relative to individuals with diverticulosis, we analyzed colon surgical resections from 24 patients (78 slides total). Of the 24 people, 13 (54%) has diverticulitis, 5 (21%) had diverticulosis without diverticulitis, 6 (25%) were without diverticular disease. Among the 13 individuals with diverticulitis, 3 (23%) had life-threatening diverticulitis requiring emergency surgery, while 10 required surgery but disease was not life-threatening and surgery timing was elective. A representative *MYH11* antibody-stained slide from a non-emergency diverticulitis patient demonstrates severe attenuation of the muscularis propria overlying the diverticulum (Figure 7). The slide does not depict the diverticulitis-causing (i.e. perforated) diverticulum, but rather an asymptomatic diverticulum in the region resected. Because the observed diverticulum is not at the exact site of perforation, it is less likely that attenuated muscularis propria is secondary to the perforation (or its associated inflammation), and more likely that it was present in the tissue prior

to the perforation, raising the possibility that this muscular change is present in many regions of the sigmoid colon of the sampled patient.

To test the hypothesis that weakness of the elastin extracellular matrix could predispose to diverticulitis, as suggested by our GWAS results, we analyzed elastin fibers in muscularis propria of sigmoid colon resections. Several areas were sampled for analysis: in diverticulitis patients, samples were taken from areas around a diverticulum in a region of diverticulitis (i.e. where the perforated diverticulum lies), and separately samples were taken from areas around a diverticulum at the surgical margin (i.e. the region furthest from the point of diverticulitis); in diverticulosis patients, samples were taken from areas around a diverticulum (i.e. areas of gross disease) and separate samples were taken from grossly healthy areas (i.e. areas without diverticulosis); in patients without diverticular disease, samples were taken from areas that appeared grossly healthy (sample sites are depicted in Supplemental Figure 4). Across all samples, the density of elastin fibers in the muscularis propria was markedly decreased among individuals who required emergency surgery for diverticulitis compared to all other groups (Figure 8a). In attempt to exclude the local effects of diverticulitis-associated inflammation and uncover potentially genetically-mediated differences between individuals with and without diverticulitis, we compared the density of elastin present at the surgical margin in patients with emergency diverticulitis (location 5 in Supplemental Figure 4) to a) the density of elastin present immediately adjacent to diverticula in individuals with diverticulosis (location 2 in Supplemental Figure 4), and b) the density of elastin present in individuals without diverticular disease (location 1 in Supplemental Figure 4). These are the appropriate anatomic sites for comparison because they create the most stringent contrast: the site most likely to have normal elastin density in individuals with diverticulitis (the site furthest from the perforation) compared to the site least

likely to have normal elastin density in individuals with diverticulosis (the site immediately adjacent to the diverticulum). To perform these comparisons, we constructed a multivariable linear regression (controlling for age and sex), which revealed that those requiring emergency surgery for diverticulitis had a 41% reduction (95% CI -66% to -17%, p = 0.001; Figure 8b) in elastin density in the muscularis propria (sampled at location 5) relative to patients without diverticular disease (sampled at location 1), and a 43% reduction (95% CI -69% to -17%, p = 0.002) in elastin density compared to those with diverticulosis (sampled at location 2). These analyses were re-run for percent of collagen, but no significant differences were observed between groups (Supplemental Figure 5).

Single cell RNAseq analysis

To identify the cell types in which diverticular disease-associated genes are normally expressed, we employed scDRS to analyze two single cell RNAseq datasets (Hickey et al. and Wright et al.) which collected data from patients without diverticular disease (Hickey et al.) or without an indication of diverticulitis for resection (but undocumented diverticular disease history, Wright et al.). In the Hickey et al. dataset, DD-associated genes in the sigmoid colon were most differentially abundant in WNT5B+ fibroblasts (also referred to as "villus fibroblasts WNT5B+") and the interstitial cells of Cajal (ICC) (Figure 9). The analysis was repeated across immune, epithelial, secretory, and enteroendocrine cells without any without identifying statistically significant single cell disease relevance scores in any of these cells. In the Wright et al. dataset, a similar analysis identified increased expression of DD-associated genes in colon to smooth muscle cells and ICC (Figure 10).

To more specifically understand which of the 214 DD-associated genes were enriched in these cell types, we constructed dot plots with k-means clustering for each dataset, which

revealed in the Hickey et al. dot plot the enrichment of cluster 3 genes in ICC cells and cluster 10 genes in WNT5B+ fibroblasts, and in the Wright et al. dot plot the enrichment of cluster 7 genes in ICC and cluster 9 genes in smooth muscle cells (Figure 11). Heatmaps constructed for the genes that comprise these clusters revealed marked expression of several DD-associated genes, including *ANO1*, in ICC cells, as well as *CACNB2* and *RYR2* in smooth muscle cells (Figure 12). *ANO1* is a calcium activated chloride channel that plays a role in pacemaker activity of the intestine,[67] while *CACNB2* and *RYR2* are both calcium channels involved in muscle contraction[44,68]. These observations suggest some diverticular disease-associated genes have critical roles in colonic motility. Cluster-level pathway analysis of DD-associated genes also implicated biological processes of nerve conduction (e.g. cytoskeleton of the presynaptic active zone), muscle contraction (e.g. sarcoplasmic reticulum), and extracellular matrix (Figure 13).

These single cell RNAseq analyses suggest that many genes associated with DD identified by our GWAS are expressed at relatively high levels in ICC, smooth muscle, and fibroblasts in the colon, and have important roles in gastrointestinal motility and extracellular matrix formation or maintenance. This implies that genetic variation that influences colonic motility and/or extracellular matrix composition may contribute to the formation of diverticular disease. These proposed biological mechanisms are illustrated in Figure 14.

DISCUSSION

The current study offers multiple lines of evidence that diverticular disease (DD) is predominantly a disorder of connective tissue biology and colonic motility, in contrast to historical hypotheses that DD is an infectious or inflammatory disease. This assertion is concordant with prior clinical associations of DD with rare connective tissue disorders; common diseases such as hernia, prolapse, and colonic dysmotility; and more recent genomic work.[25,69]

We began with the largest and most diverse GWAS meta-analysis for diverticular disease (DD) to date and identified 126 novel loci, nearly doubling the number reported in the literature. These loci were mapped to genes that were enriched for expression in the sigmoid colon (the site of the vast majority of diverticular disease) and are involved in biological processes related to extracellular matrix maintenance (particularly elastin synthesis and degradation) and colonic motility. By integrating our results with single cell RNAseq data from the sigmoid colon, we localized expression of diverticulitis-associated gene clusters to fibroblasts, interstitial cells of Cajal, and smooth muscle cells (which generate force needed for bowel motility and to resist stretching when intraluminal pressure is high). Finally, we found direct evidence of reduced elastin in the sigmoid colon of people with severe diverticulitis. These findings have important consequences, implicating novel cell types, suggesting a critical role for elastin and motility in DD, and providing a prioritized list of potential targets for medical therapy for a common disease with no effective medical treatment.

Analysis of tissue and pathway enrichment revealed two important findings. The first is that the genes at loci associated with DD are enriched not only in the sigmoid colon but also in hollow organs (e.g. artery, bladder, uterus) with prominent smooth muscle and extracellular matrix (ECM) components. The second is that many biological processes in which DD-associated genes are enriched relate to ECM formation (such as elastin fiber production) or degradation, as well as gastrointestinal motility; and that cells in which DD genes are enriched (fibroblasts, smooth muscle, and ICC) are responsible for these biological processes. For example, the interpretation of enrichment of DD-associated genes in ICC suggests the possibility of a motility-mediated component of these genes, which is reflected by the same genes being involved in neuronal transmission and digestive physiology. Alterations in motility and in

multiple colonic cell types (nerves, glia, neuroendocrine cells, etc) have long been appreciated in DD patients and specimens.[70] However, these clinical and histological studies are purely correlative and investigators have not been able to determine whether the associated changes are causes or effects of DD. In this context, these genetic analyses are important because they implicate specific cell types and do not suffer from reverse causation. Taken together, our findings suggest a structural (ECM and smooth muscle) and functional (ICC and smooth muscle) basis for DD.

This interpretation is further supported by our LDSC results showing shared genetic liability to other connective tissue and smooth muscle diseases (e.g. hernia, pelvic organ prolapse, and abdominal aortic aneurysm). Interestingly, we did not see an association on LDSC with inflammatory bowel disease, again suggesting that etiopathogenesis of diverticular disease is more similar to these connective tissue or muscle layer disorders than to inflammatory diseases of the bowel.

The hypothesis that DD stems from alterations in connective tissue biology was further supported and refined by our analysis of intraoperatively collected sigmoid colonic tissue. Specifically, our finding of dramatically reduced elastin density (41% decrease, adjusted for age and sex) in the sigmoid colon muscularis propria of people requiring emergency surgery for severe diverticulitis compared to controls without diverticular disease. Most interestingly, similar reductions in muscularis elastin density (43% decrease, adjusted for age and sex) were present in those needing emergency surgery when compared to those with diverticulosis, leading us to conclude that elastin reductions are associated with diverticulitis per se rather than solely DD writ large. This supports the possibility that loss of elastin in the muscularis propria increases risk of diverticulitis, but not to diverticulosis, an unexpected correlation. Given that diverticulitis

is the most clinically relevant manifestation of DD, this has potentially profound clinical implications by providing insight into the biological features that distinguish diverticulosis patients from diverticulitis patients. Such a finding heightens the potential that correcting elastin deficiencies may prevent diverticulitis. One potential therapy would be everolimus, a mammalian target of rapamycin (mTOR) antagonist that was shown to rescue defective elastin phenotypes in stem cell-derived vascular smooth muscle cells in individuals with *ELN* gene mutations that caused elastin insufficiency.[66] It should be noted that the direction of causality in the association of diverticulitis and decreased elastin is not definitively established in our analyses: it is possible chronic inflammation in individuals with diverticulitis leads to elastin degradation.[71] However, none of the individuals in our analysis who had emergency diverticulitis surgery had prior episodes of diverticulitis or gastrointestinal inflammatory conditions, leading us to believe it is more likely that decreased elastin density led to diverticulitis rather than the reverse.

One of the main limitations of this study is that it relies for the identification of diverticular disease on electronic medical record coding using Phecodes that do not distinguish between diverticulosis and diverticulitis. This is mitigated by the fact that diverticulosis is an asymptomatic and therefore underdiagnosed disease; DD cases come to light primarily through symptomatic diverticulitis (approximately 4% of individuals with diverticulosis)[72] or through screening colonoscopy (for which the rates are well below 100% of recommended individuals, and which is an imperfect tool to diagnose diverticulosis).[73,74] Therefore, we hypothesize that our identified DD cases are enriched for symptomatic disease, which explains the 12% DD prevalence in our population, which is substantially lower than population-level estimates of asymptomatic DD.[4] Nevertheless, disentangling the genetic etiology of symptomatic from

asymptomatic DD remains a challenge. A further limitation is that the largest contributor to our meta-analysis was a GWAS in the VA Million Veteran Program, which is overwhelmingly male,[75] and that the current study does not include significant numbers of individuals of Asian ancestry, whose anatomically right-sided diverticular disease may stem from a different genetic etiology.[76]

ACKNOWLEDGEMENTS & STATEMENTS

The authors would like express their gratitude for the participants of the Penn Medicine Biobank. The PMBB is supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

The authors thank Million Veteran Program (MVP) staff, researchers, and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study.[75] (See https://www.research.va.gov/mvp/ for more details). This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by the Veterans Administration (VA) Million Veteran Program (MVP) award #000. This research is based on dbGaP data accession number phs001672.v11.p1.

The current study was approved by the University of Pennsylvania IRB Protocol # 850391.

SOURCES OF FUNDING

C.J.N is supported by the American Society of Colon and Rectal Surgeons General Surgery Resident Research Initiation Grant, the Institute for Translational Medicine and Therapeutics of the Perelman School of Medicine at the University of Pennsylvania, and the NIH National Center for Advancing Translational Sciences (TL1TR001880). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.
M.G.L is supported by the Doris Duke Foundation (Grant 2023-0224)
J.S.D is supported by the American Heart Association (23POST1011251)
S.A.A is supported by the Sarnoff Cardiovascular Research Foundation
R.O.H is supported by the Irma and Norman Braman Endowment, Lustgarten Center Endowment, and NIH R01 DK128282.
S.M.D is supported by the US Department of Veterans Affairs Clinical Research and Development Award IK2-CX001780. This publication does not represent the views of the Department of Veterans Affairs or the US Government.
L.H.M is supported by NIH K08 DK124687

REFERENCES

1. Kim CN. What is the Difference Between Right- and Left-Sided Colonic Diverticulitis? *Ann Coloproctol*. Dec 2016;32(6):206-207. doi:10.3393/ac.2016.32.6.206
2. Telem DA, Buch KE, Nguyen SQ, Chin EH, Weber KJ, Divino CM. Current recommendations on diagnosis and management of right-sided diverticulitis. *Gastroenterol Res Pract*. 2009;2009:359485. doi:10.1155/2009/359485
3. Violi A, Cambie G, Miraglia C, et al. Epidemiology and risk factors for diverticular disease. *Acta Biomed*. Dec 17 2018;89(9-S):107-112. doi:10.23750/abm.v89i9-S.7924
4. Peery AF, Keku TO, Martin CF, et al. Distribution and Characteristics of Colonic Diverticula in a United States Screening Population. *Clin Gastroenterol Hepatol*. Jul 2016;14(7):980-985 e1. doi:10.1016/j.cgh.2016.01.020
5. Townsend C. *Sabiston textbook of surgery*. 21. ed. Elsevier; 2021:pages cm.
6. Sheth AA, Longo W, Floch MH. Diverticular disease and diverticulitis. *Am J Gastroenterol*. Jun 2008;103(6):1550-6. doi:10.1111/j.1572-0241.2008.01879.x
7. Peery AF, Barrett PR, Park D, et al. A high-fiber diet does not protect against asymptomatic diverticulosis. *Gastroenterology*. Feb 2012;142(2):266-72 e1. doi:10.1053/j.gastro.2011.10.035
8. Gaber CE, Kinlaw AC, Edwards JK, et al. Comparative Effectiveness and Harms of Antibiotics for Outpatient Diverticulitis : Two Nationwide Cohort Studies. *Ann Intern Med*. Jun 2021;174(6):737-746. doi:10.7326/M20-6315
9. Simianu VV, Strate LL, Billingham RP, et al. The Impact of Elective Colon Resection on Rates of Emergency Surgery for Diverticulitis. *Ann Surg*. Jan 2016;263(1):123-9. doi:10.1097/SLA.0000000000001053
10. Lamb MN, Kaiser AM. Elective resection versus observation after nonoperative management of complicated diverticulitis with abscess: a systematic review and meta-analysis. *Diseases of the colon and rectum*. Dec 2014;57(12):1430-40. doi:10.1097/DCR.0000000000000230
11. Bharucha AE, Parthasarathy G, Ditah I, et al. Temporal Trends in the Incidence and Natural History of Diverticulitis: A Population-Based Study. *Am J Gastroenterol*. Nov 2015;110(11):1589-96. doi:10.1038/ajg.2015.302
12. Peery AF, Keku TO, Addamo C, et al. Colonic Diverticula Are Not Associated With Mucosal Inflammation or Chronic Gastrointestinal Symptoms. *Clin Gastroenterol Hepatol*. Jun 2018;16(6):884-891 e1. doi:10.1016/j.cgh.2017.05.051
13. Reddy VB, Longo WE. The burden of diverticular disease on patients and healthcare systems. *Gastroenterol Hepatol (N Y)*. Jan 2013;9(1):21-7.
14. Rege RV, Nahrwold DL. Diverticular disease. *Current problems in surgery*. Mar 1989;26(3):133-89. doi:10.1016/0011-3840(89)90031-2
15. Strate LL, Liu YL, Syngal S, Aldoori WH, Giovannucci EL. Nut, corn, and popcorn consumption and the incidence of diverticular disease. *JAMA*. Aug 27 2008;300(8):907-14. doi:10.1001/jama.300.8.907
16. Ma W, Nguyen LH, Song M, et al. Intake of Dietary Fiber, Fruits, and Vegetables and Risk of Diverticulitis. *Am J Gastroenterol*. Sep 2019;114(9):1531-1538. doi:10.14309/ajg.0000000000000363

17. Liu PH, Cao Y, Keeley BR, et al. Adherence to a Healthy Lifestyle is Associated With a Lower Risk of Diverticulitis among Men. *Am J Gastroenterol*. Dec 2017;112(12):1868-1876. doi:10.1038/ajg.2017.398
18. Cao Y, Strate LL, Keeley BR, et al. Meat intake and risk of diverticulitis among men. *Gut*. Mar 2018;67(3):466-472. doi:10.1136/gutjnl-2016-313082
19. Strate LL, Keeley BR, Cao Y, Wu K, Giovannucci EL, Chan AT. Western Dietary Pattern Increases, and Prudent Dietary Pattern Decreases, Risk of Incident Diverticulitis in a Prospective Cohort Study. *Gastroenterology*. Apr 2017;152(5):1023-1030 e2. doi:10.1053/j.gastro.2016.12.038
20. Gueimonde M, Ouwehand A, Huhtinen H, Salminen E, Salminen S. Qualitative and quantitative analyses of the bifidobacterial microbiota in the colonic mucosa of patients with colorectal cancer, diverticulitis and inflammatory bowel disease. *World J Gastroenterol*. Aug 7 2007;13(29):3985-9. doi:10.3748/wjg.v13.i29.3985
21. Daniels L, Budding AE, de Korte N, et al. Fecal microbiome analysis as a diagnostic test for diverticulitis. *Eur J Clin Microbiol Infect Dis*. Nov 2014;33(11):1927-36. doi:10.1007/s10096-014-2162-3
22. Granlund J, Svensson T, Olen O, et al. The genetic influence on diverticular disease--a twin study. *Aliment Pharmacol Ther*. May 2012;35(9):1103-7. doi:10.1111/j.1365-2036.2012.05069.x
23. Strate LL, Erichsen R, Baron JA, et al. Heritability and familial aggregation of diverticular disease: a population-based study of twins and siblings. *Gastroenterology*. Apr 2013;144(4):736-742 e1; quiz e14. doi:10.1053/j.gastro.2012.12.030
24. Maguire LH, Handelman SK, Du X, Chen Y, Pers TH, Speliotes EK. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat Genet*. Oct 2018;50(10):1359-1365. doi:10.1038/s41588-018-0203-z
25. Wu Y, Goleva SB, Breidenbach LB, et al. 150 risk variants for diverticular disease of intestine prioritize cell types and enable polygenic prediction of disease susceptibility. *Cell Genom*. Jul 12 2023;3(7):100326. doi:10.1016/j.xgen.2023.100326
26. Schafmayer C, Harrison JW, Buch S, et al. Genome-wide association analysis of diverticular disease points towards neuromuscular, connective tissue and epithelial pathomechanisms. *Gut*. May 2019;68(5):854-865. doi:10.1136/gutjnl-2018-317619
27. Morris CA. Williams Syndrome. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, eds. *GeneReviews((R))*. 1993.
28. Royce PM, Steinmann BU. *Connective tissue and it heritable disorders : molecular, genetic, and medical aspects*. 2nd ed. Wiley-Liss; 2002:xvii, 1201 p.
29. Verma A, Huffman JE, Rodriguez A, et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*. Jul 19 2024;385(6706):eadj1182. doi:10.1126/science.adj1182
30. Karczewski G, Kanai et al. Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv*. 2024 Mar 15 2024;doi:10.1101/2024.03.13.24303864
31. Fang H, Hui Q, Lynch J, et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet*. Oct 3 2019;105(4):763-772. doi:10.1016/j.ajhg.2019.08.012

32. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. Mar 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
33. Pan-UKB. Phenotypes. Accessed 2/10/2025, https://pan.ukbb.broadinstitute.org/phenotypes/index.html
34. Wei WQ, Bastarache LA, Carroll RJ, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One*. 2017;12(7):e0175508. doi:10.1371/journal.pone.0175508
35. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. Sep 1 2010;26(17):2190-1. doi:10.1093/bioinformatics/btq340
36. Ani A, van der Most PJ, Snieder H, Vaez A, Nolte IM. GWASinspector: comprehensive quality control of genome-wide association study results. *Bioinformatics*. Apr 9 2021;37(1):129-130. doi:10.1093/bioinformatics/btaa1084
37. Yang Z, Wang C, Liu L, et al. CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses. *Nat Genet*. Jun 2023;55(6):1057-1065. doi:10.1038/s41588-023-01392-0
38. Gogarten SM, Bhangale T, Conomos MP, et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics*. Dec 15 2012;28(24):3329-31. doi:10.1093/bioinformatics/bts610
39. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*. Apr 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219
40. Mountjoy E, Schmidt EM, Carmona M, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet*. Nov 2021;53(11):1527-1533. doi:10.1038/s41588-021-00945-5
41. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun*. Feb 3 2021;12(1):764. doi:10.1038/s41467-020-20885-8
42. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. Apr 15 2013;14:128. doi:10.1186/1471-2105-14-128
43. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. May 2000;25(1):25-9. doi:10.1038/75556
44. AM OD, Nakamura H, Puri P. Altered ryanodine receptor gene expression in Hirschsprung's disease. *Pediatr Surg Int*. Sep 2019;35(9):923-927. doi:10.1007/s00383-019-04504-2
45. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. Nov 28 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
46. Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 9 2014;111(49):E5272-81. doi:10.1073/pnas.1419064111

47. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. Jul 12 2018;362:k601. doi:10.1136/bmj.k601
48. Kurilshikov A, Medina-Gomez C, Bacigalupe R, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. Feb 2021;53(2):156-165. doi:10.1038/s41588-020-00763-1
49. Palmos AB, Millischer V, Menon DK, et al. Proteome-wide Mendelian randomization identifies causal links between blood proteins and severe COVID-19. *PLoS Genet*. Mar 2022;18(3):e1010042. doi:10.1371/journal.pgen.1010042
50. Hubel C, Gaspar HA, Coleman JRI, et al. Genetic correlations of psychiatric traits with body composition and glycemic traits are sex- and age-dependent. *Nat Commun*. Dec 18 2019;10(1):5765. doi:10.1038/s41467-019-13544-0
51. Fan J, Hu J, Xue C, et al. ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet*. May 2020;16(5):e1008786. doi:10.1371/journal.pgen.1008786
52. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. Dec 2021;53(12):1712-1721. doi:10.1038/s41588-021-00978-w
53. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. Oct 2023;622(7982):329-338. doi:10.1038/s41586-023-06592-6
54. Gaziano L, Giambartolomei C, Pereira AC, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nature medicine*. Apr 2021;27(4):668-676. doi:10.1038/s41591-021-01310-z
55. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. Dec 4 2017;7(1):16878. doi:10.1038/s41598-017-17204-5
56. Hickey JW, Becker WR, Nevins SA, et al. Organization of the human intestine at single-cell resolution. *Nature*. Jul 2023;619(7970):572-584. doi:10.1038/s41586-023-05915-x
57. Wright CM, Schneider S, Smith-Edwards KM, et al. scRNA-Seq Reveals New Enteric Nervous System Roles for GDNF, NRTN, and TBX3. *Cell Mol Gastroenterol Hepatol*. 2021;11(5):1548-1592 e1. doi:10.1016/j.jcmgh.2020.12.014
58. Zhang MJ, Hou K, Dey KK, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet*. Oct 2022;54(10):1572-1580. doi:10.1038/s41588-022-01167-z
59. *scCustomize: Custom Visualizations & Functions for Streamlined Analyses of Single Cell Sequencing*. 2021. https://doi.org/10.5281/zenodo.5706430
60. Ma W, Wang Y, Nguyen LH, et al. Gut microbiome composition and metabolic activity in women with diverticulitis. *Nat Commun*. Apr 29 2024;15(1):3612. doi:10.1038/s41467-024-47859-4
61. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences of the United States of America*. Jun 25 2019;116(26):12672-12677. doi:10.1073/pnas.1904099116
62. Le Leu RK, Winter JM, Christophersen CT, et al. Butyrylated starch intake can prevent red meat-induced O6-methyl-2-deoxyguanosine adducts in human rectal tissue: a randomised clinical trial. *Br J Nutr*. Jul 2015;114(2):220-30. doi:10.1017/S0007114515001750

63. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol*. Sep 17 2015;16(1):195. doi:10.1186/s13059-015-0762-6
64. van Beek D, Verdonschot J, Derks K, et al. Allele-specific expression analysis for complex genetic phenotypes applied to a unique dilated cardiomyopathy cohort. *Sci Rep*. Jan 11 2023;13(1):564. doi:10.1038/s41598-023-27591-7
65. Sutherland TE, Dyer DP, Allen JE. The extracellular matrix and the immune system: A mutually dependent relationship. *Science*. Feb 17 2023;379(6633):eabp8964. doi:10.1126/science.abp8964
66. Kinnear C, Agrawal R, Loo C, et al. Everolimus Rescues the Phenotype of Elastin Insufficiency in Patient Induced Pluripotent Stem Cell-Derived Vascular Smooth Muscle Cells. *Arterioscler Thromb Vasc Biol*. May 2020;40(5):1325-1339. doi:10.1161/ATVBAHA.119.313936
67. Singh RD, Gibbons SJ, Saravanaperumal SA, et al. Ano1, a Ca2+-activated Cl- channel, coordinates contractility in mouse intestine by Ca2+ transient coordination between interstitial cells of Cajal. *J Physiol*. Sep 15 2014;592(18):4051-68. doi:10.1113/jphysiol.2014.277152
68. Soldatov NM. CACNB2: An Emerging Pharmacological Target for Hypertension, Heart Failure, Arrhythmia and Mental Disorders. *Curr Mol Pharmacol*. 2015;8(1):32-42. doi:10.2174/1874467208666150507093258
69. Bassotti G, Chistolini F, Morelli A. Pathophysiological aspects of diverticular disease of colon and role of large bowel motility. *World J Gastroenterol*. Oct 2003;9(10):2140-2. doi:10.3748/wjg.v9.i10.2140
70. Bassotti G, Battaglia E, Bellone G, et al. Interstitial cells of Cajal, enteric nerves, and glial cells in colonic diverticular disease. *Journal of clinical pathology*. Sep 2005;58(9):973-7. doi:10.1136/jcp.2005.026112
71. Heinz A. Elastases and elastokines: elastin degradation and its significance in health and disease. *Crit Rev Biochem Mol Biol*. Jun 2020;55(3):252-273. doi:10.1080/10409238.2020.1768208
72. Strate LL, Morris AM. Epidemiology, Pathophysiology, and Treatment of Diverticulitis. *Gastroenterology*. Apr 2019;156(5):1282-1298 e1. doi:10.1053/j.gastro.2018.12.033
73. Engel-Nitz NM, Miller-Wilson LA, Le L, Limburg P, Fisher DA. Colorectal screening among average risk individuals in the United States, 2015-2018. *Prev Med Rep*. Feb 2023;31:102082. doi:10.1016/j.pmedr.2022.102082
74. Niikura R, Nagata N, Shimbo T, Akiyama J, Uemura N. Colonoscopy can miss diverticula of the left colon identified by barium enema. *World J Gastroenterol*. Apr 21 2013;19(15):2362-7. doi:10.3748/wjg.v19.i15.2362
75. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. Feb 2016;70:214-23. doi:10.1016/j.jclinepi.2015.09.016
76. Choe EK, Lee JE, Chung SJ, et al. Genome-wide association study of right-sided colonic diverticulosis in a Korean population. *Sci Rep*. May 14 2019;9(1):7360. doi:10.1038/s41598-019-43692-8

TABLES AND FIGURES

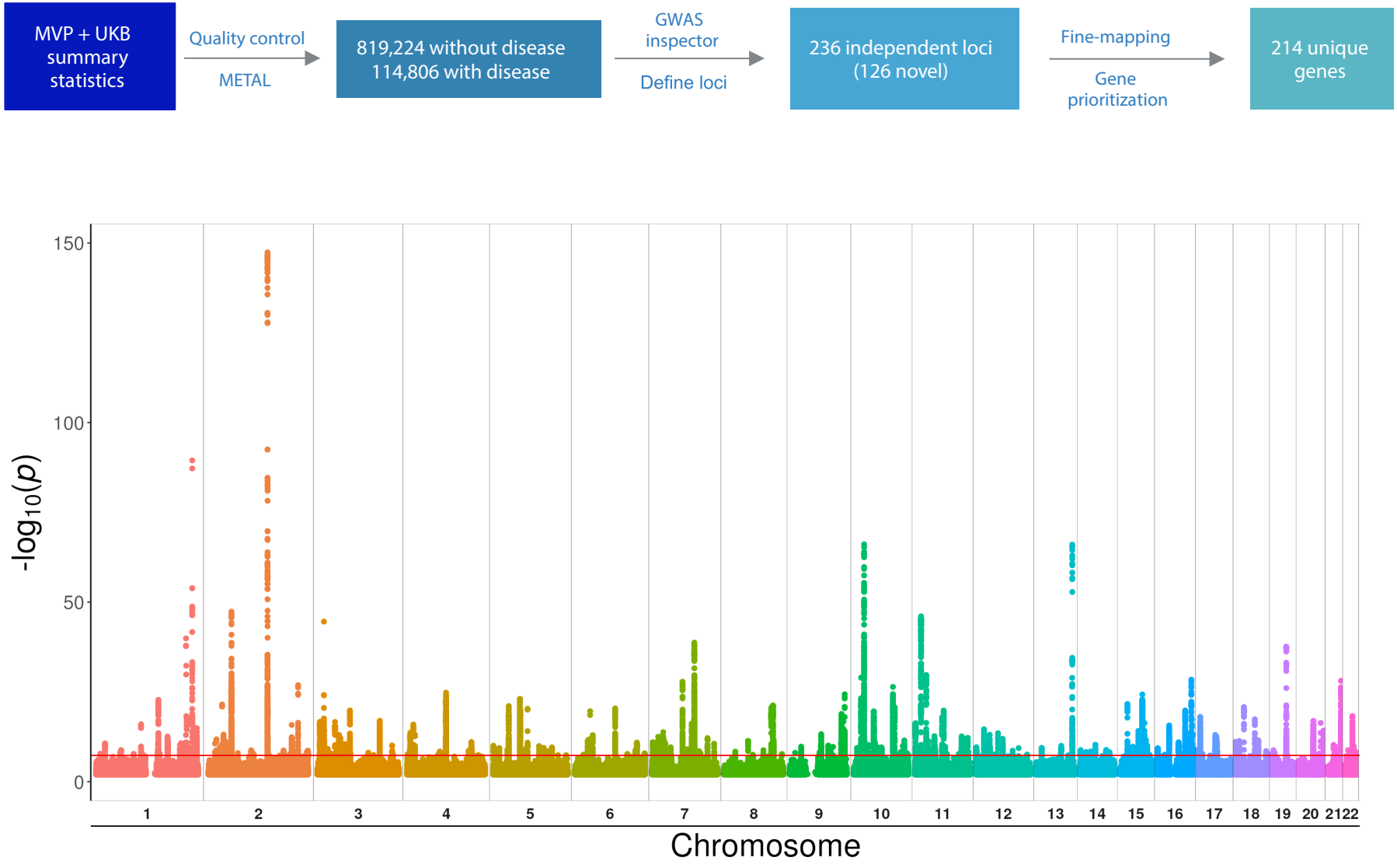
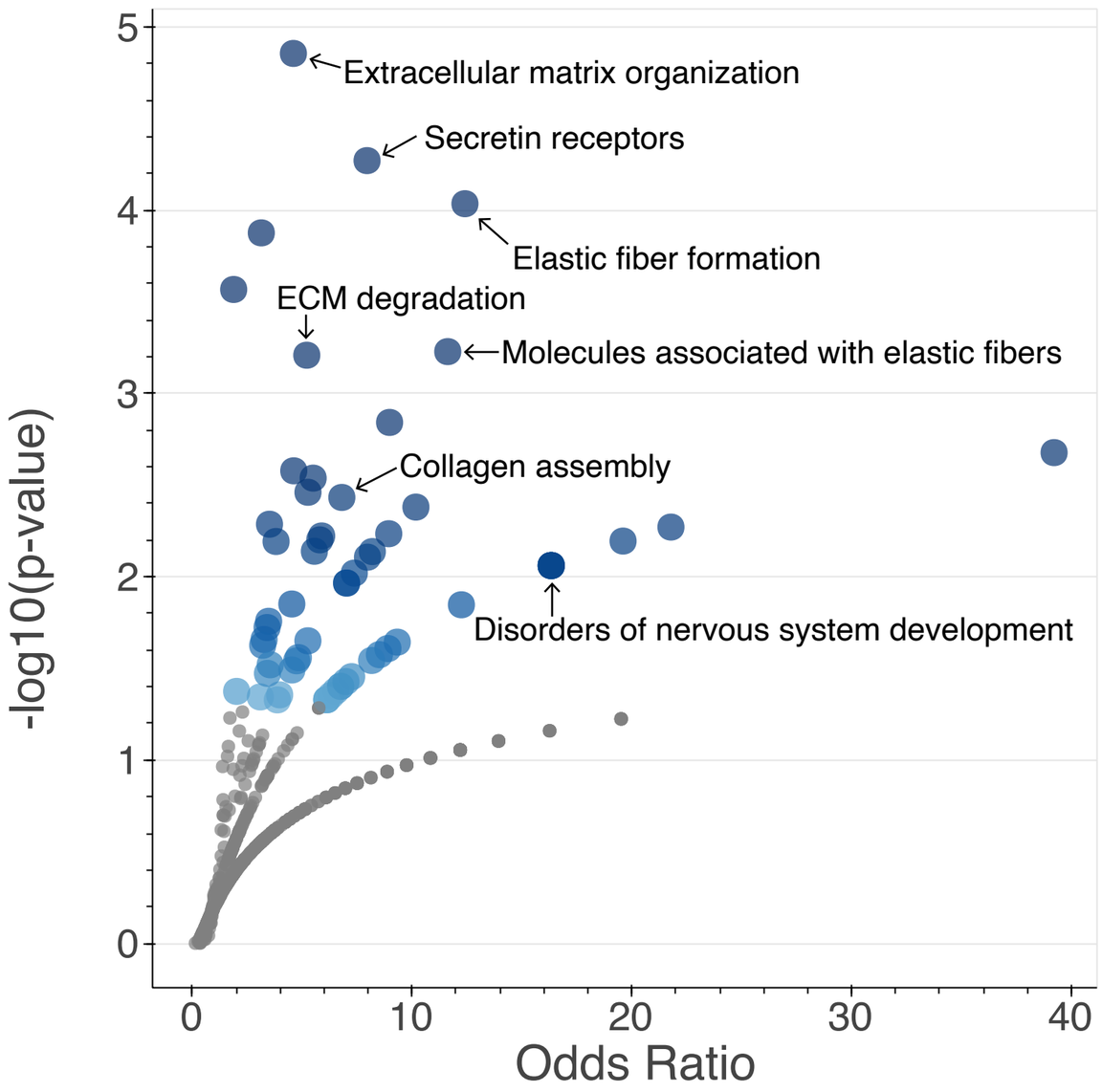
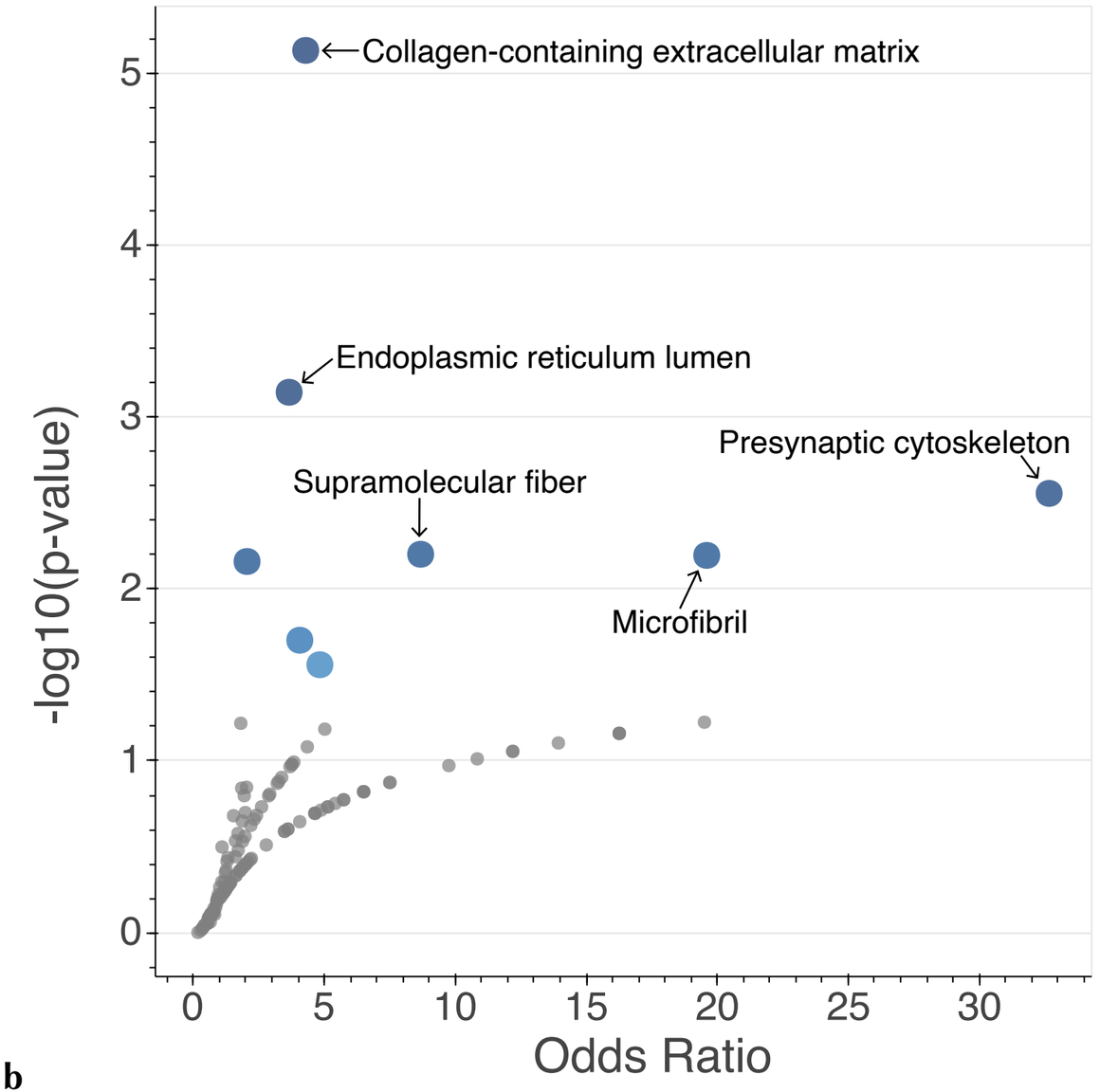


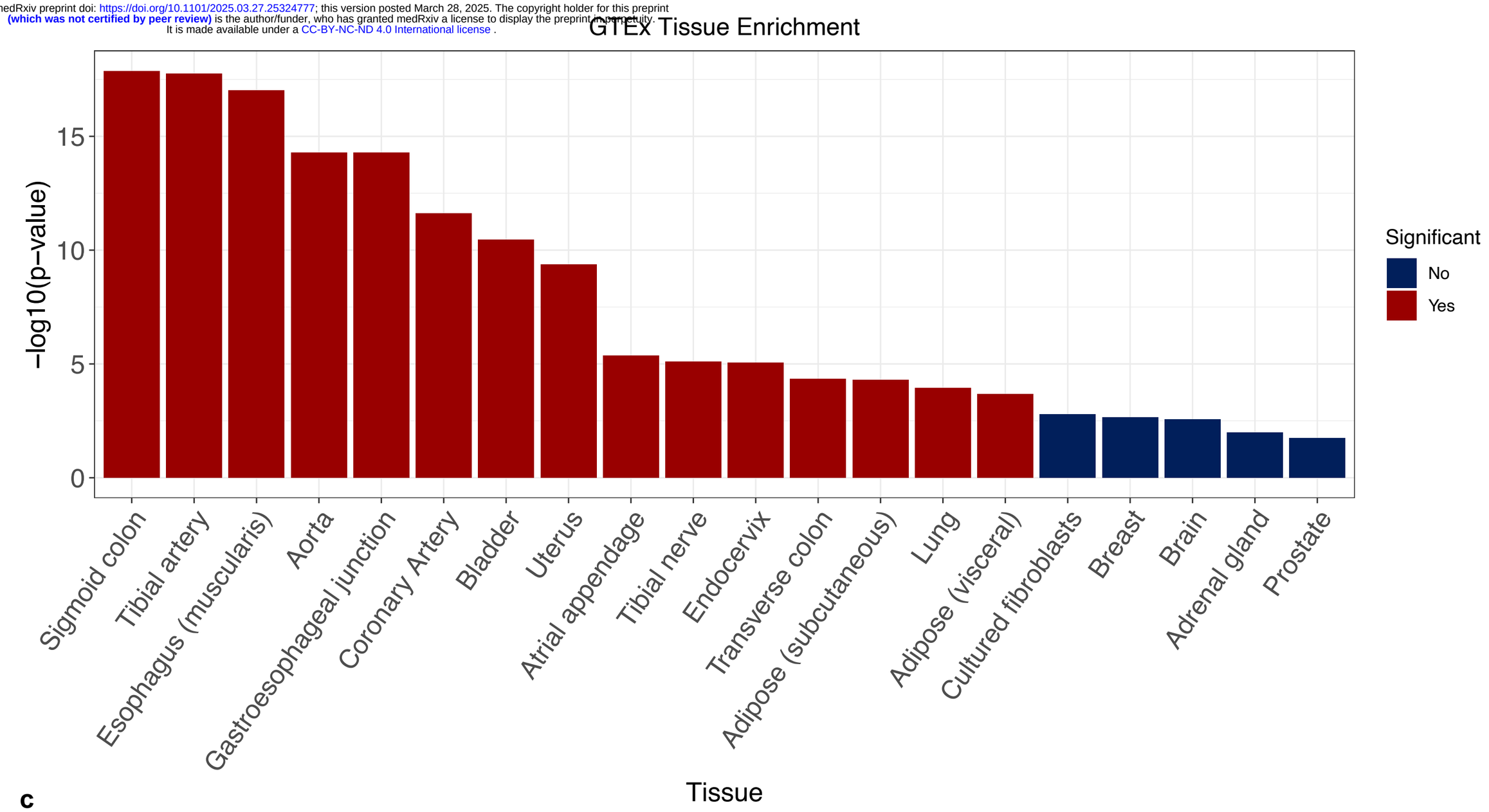
Figure 1. Study overview. a, Flowchart of GWAS and gene prioritization. **b,** Manhattan plot demonstrates results of GWAS meta-analysis.



a



b



c

Figure 2. Enrichment analysis. Prioritized gene enrichment in **a,** Reactome Pathway (2024), **b,** GO Cellular Component (2023), and **c,** GTEx (v8) tissues. In volcano plots (a and b), each dot represents a term from target gene set in which the prioritized genes were enriched; blue dots are significant (p-value < 0.05); ECM = extracellular matrix. In c, significance is determined by adjusted p-value < 0.05.

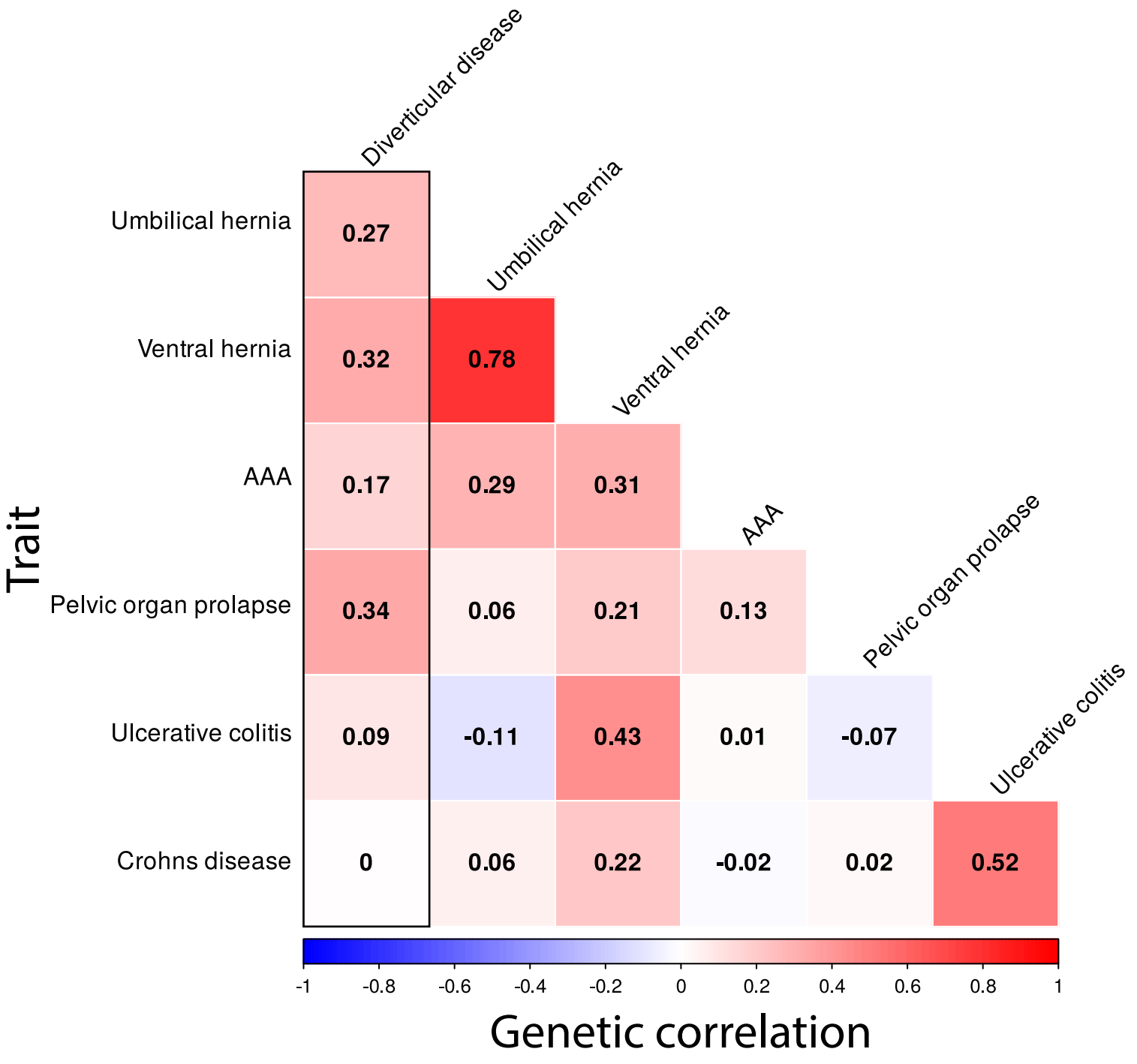


Figure 3. LD score regression shows genetic correlation between diverticular disease and other putative connective tissue disease phenotypes.

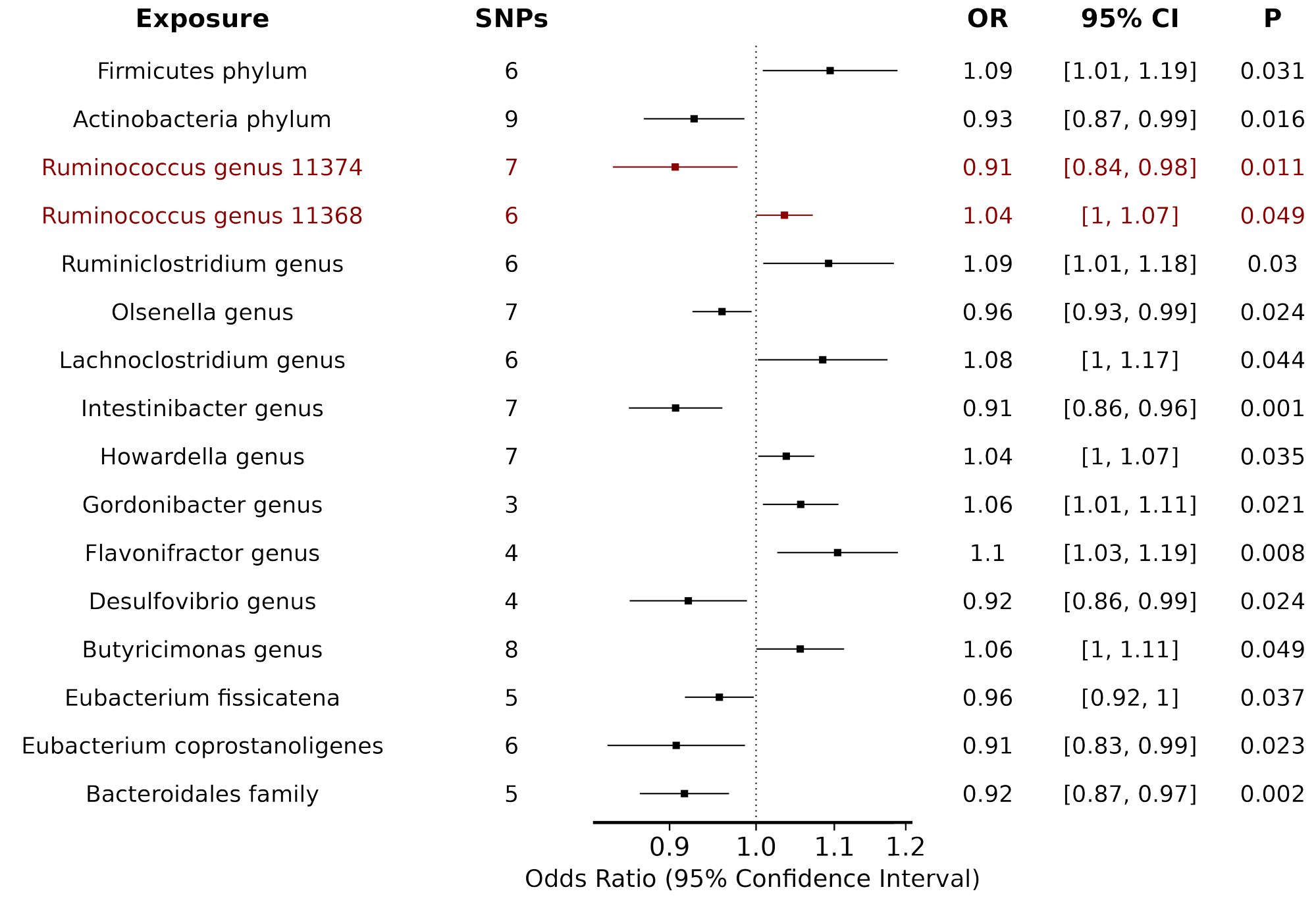


Figure 4. Microbiome Mendelian randomization (MR) results. In all cases, inverse variance weighted MR was performed for the outcome of diverticular disease. Odds ratio > 1 indicates association with increased diverticular disease liability.

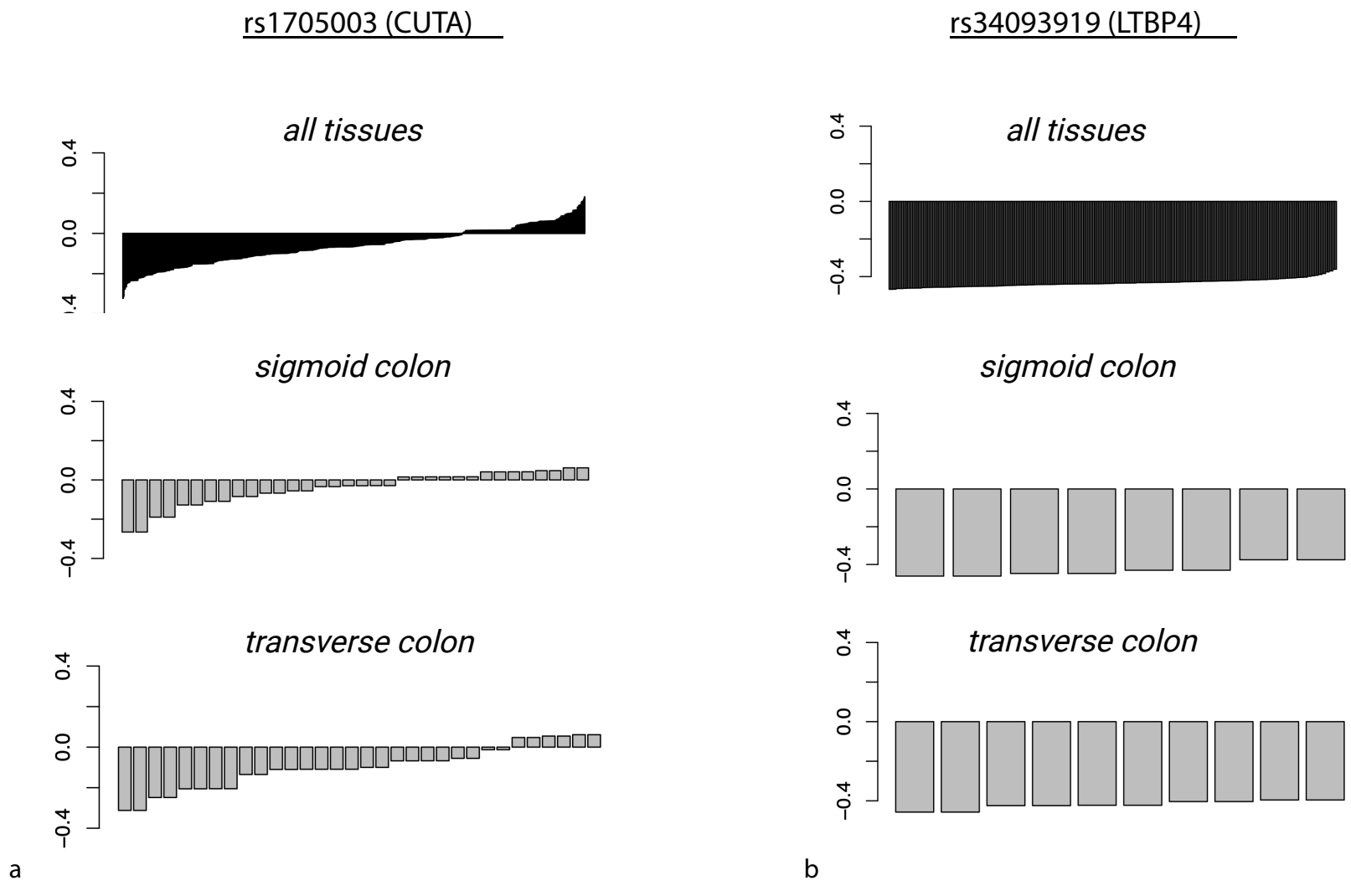


Figure 5. Allele specific expression in GTEx of two genetic variants (**a** and **b**) across multiple tissues. Each bar in the graph shows expression in a single individual. Y-axis reveals the relative abundance of expression (range -0.5, 0.5) where |0.5| demonstrates exclusive expression of one allele.

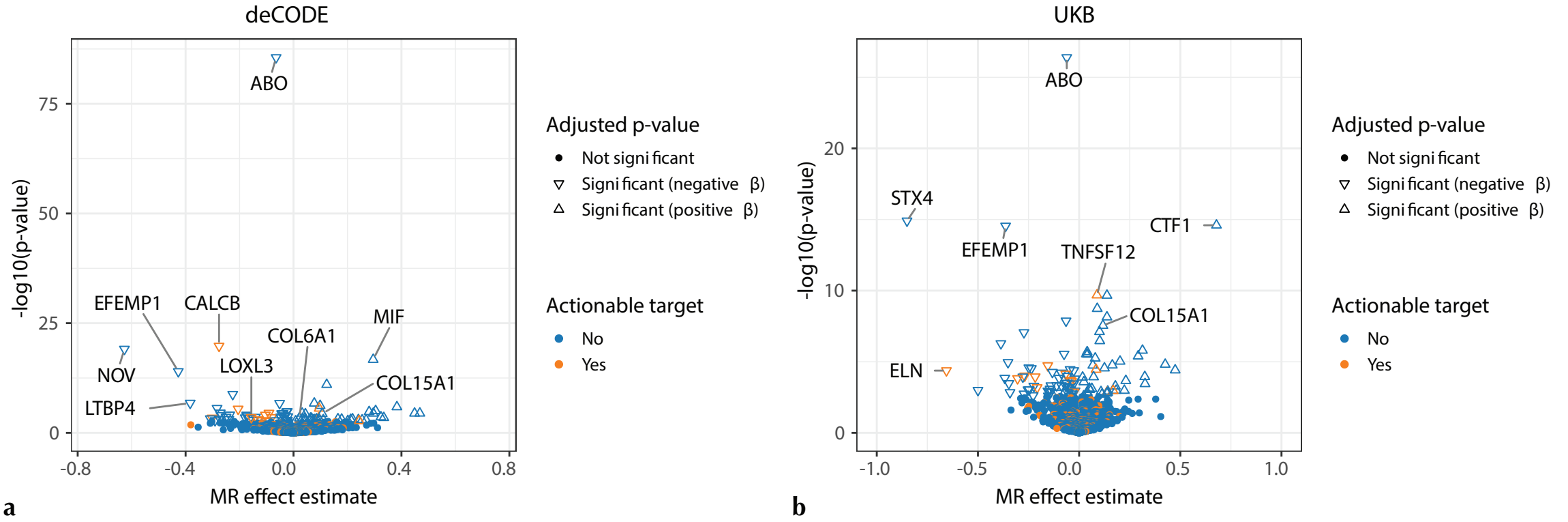
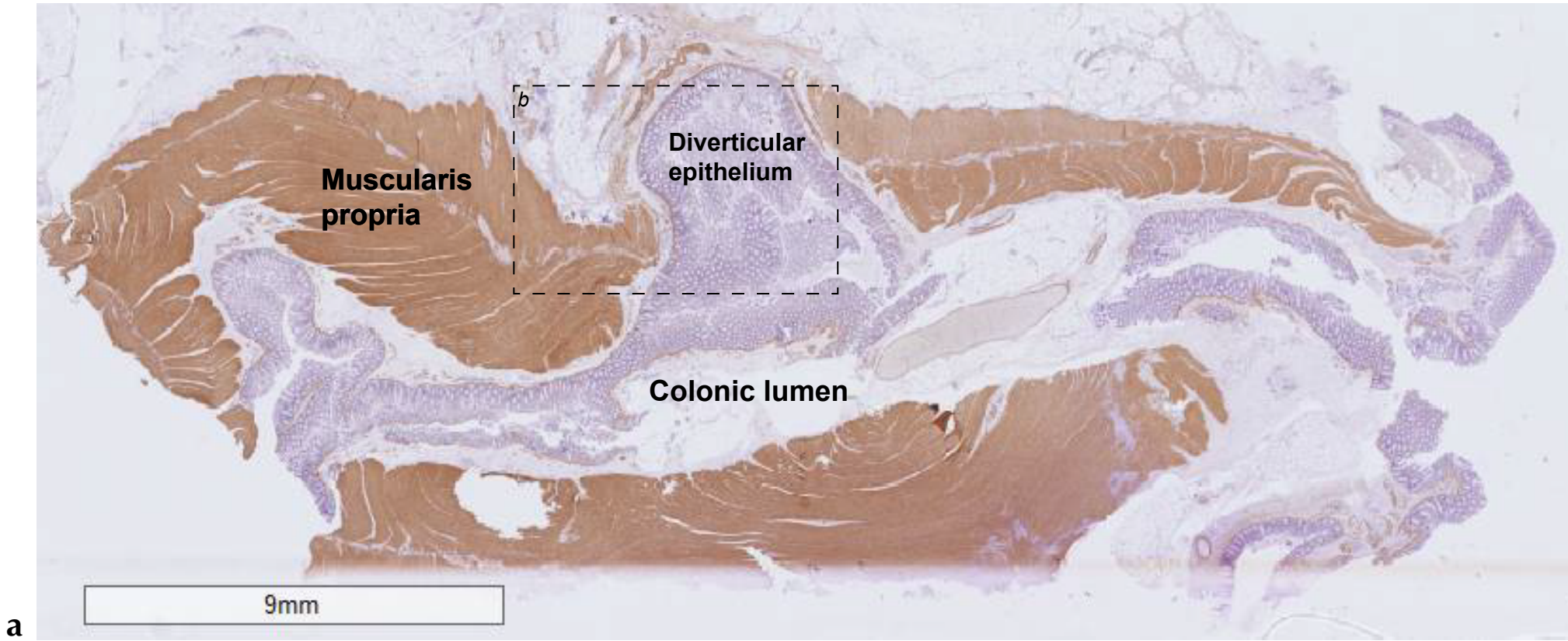
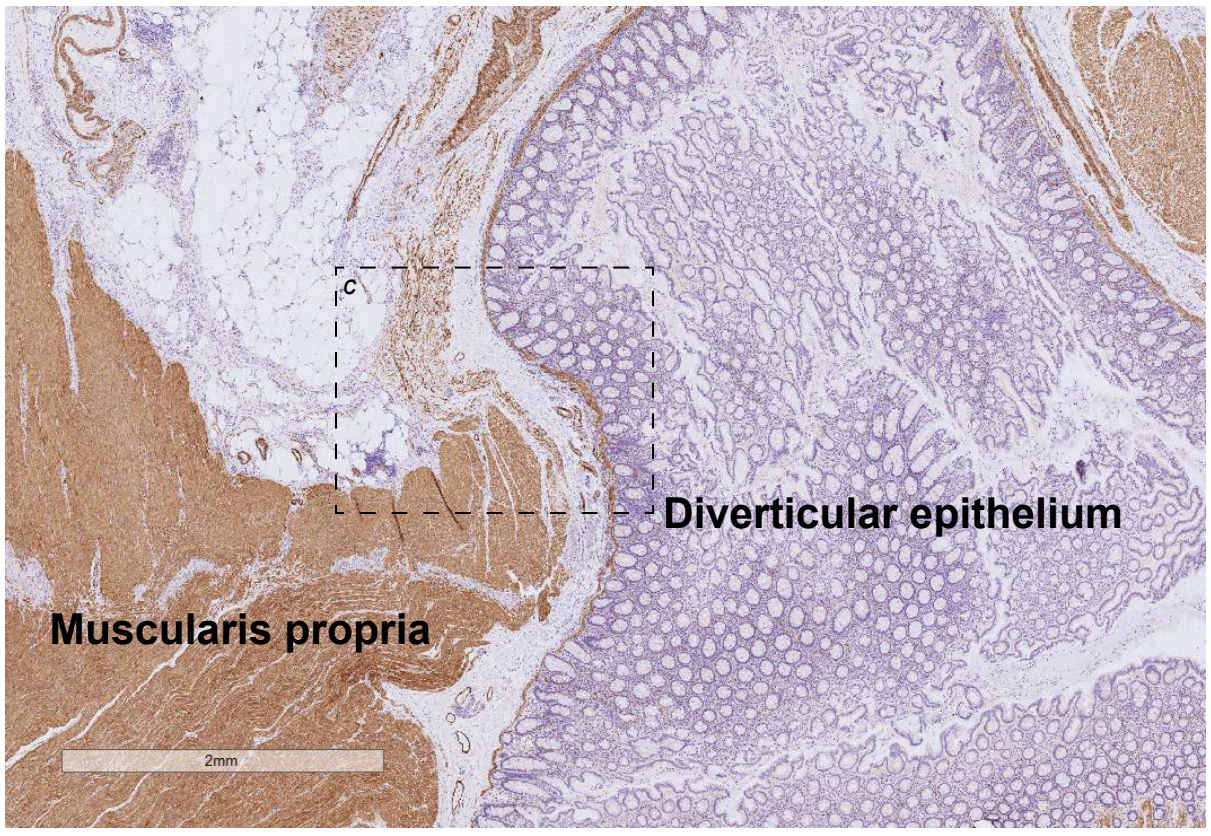


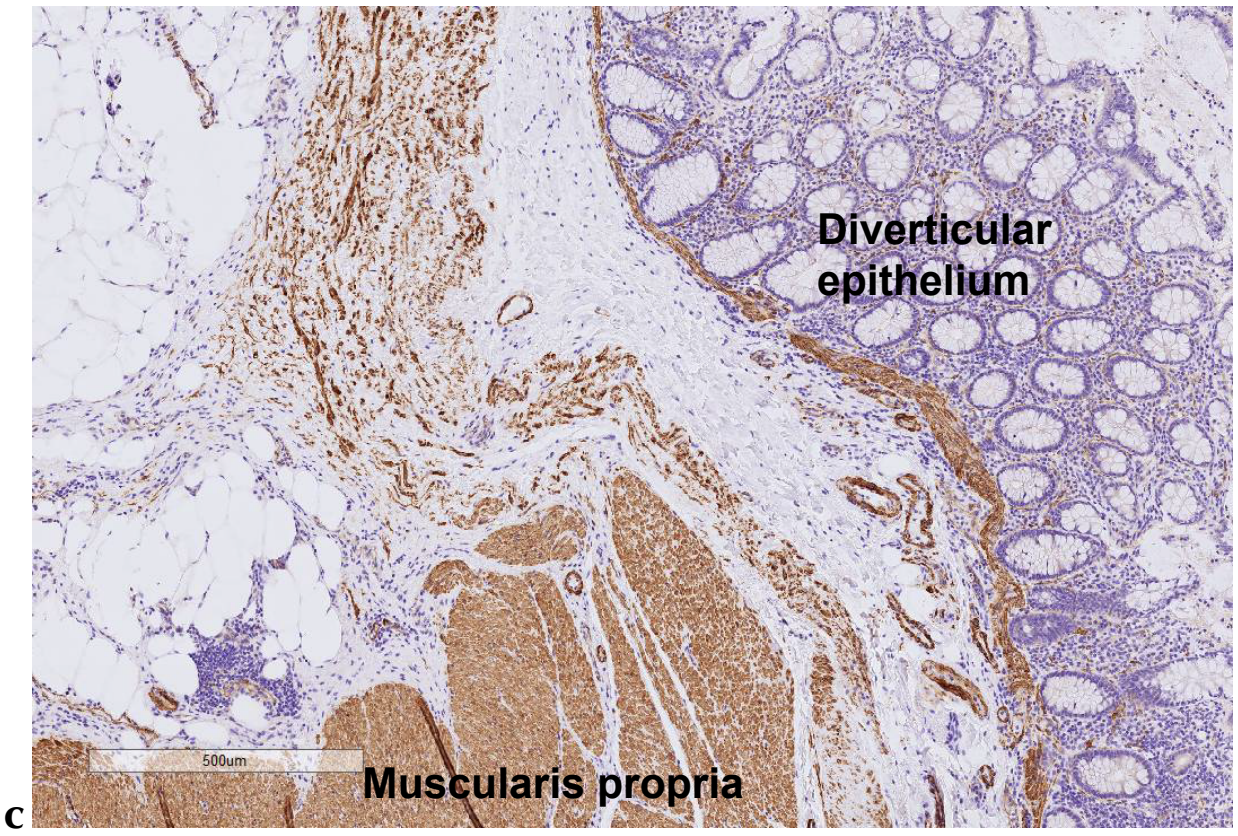
Figure 6. pQTL Mendelian randomization identifies 157 proteins associated with diverticular disease in the **a,** deCODE and **b,** UKB-PPP databases. MR effect estimate is the natural logarithm of the odds ratio. Positive beta means increase serum protein abundance is positively associated with diverticular disease.



a

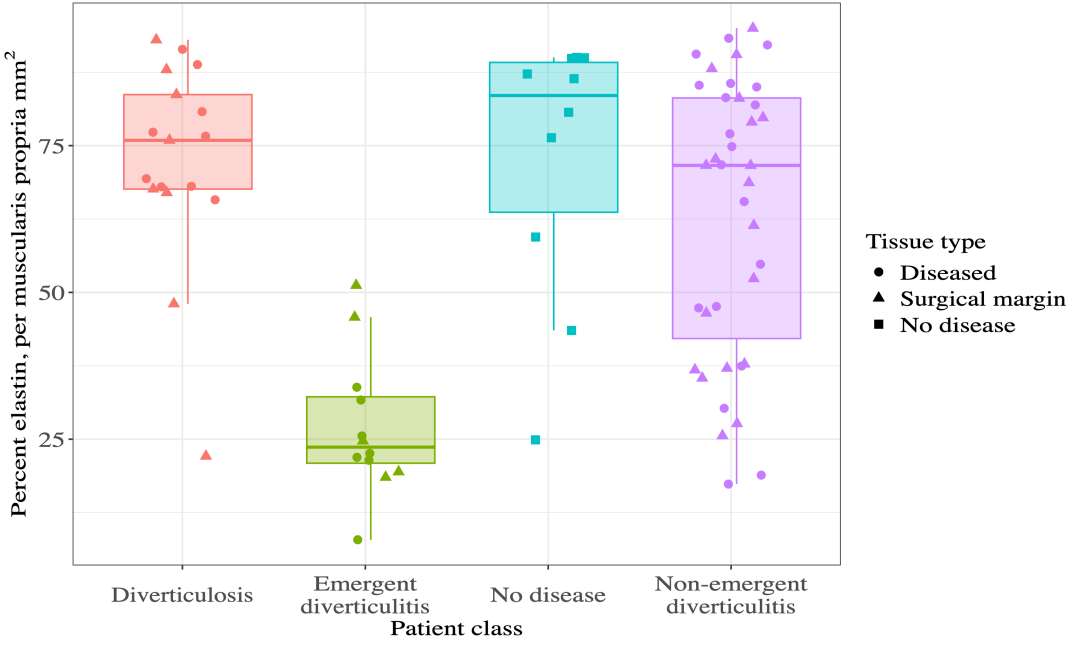


b



c

Figure 7. MYH11 antibody staining (brown) of colonic tissue resected during surgery from a patient with diverticular disease, demonstrating thinning of the muscularis propria overlying the diverticulum at **a,** 1x magnification, **b,** 5x magnification, and **c,** 20x magnification. The area outlined with dashed lines in panel a shows the borders of panel b, and the area outlined with dashed lines in panel b shows the borders of panel c. Hematoxylin (purple) was used as a counter-stain.



a

Characteristic	Beta	95% CI[1]	p-value
Sex	-1.2	-14, 11	0.9
Age (years)	0.14	-0.33, 0.60	0.6
Patient Classification			
No disease	—	—	
Diverticulosis[2]	1.2	-21, 23	>0.9
Emergent diverticulitis[3]	-41	-66, -17	0.001
Non-emergent diverticulitis	-10	-26, 5.5	0.2

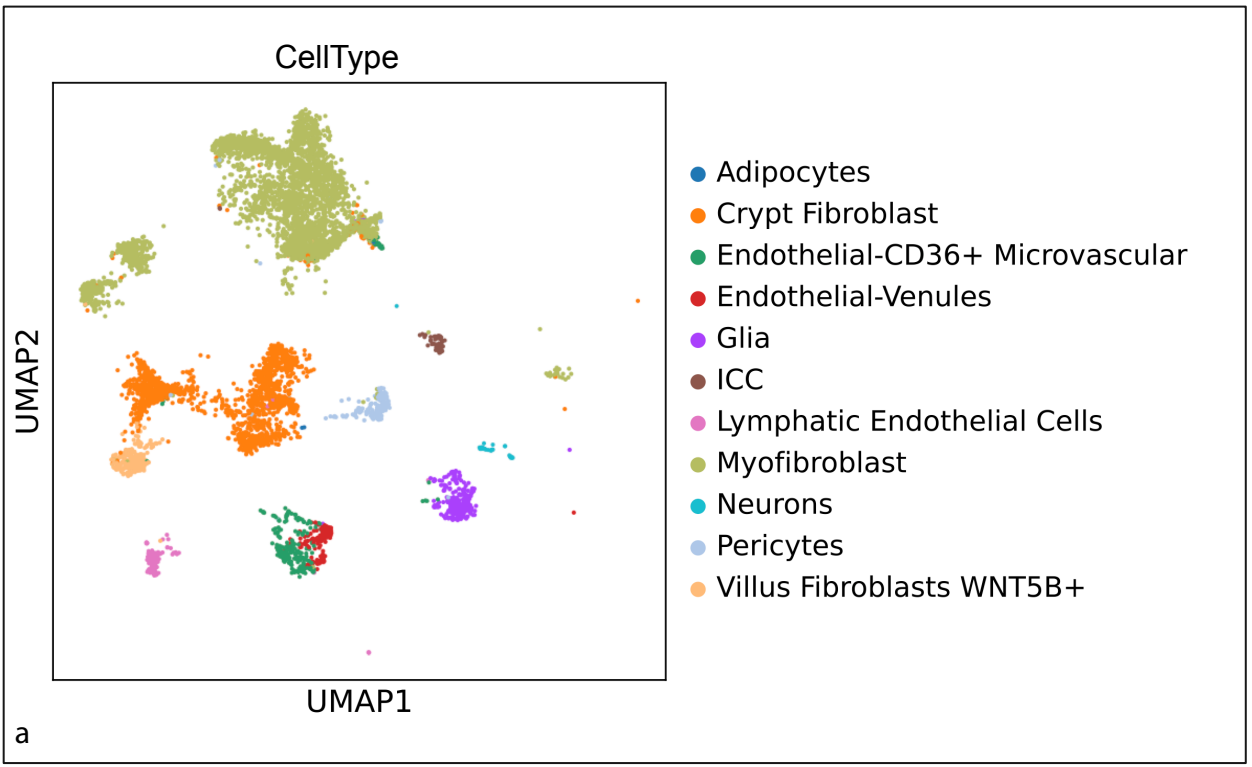
[1] CI = Confidence Interval

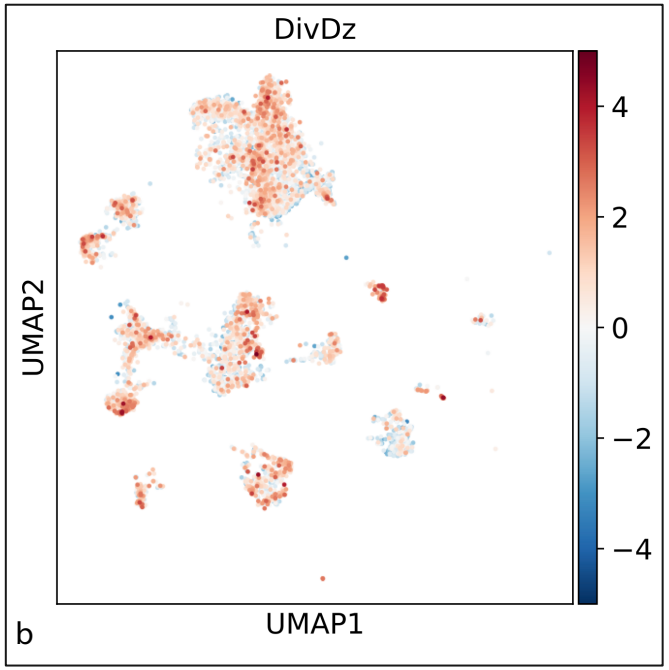
[2] Tissue from peri-diverticular area

[3] Tissue from surgical margin

b

Figure 8. Elastin quantification. a, Each dot represents the measured elastin density on a prepared slide, with boxplot showing median elastin density (horizontal line) by patient class, interquartile range (IQR; the borders of the box) and whiskers extending 1.5*IQR from the end of the box. **b,** Multivariable linear regression showing a 41% reduction in elastin density among patients undergoing emergency diverticulitis surgery compared to those without disease. Peri-diverticular area corresponds to location 2 in Extended Data Fig.6; surgical margin corresponds to location 5 in Extended Data Fig.6.





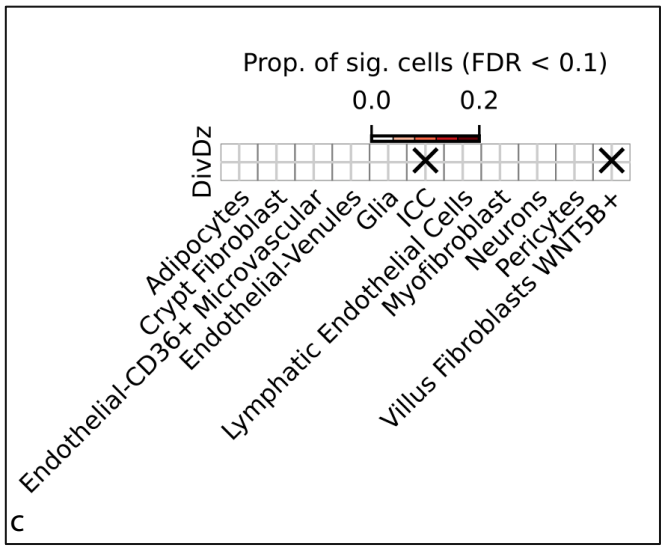


Figure 9. Cell-level localization of diverticular disease-associated genes obtained with scDRS in the Hickey et al. single-cell RNAseq dataset showing **a)** Dimensional reduction of analyzed cells, **b)** Relative cell-level expression of diverticular-disease associated genes, **c)** Cell type-gene expression associations, with X denoting significantly (FDR < 0.01) increased expression of diverticular disease (DivDz)-associated genes in the indicated cell type. Boxes in white do not have a high proportion of cells with significant (Prop. of sig. cells) gene expression enrichment.

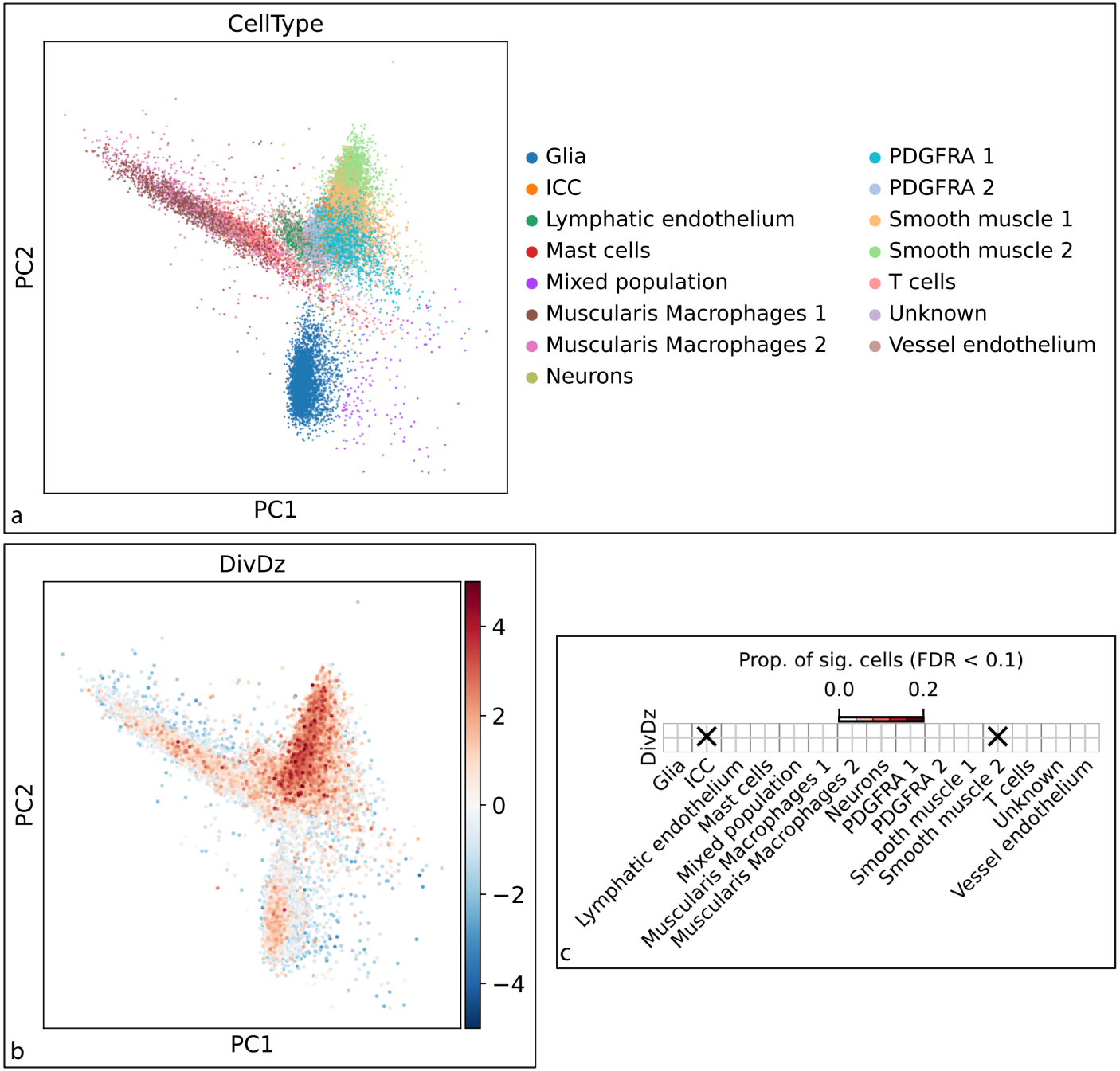


Figure 10. Cell-level localization of diverticular disease-associated genes obtained with scDRS in the Wright et al. single-cell RNAseq dataset showing **a)** Dimensional reduction of analyzed cells, **b)** Relative cell-level expression of diverticular-disease associated genes, **c)** Cell type-gene expression associations, with X denoting significantly (FDR < 0.01) increased expression of diverticular disease (DivDz)-associated genes in the indicated cell type. Boxes in white do not have a high proportion of cells with significant (Prop. of sig. cells) gene expression enrichment.

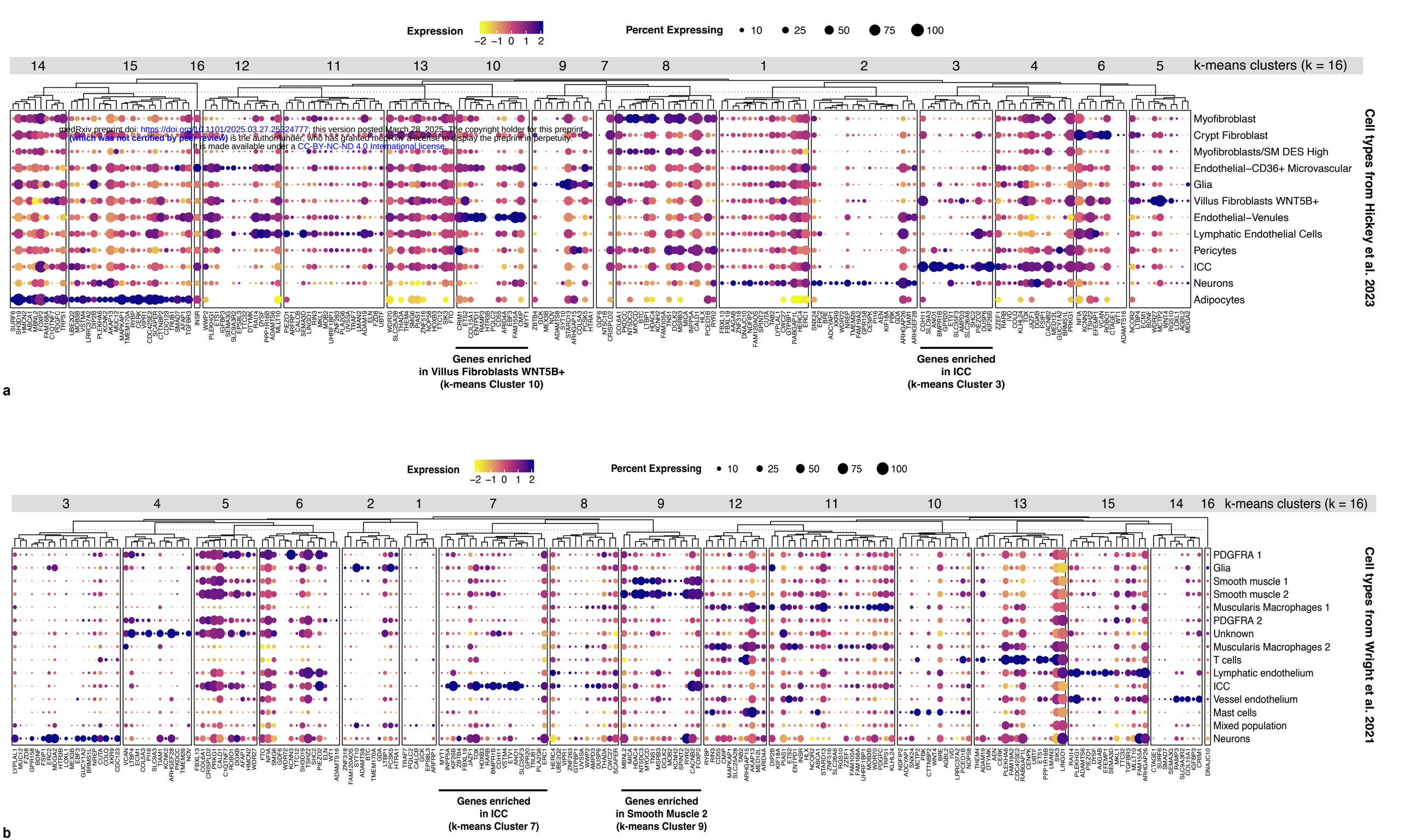
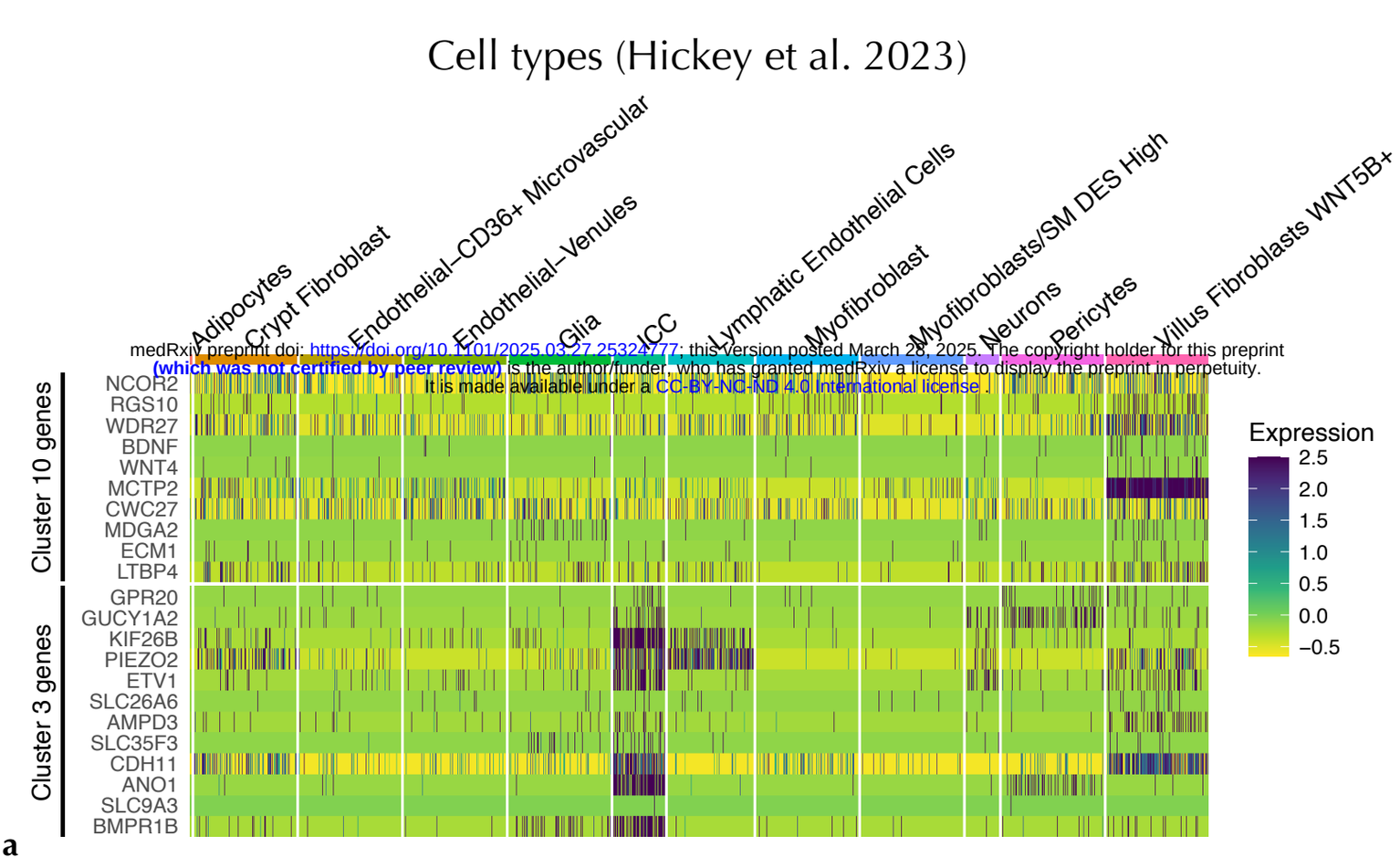
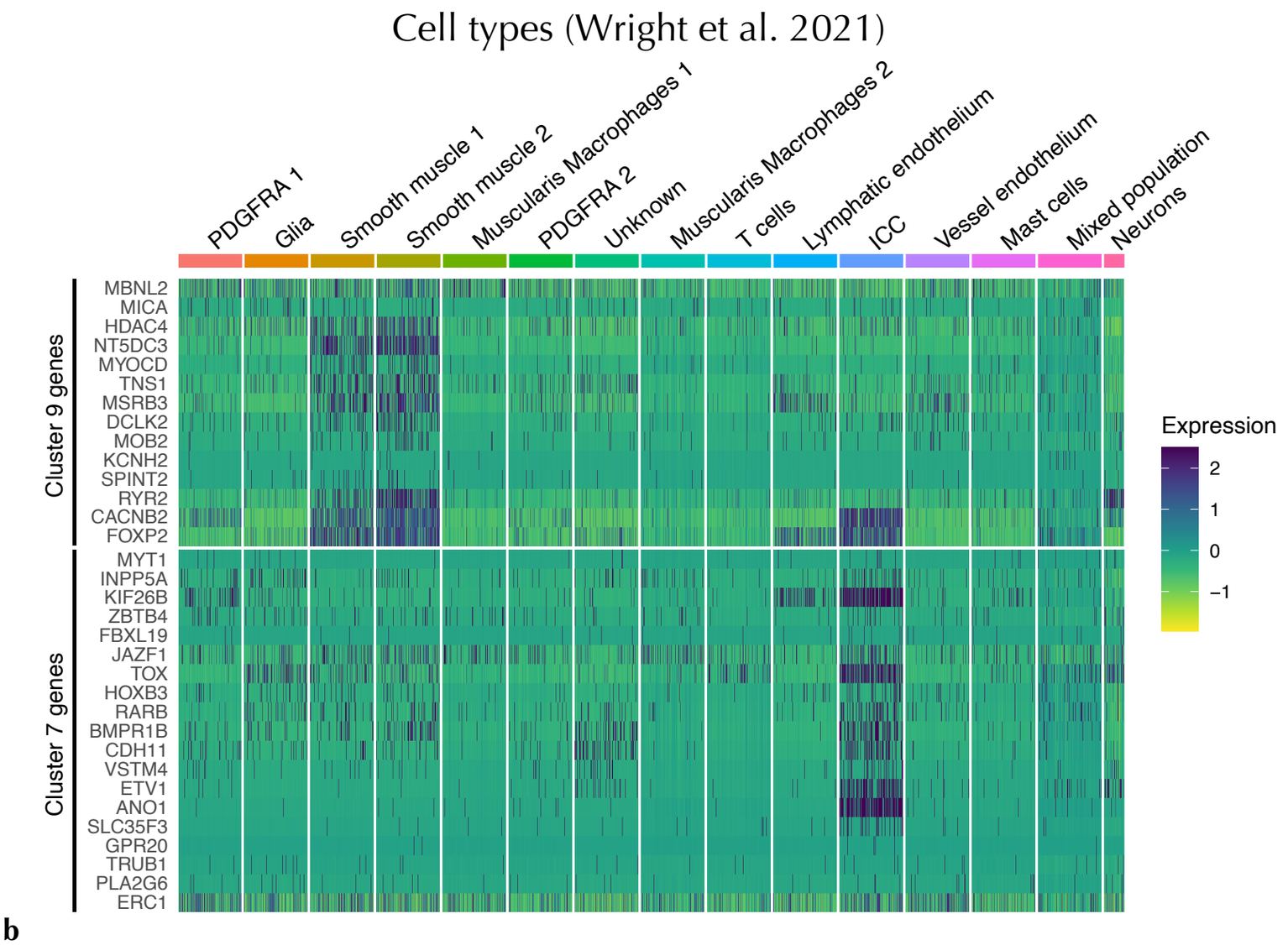


Figure 11. Dot plots showing expression patterns of diverticular disease-associated genes with k-means clustering applied, in the **a,** Hickey et al. and **b,** Wright et al. scRNAseq datasets. Mean expression for each gene across all samples is set to zero, and scaled expression is shown as standard deviations above or below the mean.



a



b

Figure 12. Heatmaps showing expression of clusters of genes in **a,** Hickey et al. WNT5B+ fibroblasts (cluster 10) and ICC (cluster 3) and **b,** Wright et al. smooth muscle cells (cluster 9) and ICC (cluster 7). Mean expression for each gene across all samples is set to zero, and scaled expression is shown as standard deviations above or below the mean.

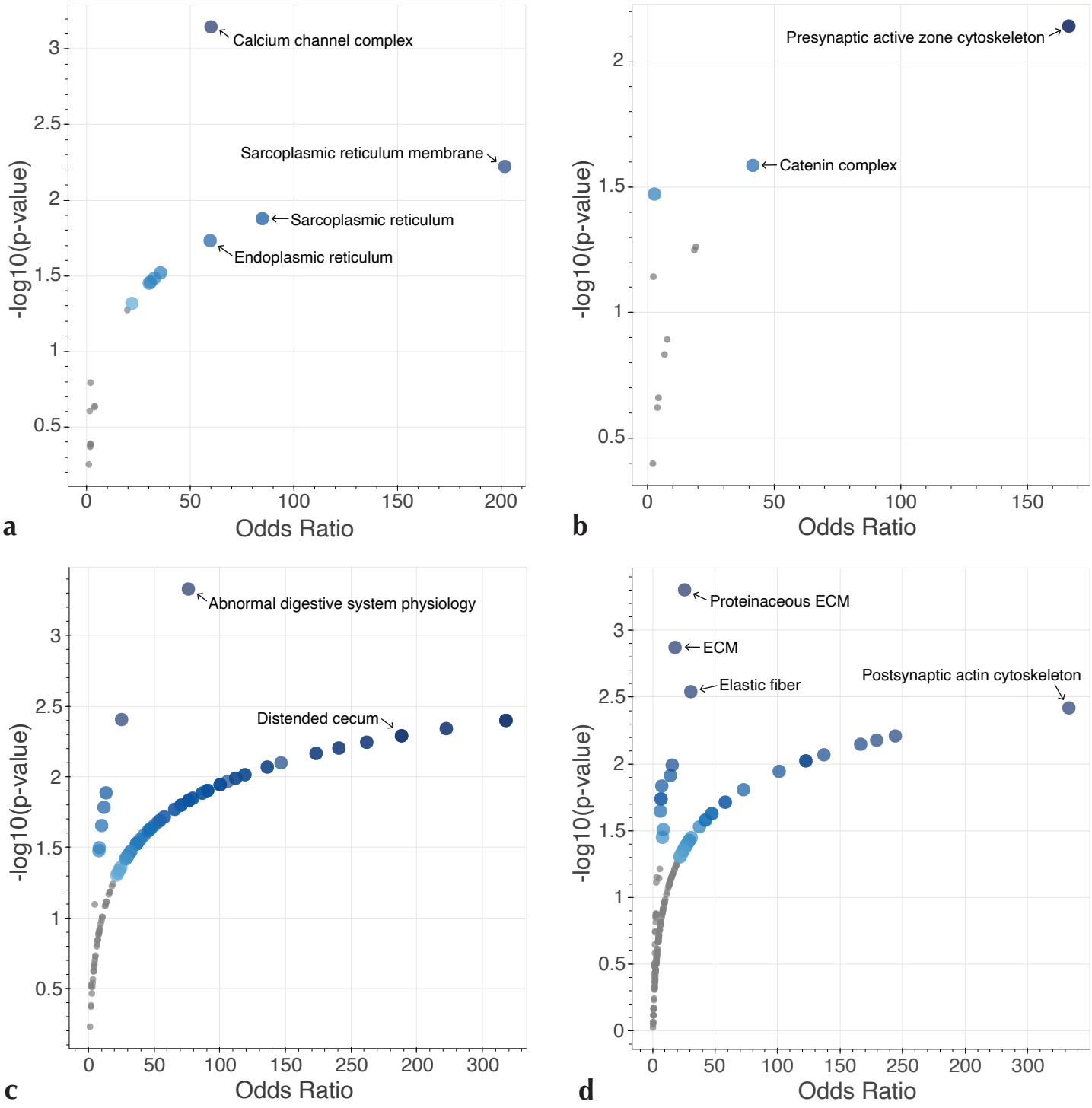


Figure 13. Pathway analysis of clusters of genes enriched in **a,** smooth muscle cells (cluster 9, Wright et al.), in gene ontology (GO) cellular component 2023, **b,** interstitial cells of Cajal (ICC) (cluster 7, Wright et al.), in GO cellular component 2023, **c,** ICC (cluster 3, Hickey et al.), in the MGI mammalian phenotype enrichr pathway, **d,** WNT5B+ fibroblasts (cluster 10, Hickey et al.), in the Jensen compartments enrichr pathway

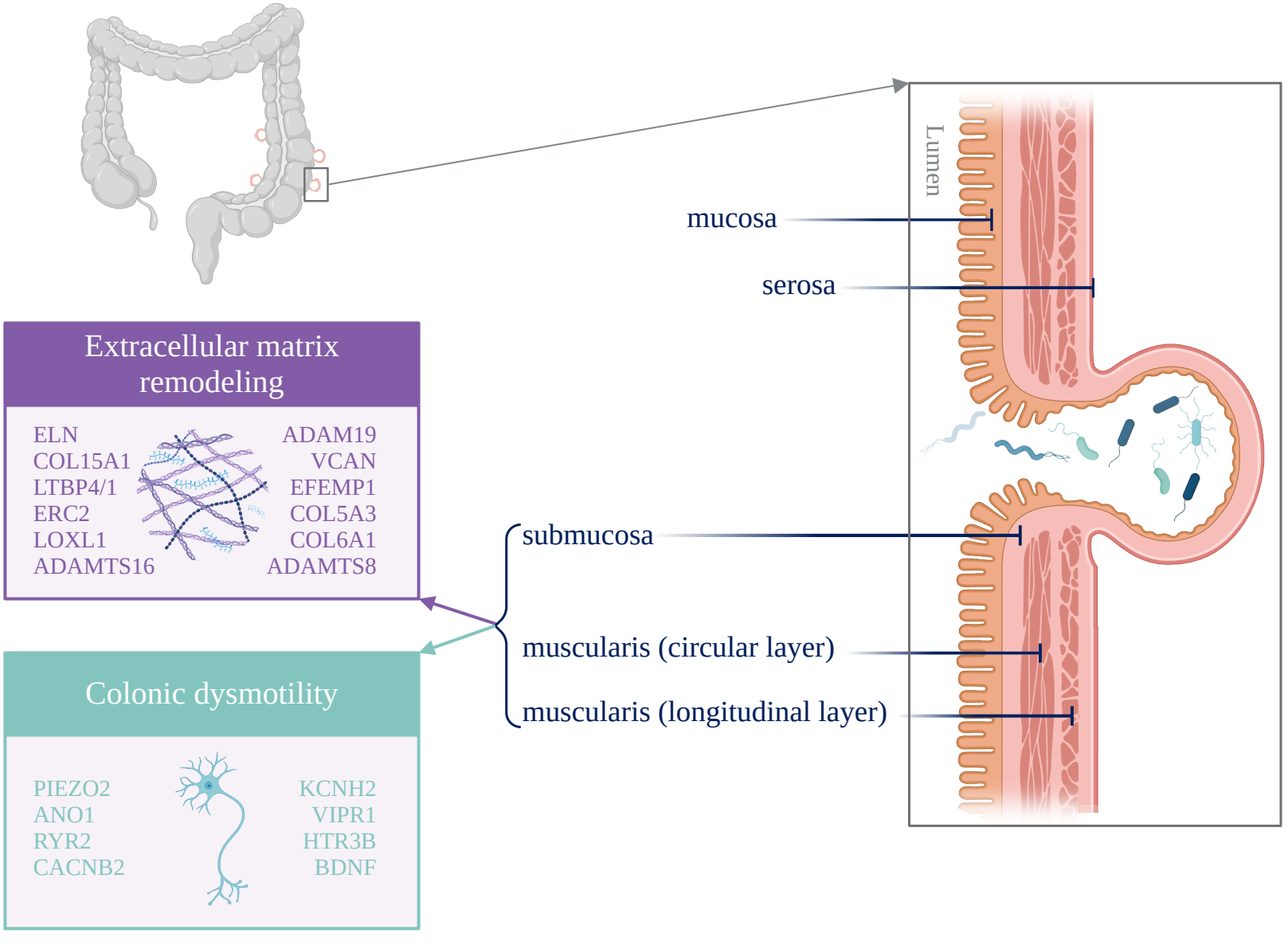


Figure 14. Proposed biological changes through which the genes identified in this GWAS may increase liability to diverticular disease.

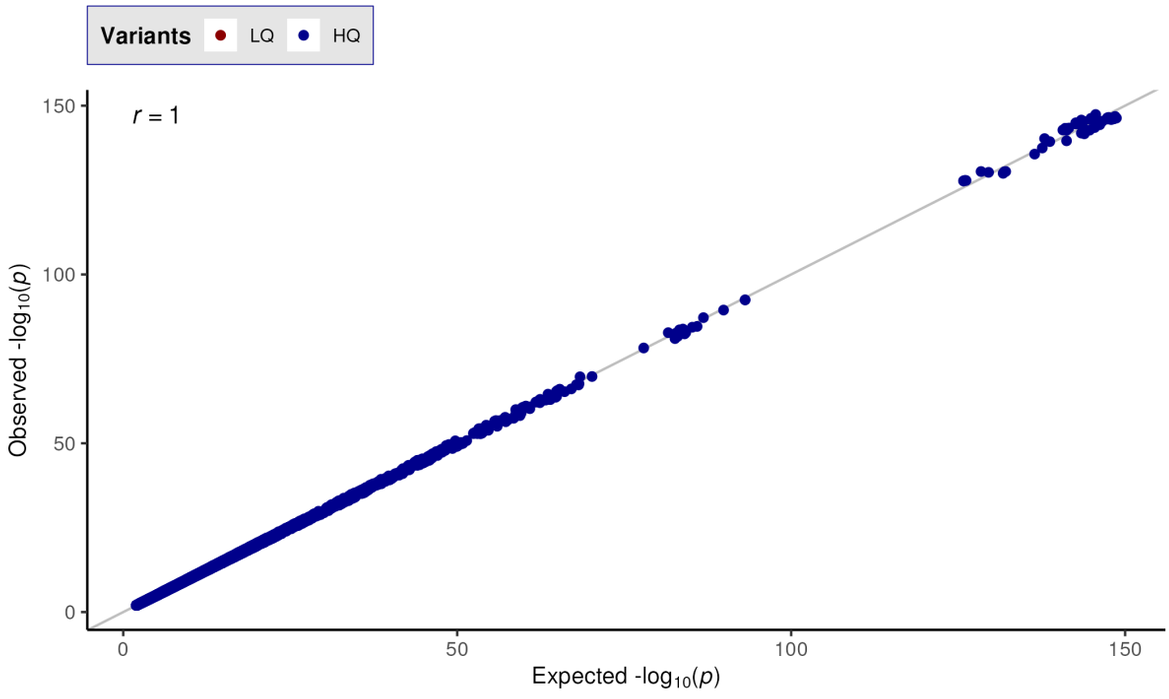
SUPPLEMENTAL DATA

Quality Check Report

GWASinspector package

P-value correlation (observed vs expected)

	Value
Included variants	44,975,206 (100%)
r	1

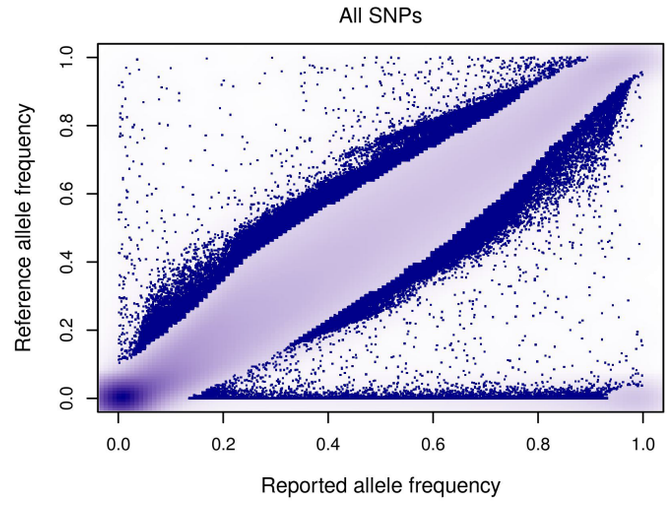


Distribution statistics

	Value
Skewness	-1.39
Skewness (HQ)	1.14
Kurtosis	110
Kurtosis (HQ)	62.4
Visscher's stat	91.1
Visscher's stat (HQ)	91.1
Lambda - total	1.16

* **HQ** = High-Quality variants

Result of matching with reference dataset



Supplemental Figure 1. GWASinspector results showing quality control metrics for the GWAS meta-analysis.

MYH STAINING PROTOCOL

Five-micron sections of formalin-fixed paraffin-embedded tissue were stained using antibody against _MYH11 from Sigma cat# HPA015310, at 1:100). Staining was done on a Leica Bond-IIITM instrument using the Bond Polymer Refine Detection System (Leica Microsystems DS9800). Heat-induced epitope retrieval was done for 20 minutes with ER1 solution (Leica Microsystems, AR9961). The experiment was done at room temperature. Slides are washed three times between each step with bond wash buffer or water.

ELASTIC STAINING PROTOCOL

Specimen Criteria: appropriately fixed, paraffin embedded, 4-6 µm tissue section
Control: Aorta, Normal Lung

PSCS – Elastic Stain Protocol with Verhoeff's Elastic Stain Kit, Cat# KTVELPT

1. Deparaffinize in Xylene – 3 changes @ 4 mins each.
2. Rinse in 100% EtOH – 3 changes @ 3 mins each.
3. Rinse in Running Tap Water – 1 min.
4. Immerse in Verhoeff's Elastic Stain – 15 mins.
5. Rinse in Running Tap Water – 1 min.- Check the staining intensity
6. Immerse in 2% Ferric Chloride – pour in and out.
7. Rinse in Running Tap Water – 1 min. Check the staining intensity
8. Immerse in 5% Sodium Thiosulfate – pour in and out.
9. Rinse in Running Tap Water – 1 min.
10. Immerse in Van Gieson's Stain – pour in and out- Check the staining intensity
11. Dehydrate in 100% EtOH – 3 changes @ pour in and out.
12. Clear in Xylene – 3 changes @ 30s each.
13. Coverslip according to protocol.

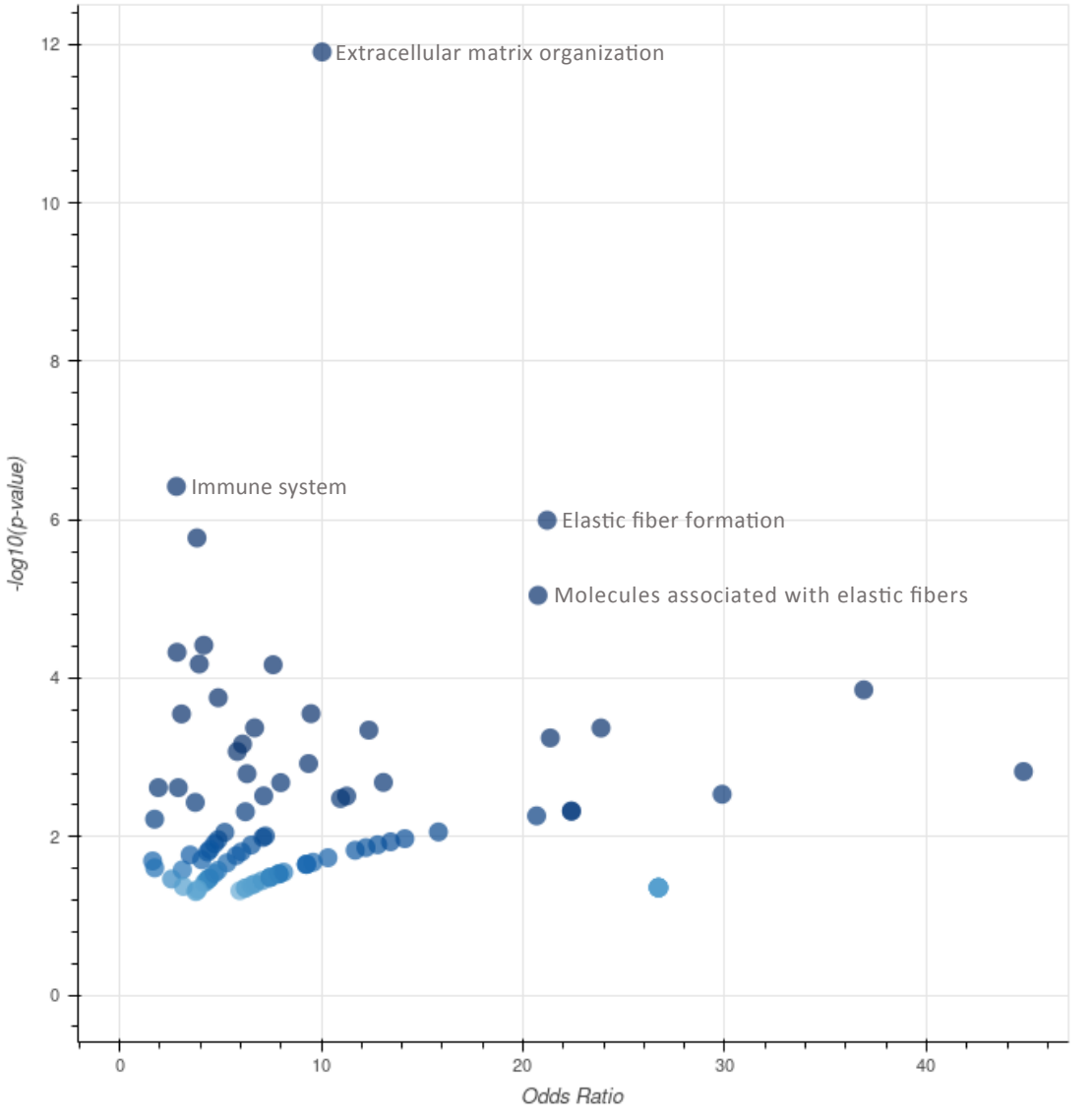
Preparation of Verhoeff's Elastic Stain:
Solution expires after 1 use. Prepare just before use.

1. Add 40 mL of 5% Alcoholic Hematoxylin.
2. Add 20 mL of 10% Ferric Chloride & mix.
3. Add 20 mL of Universal Iodine Solution & mix.
4. Filter using a paper towel or filter paper.

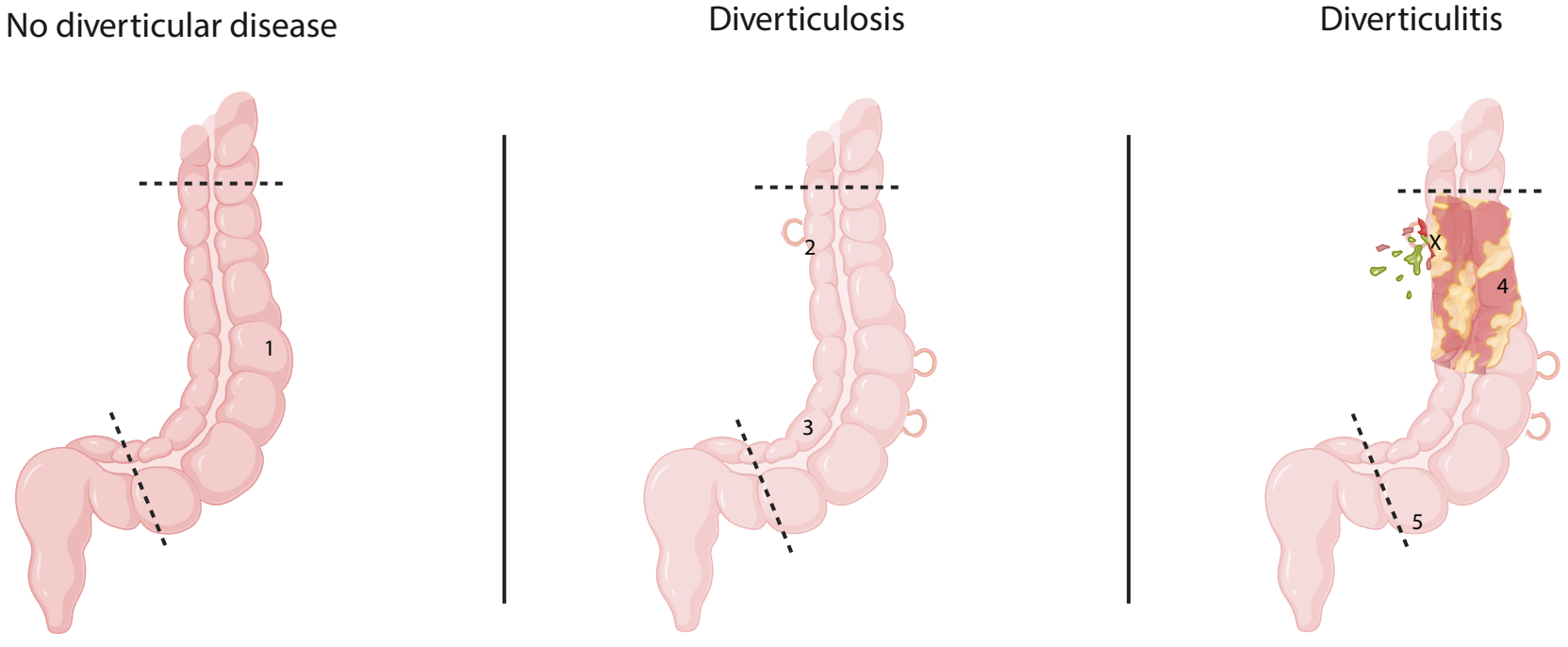
Preparation of 2% Ferric Chloride:
Solution expires after 1 use. Prepare just before use.

1. Add 40 mL of DI water.
2. Add 10 mL of 10% Ferric Chloride & mix.

Supplemental Figure 2. Immunohistochemistry staining protocols. (Note: only first page displayed.)



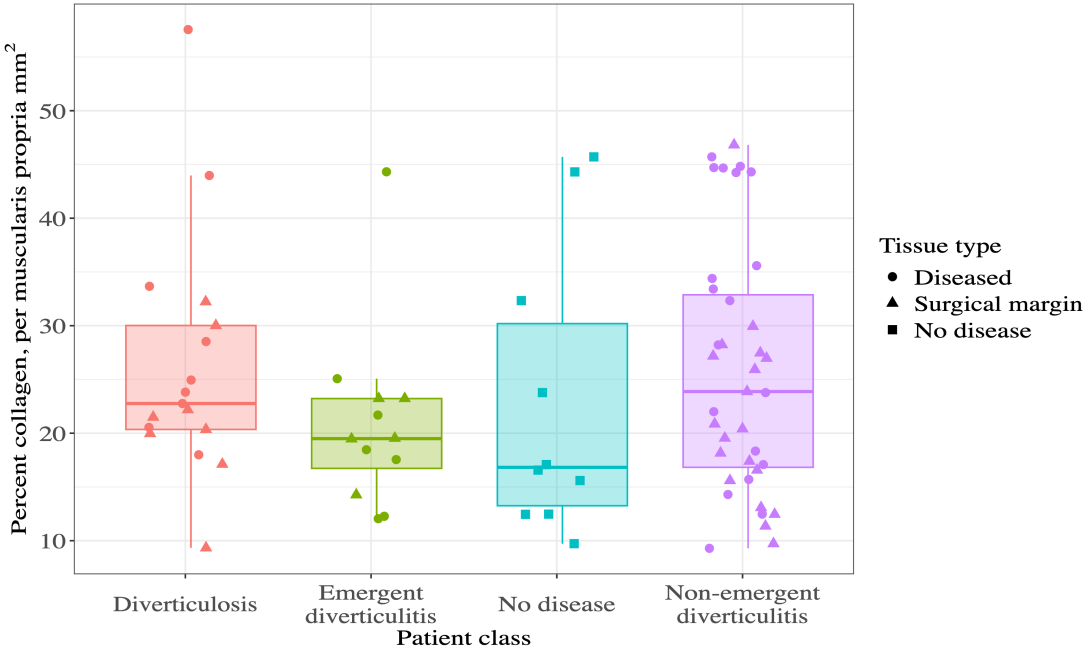
Supplemental Figure 3. Enrichment analysis of significant pQTL MR proteins in the Reactome database. Blue dots are significant (p-value < 0.05).



Legend

Dotted lines denote start and end of surgical specimen (cut ends of the colon)
1 - Site of grossly healthy sample
2 - Site of peri-diverticular sample
3 - Site of grossly healthy sample
X - Location of perforated diverticulum, with spillage of intraluminal contents and surrounding inflammation (diverticulitis)
4 - Site of sample in region of gross disease (i.e. within area of diverticulitis)
5 - Site of sample at surgical margin (i.e. furthest point away from location of diverticulitis)

Supplemental Figure 4. Locations of samples taken from surgical specimens for immunohistochemistry staining and elastin quantification.



a

Characteristic	Beta	95% CI[1]	p-value
Sex	-0.57	-7.3, 6.1	0.9
Age (years)	-0.02	-0.27, 0.23	0.9
Patient Classification			
No disease	—	—	
Diverticulosis[2]	7.2	-4.7, 19	0.2
Emergent diverticulitis[3]	-3.0	-16, 10	0.6
Non-emergent diverticulitis	2.6	-5.9, 11	0.5

[1] CI = Confidence Interval

[2] Tissue from peri-diverticular area

[3] Tissue from surgical margin

b

Supplemental Figure 5. Collagen quantification a, Each dot represents the measured collagen density on a prepared slide, with boxplot showing median collagen density (horizontal line) by patient class, interquartile range (IQR; the borders of the box) and whiskers extending 1.5*IQR from the end of the box. **b,** Multivariable linear regression showing no significant difference in collagen density among patients undergoing emergency diverticulitis surgery compared to those without disease. Peri-diverticular area corresponds to location 2 in Extended Data Fig.6; surgical margin corresponds to location 5 in Extended Data Fig.6.

CHR	POSITION	rsid	EFFECT_ALL	OTHER_ALL	STDERR	Pval	EFFECT	Novel
1	19648300	rs114891776	A	G	0.0102	2.09E-08	0.057	Yes
1	22442025	rs1474646	G	A	0.0054	2.21E-11	-0.0363	
1	59653742	rs3889199	G	A	0.0054	8.72E-09	-0.0312	Yes
1	61831890	rs1125777	T	C	0.0048	1.58E-09	0.0291	Yes
1	92315658	rs12094295	A	G	0.0067	1.75E-08	0.0377	Yes
1	110364821	rs2813773	A	G	0.0052	1.01E-16	0.0431	
1	112832256	rs4268356	C	G	0.0089	1.53E-08	-0.0503	Yes
1	150479605	rs114293104	T	C	0.0179	3.58E-11	-0.1187	Yes
1	151904109	rs56313205	T	C	0.0053	1.70E-23	-0.0531	
1	154791995	rs1218600	A	C	0.0055	1.30E-08	-0.0313	Yes
1	174147516	rs4650958	A	G	0.005	2.79E-13	-0.0369	Yes
1	204319907	rs7513240	T	C	0.0049	6.34E-11	-0.0317	Yes
1	207481316	rs6698351	A	G	0.0056	1.91E-11	-0.0375	Yes
1	215474906	rs11120546	G	A	0.0046	3.10E-09	-0.0271	Yes
1	219294570	rs61823192	T	C	0.0164	1.27E-40	-0.2193	
1	221073294	rs2784274	T	C	0.0043	1.94E-18	-0.0374	
1	234352899	rs4333882	G	A	0.0054	3.47E-90	0.109	
1	237275789	rs12139283	A	C	0.0047	1.80E-10	-0.03	Yes
1	237989132	rs12024181	T	A	0.005	1.51E-08	-0.0286	Yes
1	245773041	rs12041565	T	C	0.0059	1.38E-15	0.0474	
2	15958795	rs4471862	A	G	0.0051	2.36E-11	0.0343	Yes
2	18868716	rs11096554	G	A	0.0051	4.77E-12	0.0354	
2	20039356	rs12614685	T	C	0.0047	2.05E-09	0.0279	Yes
2	28288392	rs10173528	T	C	0.0048	2.07E-10	-0.0307	
2	33361425	rs6714546	G	A	0.005	3.15E-22	-0.0482	
2	36768447	rs10084382	T	C	0.0049	2.81E-13	0.0361	
2	42131643	rs11694852	C	T	0.0048	6.32E-09	-0.028	
2	43306379	rs895518	G	A	0.005	6.60E-10	0.0308	Yes
2	56040099	rs10199082	C	T	0.007	4.61E-48	0.1024	
2	71675642	rs10187616	T	C	0.0065	1.11E-08	0.0369	Yes
2	104454345	rs1441098	T	A	0.0049	1.82E-09	-0.0294	Yes
2	144305674	rs6704689	A	G	0.0058	4.11E-148	-0.1494	
2	183509955	rs10177196	C	T	0.0046	2.06E-09	-0.0277	Yes
2	203131937	rs7581542	G	A	0.0071	1.60E-16	-0.0583	Yes
2	218796452	rs11895557	A	G	0.0055	1.36E-27	-0.0595	
2	240076712	rs3791503	A	G	0.0049	5.26E-09	0.0288	Yes
2	242630864	rs75447999	C	A	0.0045	6.60E-09	0.0259	Yes
3	5900429	rs6768341	G	A	0.0068	2.06E-17	-0.058	
3	7904532	rs13323901	A	C	0.0058	1.81E-10	0.037	Yes
3	15502681	rs7609897	T	G	0.0058	2.60E-45	-0.0823	
3	25209801	rs7645872	T	C	0.0048	3.34E-09	0.0283	Yes
3	30492708	rs11705959	A	C	0.0048	1.14E-08	0.0275	Yes
3	42496944	rs1354993	T	C	0.0052	2.63E-16	0.0425	
3	48668170	rs114135552	G	C	0.011	1.32E-09	0.067	Yes
3	52467263	rs648514	A	G	0.0046	9.79E-14	0.0342	
3	55602137	rs61613824	A	T	0.0048	7.81E-13	-0.0341	Yes

Supplemental Table 1. 236 genome-wide significant loci disocvered in the genome-wide association study meta-analysis. (Note: only first page displayed.)

Locus (chr_pos)	Signal					
	CHR	POSITION	RSID	EFFECT_ALL	OTHER_ALL	Prioritized Gene
10_101394572	10	101393872	rs7911100	A	G	SLC25A28
10_101394572	10	101344263	rs35564340	A	G	SLC25A28
10_116722834	10	116686790	rs185963450	A	T	TRUB1
10_121290571	10	121290571	rs4752324	C	T	RGS10
10_12280296	10	12280296	rs10906104	G	A	CDC123
10_124222105	10	124220667	rs55928386	T	C	HTRA1
10_125134743	10	125134743	rs4980285	A	C	
10_131752583	10	131752583	rs11017029	T	C	EBF3
10_134420997	10	134420997	rs12359407	C	G	INPP5A
10_18440444	10	18223077	rs691444	C	G	CACNB2
10_18440444	10	18440444	rs1888693	A	G	CACNB2
10_18440444	10	18467169	rs143821222	C	G	CACNB2
10_21935248	10	21820650	rs564819152	A	G	MLLT10
10_25811435	10	25728493	rs11014558	T	G	GPR158
10_25811435	10	25811435	rs7086249	C	T	GPR158
10_36461567	10	36461567	rs1451193	A	G	FZD8
10_36461567	10	36410710	rs11010468	T	C	
10_50324685	10	50324685	rs11100988	T	G	VSTM4
10_53271054	10	53450806	rs75455260	C	T	PRKG1
10_53271054	10	53264935	rs144359077	C	G	PRKG1
10_63243409	10	63237354	rs2787724	A	G	TMEM26
10_69758264	10	69758264	rs12783464	G	A	HERC4
10_69758264	10	69548701	rs113776658	T	C	HERC4
10_8710715	10	8718004	rs9733843	G	A	
10_97516612	10	97323370	rs12355467	A	G	ENTPD1
10_97516612	10	97375268	rs11188403	A	G	ENTPD1
10_97516612	10	97516612	rs74671500	T	C	ENTPD1
10_97516612	10	97323371	rs12358528	A	G	ENTPD1
10_97516612	10	97447487	rs11188434	A	T	ENTPD1
11_10331664	11	10331664	rs4444073	C	A	AMPD3
11_106818037	11	106818037	rs1455588	A	G	GUCY1A2
11_113231691	11	113204266	rs11214577	T	C	TTC12
11_113782139	11	113791278	rs12273321	A	T	HTR3B
11_116715784	11	116715784	rs142174850	A	G	SIK3
11_128422541	11	128420873	rs4245078	C	T	ETS1
11_130275749	11	130491732	rs7110242	G	T	ADAMTS8
11_130275749	11	130275749	rs11222085	C	T	ADAMTS8
11_1507731	11	1505454	rs2334652	G	A	MOB2
11_15089761	11	15215059	rs10832370	A	T	CALCB
11_15089761	11	15087959	rs11023422	C	A	CALCB

Supplemental Table 2. Genes prioritized for each signal (index variants and secondary signals). Locus column displays the locus in which the signal resides. (Note: only first page displayed.)

Method	Phenotype	Source study	Data access link
LDSC	Abominal Aortic Aneurysm (AAA)	PMID:37845353	https://csg.sph.umich.edu/willer/public/AAAgen2023/
	Ventral hernia (ICD 10: K43)	Neale lab consortium (https://nealelab.github.io/UKBB_ldsc/downloads.htm	https://www.dropbox.com/s/pcsb8gmc1mkz9vf/K43.ldsc.imputed_v3.both_sexes.tsv.bgz?dl=1
	Umbilical hernia (ICD 10: K42)	Neale lab consortium (https://nealelab.github.io/UKBB_ldsc/downloads.htm	https://www.dropbox.com/s/yhcnh1g8ibgl0ad/K42.ldsc.imputed_v3.both_sexes.tsv.bgz?dl=1
	Pelvic organ prolapse (ICD 10: N81), female	Neale lab consortium (https://nealelab.github.io/UKBB_ldsc/downloads.htm	https://www.dropbox.com/s/zom3k4433r53sq4/N81.ldsc.imputed_v3.female.tsv.bgz?dl=1
	Ulcerative colitis (ICD 10: K51)	Neale lab consortium (https://nealelab.github.io/UKBB_ldsc/downloads.htm	https://www.dropbox.com/s/c17owvqlw26vulr/K51.ldsc.imputed_v3.both_sexes.tsv.bgz?dl=1
MR	Microbiome	PMID:33462485	https://mibiogen.gcc.rug.nl/
	Plasma proteome	PMID: 34857953	https://www.decode.com/summarydata/
		PMID:37794186	https://metabolomips.org/ukbbpgwas/
RNAseq	Healthy human colon	PMID:37468586 (Hickey et al.)	https://datadryad.org/dataset/doi:10.5061/dryad.8pk0p2ns8
	Healthy human colon	PMID:33444816 (Wright et al.)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156905

LDSC = Linkage disequlibrium score regression
MR = Mendelian randomization
PMID = PubMed identifier

Supplemental Table 3. Input data sources.

Exposure	Beta	P-value	StdErr	CI_upper	CI_lower	Outcome	Num_SNPs	F_stat	Aptamer_id	P_adj	Negative_log_pva	Actionable_target
ABO	-0.065	2.83E-86	0.0033	-0.058	-0.071	DivDz	148	4008.01	9253_52	4.99E-83	85.55	no
CALCB	-0.276	1.70E-20	0.0297	-0.218	-0.334	DivDz	15	71.53	17170_15	1.50E-17	19.77	yes
NOV	-0.627	8.15E-20	0.0688	-0.492	-0.762	DivDz	1	69.59	2737_22	4.80E-17	19.09	no
MIF	0.295	1.95E-17	0.0348	0.363	0.227	DivDz	3	159.76	8221_19	8.60E-15	16.71	no
EFEMP1	-0.427	1.03E-14	0.0552	-0.319	-0.535	DivDz	9	59.95	8480_29	3.65E-12	13.99	no
ANTXR2	0.123	9.95E-12	0.0181	0.158	0.088	DivDz	17	65.27	15559_5	2.93E-09	11.00	no
ITIH4	-0.225	1.65E-09	0.0374	-0.152	-0.299	DivDz	5	58.28	4811_33	4.17E-07	8.78	no
DPT	-0.052	1.89E-07	0.0099	-0.032	-0.071	DivDz	61	79.38	4979_34	3.34E-05	6.72	no
LTBP4	-0.383	1.67E-07	0.0732	-0.240	-0.527	DivDz	2	40.60	13133_73	3.34E-05	6.78	no
CHRDL2	0.077	1.81E-07	0.0147	0.106	0.048	DivDz	29	76.26	6086_15	3.34E-05	6.74	no
MFAP4	0.099	6.26E-07	0.0198	0.137	0.060	DivDz	11	102.26	5636_10	0.00010053	6.20	no
LIMA1	0.383	1.23E-06	0.0790	0.538	0.228	DivDz	1	55.27	11543_84	0.00018072	5.91	no
TNFSF12	0.095	2.58E-06	0.0202	0.135	0.055	DivDz	30	64.22	5939_42	0.00032541	5.59	yes
TIMP2	-0.283	2.49E-06	0.0602	-0.165	-0.401	DivDz	2	57.87	2278_61	0.00032541	5.60	no
LINGO1	-0.205	3.69E-06	0.0444	-0.118	-0.292	DivDz	3	56.76	6620_82	0.00043439	5.43	yes
MUSK	0.304	9.16E-06	0.0686	0.439	0.170	DivDz	2	43.79	11547_84	0.00101071	5.04	no
CTRB2	-0.023	1.34E-05	0.0053	-0.013	-0.034	DivDz	83	452.24	5648_28	0.00139399	4.87	no
GSTT1	-0.026	1.60E-05	0.0061	-0.014	-0.038	DivDz	76	295.87	19230_12	0.00148661	4.80	no
USP8	0.280	1.57E-05	0.0649	0.407	0.153	DivDz	1	86.96	13450_49	0.00148661	4.80	no
AES	0.449	3.47E-05	0.1084	0.661	0.236	DivDz	1	43.04	18396_10	0.00248948	4.46	no
WIF1	0.296	3.52E-05	0.0715	0.436	0.156	DivDz	2	34.52	16070_7	0.00248948	4.45	no
PDE5A	-0.091	3.19E-05	0.0219	-0.048	-0.134	DivDz	26	45.72	5256_86	0.00248948	4.50	yes
PAM	0.032	3.50E-05	0.0078	0.047	0.017	DivDz	33	630.33	5620_13	0.00248948	4.46	no
NEU1	0.468	3.39E-05	0.1129	0.690	0.247	DivDz	1	34.00	15426_5	0.00248948	4.47	no
MDH1	-0.272	2.87E-05	0.0649	-0.144	-0.399	DivDz	1	77.31	3853_56	0.00248948	4.54	no
ADAMTS13	-0.038	4.06E-05	0.0093	-0.020	-0.056	DivDz	103	272.75	3175_51	0.00264117	4.39	no
ELANE	-0.043	4.34E-05	0.0104	-0.022	-0.063	DivDz	63	79.83	13671_40	0.00264117	4.36	no
COL15A1	0.106	4.24E-05	0.0260	0.157	0.055	DivDz	20	50.88	8974_172	0.00264117	4.37	no
THBS2	-0.044	4.26E-05	0.0108	-0.023	-0.065	DivDz	78	102.75	3339_33	0.00264117	4.37	no
LGALS3	0.044	4.92E-05	0.0108	0.065	0.023	DivDz	20	165.91	3066_12	0.0028942	4.31	no
DNAJB12	-0.236	8.15E-05	0.0600	-0.119	-0.354	DivDz	1	92.40	8006_12	0.00464182	4.09	no
RCN1	-0.173	8.66E-05	0.0442	-0.087	-0.260	DivDz	5	55.52	9322_15	0.00475534	4.06	no
MAPK9	-0.104	8.89E-05	0.0266	-0.052	-0.157	DivDz	12	43.02	15604_18	0.00475534	4.05	yes
PPP2R5A	-0.269	0.00010526	0.0694	-0.133	-0.405	DivDz	2	37.44	13622_16	0.00546723	3.98	no
FABP2	-0.031	0.0001265	0.0081	-0.015	-0.047	DivDz	56	354.59	15385_116	0.00638258	3.90	no
FUT3	-0.030	0.00013611	0.0079	-0.015	-0.046	DivDz	59	135.19	4548_4	0.00667676	3.87	no
CILP2	-0.173	0.00015005	0.0457	-0.084	-0.263	DivDz	3	51.04	8841_65	0.00716174	3.82	no
SEMA3E	-0.021	0.0002013	0.0057	-0.010	-0.032	DivDz	96	419.11	5363_51	0.0093551	3.70	no
CBR3	-0.048	0.00022751	0.0129	-0.022	-0.073	DivDz	36	106.79	14091_42	0.01030191	3.64	no
XPNPEP1	-0.250	0.00025122	0.0683	-0.116	-0.384	DivDz	1	55.84	3481_87	0.01082068	3.60	no
TTL	0.324	0.00025028	0.0884	0.497	0.150	DivDz	1	52.14	13973_62	0.01082068	3.60	no
CACYBP	-0.141	0.0002668	0.0387	-0.065	-0.217	DivDz	1	220.04	12432_23	0.01121831	3.57	no
LOXL3	-0.156	0.00030606	0.0433	-0.071	-0.241	DivDz	2	101.33	15427_35	0.01256984	3.51	yes
ATP1B2	-0.081	0.00032058	0.0226	-0.037	-0.125	DivDz	39	51.67	7218_87	0.01286686	3.49	yes
FAM213A	-0.179	0.0003694	0.0502	-0.080	-0.277	DivDz	4	41.74	13423_94	0.01388011	3.43	no
CD209	-0.046	0.00036355	0.0130	-0.021	-0.072	DivDz	34	89.63	3029_52	0.01388011	3.44	no
DNAJB1	0.335	0.00035504	0.0937	0.518	0.151	DivDz	1	40.61	3852_19	0.01388011	3.45	no
MAP3K3	-0.302	0.00041035	0.0854	-0.134	-0.469	DivDz	1	35.77	12990_39	0.01509763	3.39	yes
COL6A1	0.022	0.00043974	0.0062	0.034	0.010	DivDz	98	763.01	16828_8	0.01553148	3.36	no
EPHB2	0.093	0.00043308	0.0264	0.145	0.041	DivDz	16	53.57	8225_86	0.01553148	3.36	yes
CD55	0.063	0.00047703	0.0180	0.098	0.028	DivDz	31	109.15	5069_9	0.01651828	3.32	no
GSTT2B	0.067	0.00050187	0.0191	0.104	0.029	DivDz	26	93.79	11273_176	0.01704439	3.30	no
PRKCSH	-0.163	0.00053811	0.0470	-0.070	-0.255	DivDz	3	56.75	5687_5	0.01782263	3.27	no
RETN	0.041	0.00054497	0.0118	0.064	0.018	DivDz	42	80.84	3046_31	0.01782263	3.26	no
PRPSAP2	-0.308	0.00059483	0.0898	-0.132	-0.485	DivDz	1	51.99	19150_20	0.01909935	3.23	no
NAPG	0.219	0.00071704	0.0646	0.345	0.092	DivDz	1	79.05	17773_26	0.02146264	3.14	no
FBLN1	0.085	0.0007139	0.0252	0.135	0.036	DivDz	15	42.89	6470_19	0.02146264	3.15	no
ADK	-0.273	0.00070189	0.0805	-0.115	-0.430	DivDz	2	30.10	19207_119	0.02146264	3.15	no
PPP1R14A	0.165	0.00068573	0.0485	0.260	0.070	DivDz	14	49.31	19117_3	0.02146264	3.16	no
BOLA1	0.032	0.00078765	0.0096	0.051	0.013	DivDz	18	1157.09	15370_5	0.02318303	3.10	no
CRAT	0.178	0.00081523	0.0533	0.283	0.074	DivDz	2	67.24	12637_7	0.0236016	3.09	no
PYY	-0.240	0.00083225	0.0717	-0.099	-0.380	DivDz	5	30.68	3727_35	0.02370561	3.08	no
TXNDC12	0.212	0.00093296	0.0640	0.337	0.086	DivDz	1	121.14	19334_62	0.02615249	3.03	no
MCL1	-0.133	0.00100786	0.0404	-0.054	-0.212	DivDz	2	85.99	10396_6	0.02781056	3.00	yes
HSPA1A	0.240	0.00105793	0.0734	0.384	0.097	DivDz	2	33.98	4124_24	0.02830772	2.98	no
PXDN	0.078	0.00104671	0.0238	0.125	0.031	DivDz	12	51.62	13463_1	0.02830772	2.98	no
ENO2	0.110	0.00111705	0.0339	0.177	0.044	DivDz	2	153.87	10339_48	0.02859004	2.95	no
SCARA5	0.072	0.00110304	0.0220	0.115	0.029	DivDz	33	60.27	10419_1	0.02859004	2.96	no
ANXA1	0.096	0.00110562	0.0294	0.154	0.038	DivDz	4	77.84	4960_72	0.02859004	2.96	no
RELT	0.269	0.0011953	0.0831	0.432	0.106	DivDz	1	46.03	14112_40	0.03015564	2.92	no
NUDCD3	0.254	0.00122355	0.0786	0.408	0.100	DivDz	1	74.18	4254_6	0.03043358	2.91	no
HSP90B1	-0.023	0.00139488	0.0071	-0.009	-0.036	DivDz	87	192.24	6393_63	0.0342134	2.86	no
PPCS	-0.149	0.00143508	0.0468	-0.057	-0.241	DivDz	1	147.81	17843_30	0.03471725	2.84	no
SPINT2	0.042	0.00145657	0.0131	0.067	0.016	DivDz	58	129.73	2843_13	0.03476086	2.84	no
THBS1	-0.291	0.00155811	0.0920	-0.111	-0.471	DivDz	1	34.56	3474_19	0.03587231	2.81	no
FLT4	-0.119	0.00156408	0.0377	-0.045	-0.193	DivDz	11	36.98	16035_8	0.03587231	2.81	yes
ARFIP1	-0.069	0.0015598	0.0217	-0.026	-0.111	DivDz	13	60.56	13488_3	0.03587231	2.81	no

Supplemental Table 4. Full results of pQTL MR using deCODE data. (Note: only first page displayed.)

Exposure	Beta	P-value	StdErr	CI_upper	CI_lower	Outcome	Num_SNPs	F_stat	Aptamer_id	P_adj	Negative_log_pval	Actionable_target
ABO	-0.061	4.07E-27	0.0057	-0.050	-0.072	DivDz	114	222.62	OID30675	8.03E-24	26.39	no
STX4	-0.851	1.24E-15	0.1063	-0.642	-1.059	DivDz	1	30.79	OID21326	1.23E-12	14.91	no
CTF1	0.679	2.51E-15	0.0858	0.847	0.511	DivDz	1	45.86	OID20061	1.42E-12	14.60	no
EFEMP1	-0.363	2.87E-15	0.0460	-0.273	-0.453	DivDz	19	51.48	OID20414	1.42E-12	14.54	no
TNFSF12	0.088	2.06E-10	0.0139	0.115	0.061	DivDz	38	81.95	OID20624	7.10E-08	9.69	yes
PI16	0.138	2.16E-10	0.0217	0.180	0.095	DivDz	37	65.99	OID30676	7.10E-08	9.67	no
CHRDL2	0.090	1.84E-09	0.0149	0.119	0.060	DivDz	23	100.07	OID20223	5.18E-07	8.74	no
C2	0.138	7.23E-09	0.0238	0.184	0.091	DivDz	3	319.36	OID20410	1.79E-06	8.14	no
DPT	-0.065	1.35E-08	0.0114	-0.042	-0.087	DivDz	59	95.37	OID20278	2.96E-06	7.87	no
COL15A1	0.118	2.85E-08	0.0212	0.159	0.076	DivDz	21	78.42	OID30396	5.63E-06	7.55	no
TNFRSF6B	0.101	7.73E-08	0.0188	0.138	0.064	DivDz	63	55.55	OID20964	1.39E-05	7.11	no
USP28	-0.273	8.84E-08	0.0510	-0.173	-0.373	DivDz	3	59.28	OID31084	1.46E-05	7.05	no
MFAP4	0.102	3.40E-07	0.0200	0.141	0.063	DivDz	15	88.44	OID30598	5.16E-05	6.47	no
SLC4A1	-0.387	5.26E-07	0.0771	-0.236	-0.538	DivDz	2	38.96	OID30369	7.42E-05	6.28	no
APOA1	0.313	1.59E-06	0.0653	0.441	0.185	DivDz	4	29.83	OID30769	0.00020941	5.80	no
AAMDC	0.040	2.05E-06	0.0084	0.056	0.023	DivDz	42	184.60	OID30236	0.00025263	5.69	no
NID2	0.041	3.01E-06	0.0089	0.059	0.024	DivDz	83	93.00	OID21085	0.00031328	5.52	no
PAM	0.036	2.86E-06	0.0076	0.050	0.021	DivDz	39	216.71	OID20291	0.00031328	5.54	no
CAMKK1	-0.074	2.89E-06	0.0159	-0.043	-0.106	DivDz	11	130.26	OID21182	0.00031328	5.54	no
MANF	0.292	4.03E-06	0.0633	0.416	0.168	DivDz	3	37.58	OID20707	0.00039831	5.39	no
SCARA5	0.079	5.64E-06	0.0175	0.114	0.045	DivDz	45	61.74	OID21058	0.00053112	5.25	no
SPINT2	0.202	9.36E-06	0.0456	0.292	0.113	DivDz	14	54.11	OID20720	0.000841	5.03	no
GPR158	-0.351	1.12E-05	0.0800	-0.195	-0.508	DivDz	1	62.10	OID31013	0.00096422	4.95	no
PRKAR2A	0.425	1.55E-05	0.0984	0.618	0.232	DivDz	1	42.00	OID31119	0.00127635	4.81	no
TFPI	0.068	1.78E-05	0.0158	0.099	0.037	DivDz	32	80.69	OID20388	0.00135303	4.75	no
PPP1R14A	0.164	1.74E-05	0.0382	0.239	0.089	DivDz	19	51.75	OID31011	0.00135303	4.76	no
MAPK9	-0.155	1.87E-05	0.0362	-0.084	-0.226	DivDz	8	33.85	OID20557	0.00137087	4.73	yes
HSPA1A	-0.248	2.67E-05	0.0591	-0.132	-0.364	DivDz	4	32.86	OID20718	0.0018846	4.57	no
ADM	0.125	2.79E-05	0.0299	0.184	0.067	DivDz	15	56.88	OID21467	0.00190323	4.55	no
AGER	-0.236	3.02E-05	0.0566	-0.125	-0.347	DivDz	18	106.98	OID20756	0.00199062	4.52	no
SUSD2	-0.045	3.50E-05	0.0108	-0.023	-0.066	DivDz	41	146.03	OID21098	0.00216345	4.46	no
SHBG	0.084	3.50E-05	0.0202	0.123	0.044	DivDz	33	69.37	OID30685	0.00216345	4.46	yes
PHLDB1	0.475	3.87E-05	0.1153	0.701	0.249	DivDz	1	33.63	OID31467	0.00231452	4.41	no
ELN	-0.656	4.20E-05	0.1601	-0.342	-0.970	DivDz	5	61.13	OID30301	0.00237328	4.38	yes
LRRC37A2	-0.027	4.19E-05	0.0065	-0.014	-0.040	DivDz	45	524.88	OID31440	0.00237328	4.38	no
OGA	-0.132	5.44E-05	0.0327	-0.068	-0.196	DivDz	9	44.35	OID31465	0.00298715	4.26	no
PKD1	-0.088	6.17E-05	0.0220	-0.045	-0.131	DivDz	28	95.48	OID30328	0.00329729	4.21	no
FLT4	-0.066	7.20E-05	0.0167	-0.034	-0.099	DivDz	43	51.99	OID21491	0.00374569	4.14	yes
TNF	-0.269	9.99E-05	0.0692	-0.134	-0.405	DivDz	1	97.72	OID20074	0.00506005	4.00	yes
CD302	-0.058	0.00010259	0.0149	-0.029	-0.087	DivDz	27	73.31	OID21328	0.00506806	3.99	no
NUDT5	-0.278	0.00011648	0.0722	-0.137	-0.419	DivDz	2	40.84	OID21115	0.00526572	3.93	no
RELT	0.072	0.00011725	0.0186	0.108	0.035	DivDz	30	97.67	OID21142	0.00526572	3.93	no
TNF	-0.217	0.00011438	0.0563	-0.107	-0.328	DivDz	3	51.50	OID20473	0.00526572	3.94	yes
ZPR1	0.324	0.00011243	0.0840	0.489	0.160	DivDz	1	43.14	OID30801	0.00526572	3.95	no
GSTT2B	0.017	0.00012345	0.0045	0.026	0.008	DivDz	93	313.37	OID31097	0.00542077	3.91	no
AP3B1	-0.367	0.00014242	0.0965	-0.178	-0.556	DivDz	1	42.31	OID31478	0.00599076	3.85	no
NCAN	-0.099	0.00014249	0.0261	-0.048	-0.150	DivDz	12	56.72	OID21055	0.00599076	3.85	no
AKT3	-0.304	0.00015158	0.0803	-0.147	-0.462	DivDz	1	48.68	OID21197	0.00624017	3.82	yes
AFAP1	-0.022	0.00016062	0.0057	-0.010	-0.033	DivDz	134	171.20	OID30569	0.00647739	3.79	no
AOC3	-0.043	0.0001975	0.0116	-0.020	-0.066	DivDz	29	212.85	OID20389	0.00780517	3.70	yes
HEXIM1	0.231	0.00021376	0.0623	0.353	0.109	DivDz	3	42.68	OID20589	0.00828215	3.67	no
IL12B	-0.037	0.00023208	0.0102	-0.018	-0.057	DivDz	55	102.47	OID20666	0.00881905	3.63	yes
LPA	0.023	0.00026401	0.0064	0.036	0.011	DivDz	105	189.16	OID30747	0.00984293	3.58	no
GMPR	-0.065	0.00032033	0.0181	-0.030	-0.100	DivDz	26	54.88	OID20641	0.01172177	3.49	no
BTD	-0.060	0.00032844	0.0168	-0.027	-0.093	DivDz	65	108.71	OID30754	0.01179993	3.48	no
SMARCA2	-0.347	0.00033758	0.0969	-0.157	-0.537	DivDz	1	37.97	OID20796	0.01191159	3.47	no
PRSS53	-0.027	0.00035574	0.0075	-0.012	-0.042	DivDz	61	243.22	OID31093	0.01211953	3.45	no
DNAJB1	0.326	0.00035504	0.0912	0.504	0.147	DivDz	1	49.78	OID21363	0.01211953	3.45	no
LAMA4	-0.088	0.00047783	0.0251	-0.038	-0.137	DivDz	16	46.15	OID20769	0.01565079	3.32	no
MINDY1	-0.111	0.00047122	0.0317	-0.049	-0.173	DivDz	5	98.32	OID31399	0.01565079	3.33	no
NCF2	-0.139	0.00048315	0.0399	-0.061	-0.218	DivDz	3	117.73	OID20567	0.01565079	3.32	no
PLXDC2	0.179	0.00052046	0.0516	0.280	0.078	DivDz	9	41.79	OID31101	0.01632419	3.28	no
IGHMBP2	-0.224	0.00051932	0.0644	-0.097	-0.350	DivDz	1	101.34	OID30123	0.01632419	3.28	no
SCARF2	-0.065	0.00053401	0.0188	-0.028	-0.102	DivDz	22	118.33	OID21048	0.01648771	3.27	no
CD209	-0.047	0.00062875	0.0137	-0.020	-0.074	DivDz	42	119.37	OID20262	0.01911413	3.20	no
TNF	-0.207	0.00065393	0.0609	-0.088	-0.327	DivDz	3	42.63	OID21237	0.01957828	3.18	yes
AHNAK2	-0.054	0.0006841	0.0160	-0.023	-0.086	DivDz	23	124.75	OID30244	0.02017575	3.16	no
NUCB2	0.050	0.00074711	0.0150	0.080	0.021	DivDz	44	84.54	OID21448	0.02171023	3.13	no
CD55	0.063	0.00077752	0.0188	0.100	0.026	DivDz	39	81.89	OID20377	0.02226626	3.11	no
CHMP1A	-0.144	0.00082482	0.0429	-0.059	-0.228	DivDz	11	22.78	OID20930	0.02328349	3.08	no
ARSB	0.052	0.00090992	0.0157	0.083	0.021	DivDz	37	70.60	OID21331	0.02499275	3.04	no
PPCDC	-0.045	0.00093596	0.0135	-0.018	-0.071	DivDz	26	106.28	OID21027	0.02499275	3.03	no
EBAG9	-0.247	0.00092062	0.0744	-0.101	-0.392	DivDz	1	78.90	OID20918	0.02499275	3.04	no
LYSMD3	-0.253	0.00093455	0.0764	-0.103	-0.402	DivDz	1	47.98	OID30950	0.02499275	3.03	no
ZBTB17	-0.500	0.00101724	0.1521	-0.202	-0.798	DivDz	1	37.34	OID20082	0.02680092	2.99	no
MIF	0.183	0.00104905	0.0558	0.292	0.074	DivDz	3	43.15	OID21073	0.02692107	2.98	no
CEP170	0.223	0.00104681	0.0682	0.357	0.090	DivDz	2	46.06	OID30240	0.02692107	2.98	no

Supplemental Table 5. Full results of pQTL MR using UK Biobank Pharma Proteomics Project (UKB-PPP) data. (Note: only first page displayed.)